The Alan Turing Institute

Towards Openness Beyond Open Access

User Journeys through 3 Open AI Collaboratives

Jennifer Ding | The Alan Turing Institute

Christopher Akiki, Yacine Jernite, Anne Lee Steele, Temi Popo

@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

Overview

1. Open Science at The Alan Turing Institute
2. Approaches to Open Source in AI
3. The Open AI Collaborative
4. User Journeys through 3 Open AI Collaboratives

@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

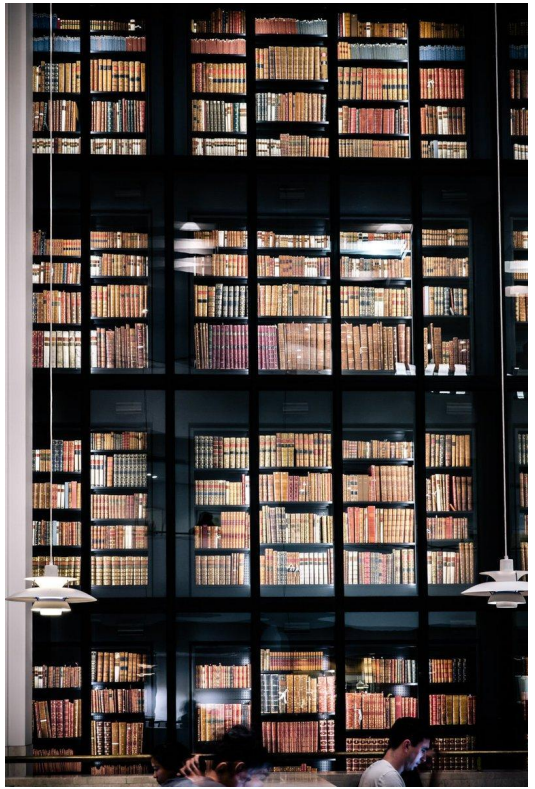
Overview

1. ***Open Science at The Alan Turing Institute***
2. Approaches to Open Source in AI
3. The Open AI Collaborative
4. User Journeys through 3 Open AI Collaboratives

@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

The Alan Turing Institute

- UK’s national institute for DS & AI
- Advance world-class research and apply it to real-world problems

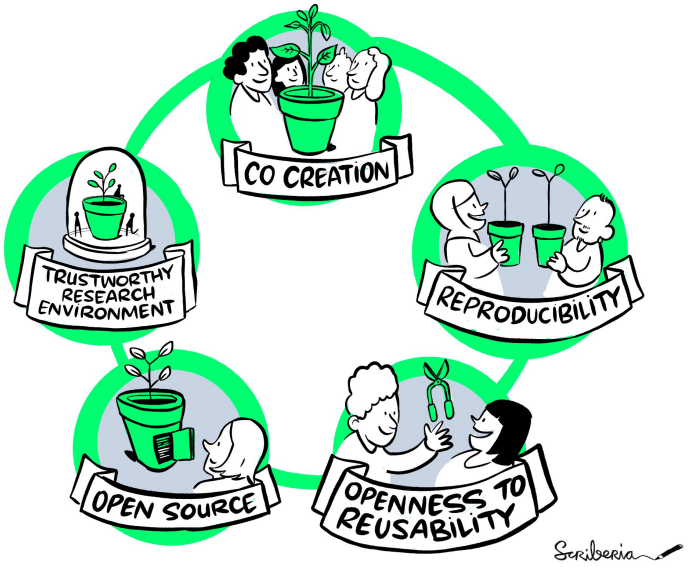


@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

RAMs in the Alan Turing Institute

- Positioned horizontally across Turing to drive culture change by advocating for a **Team Science** mentality & open ways of working

- Operationalise best practices in interoperability, reproducibility and re-usability on research teams with a focus on **real-world users & impact**



@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

Overview

1. Open Science at The Alan Turing Institute
2. ***Approaches to Open Source in AI***
3. The Open AI Collaborative
4. User Journeys through 3 Open AI Collaboratives

@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

Open ML Access

scikit-learn is a Python module for machine learning built on top of SciPy and is distributed under the 3-Clause BSD license.

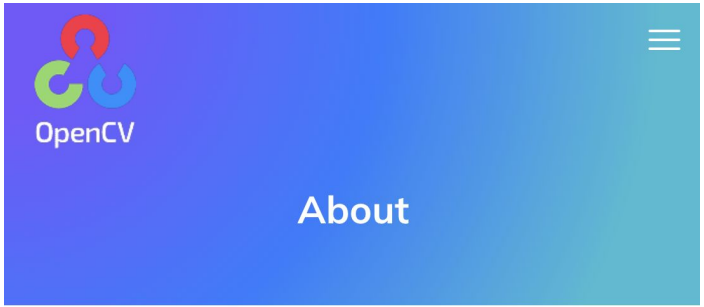
The project was started in 2007 by David Cournapeau as a Google Summer of Code project, and since then many volunteers have contributed. See the About us page for a list of core contributors.

It is currently maintained by a team of volunteers.

```
                              YOLO LICENSE
                          Version 2, July 29 2016

THIS SOFTWARE LICENSE IS PROVIDED "ALL CAPS" SO THAT YOU KNOW IT IS SUPER
SERIOUS AND YOU DON'T MESS AROUND WITH COPYRIGHT LAW BECAUSE YOU WILL GET IN
TROUBLE HERE ARE SOME OTHER BUZZWORDS COMMONLY IN THESE THINGS WARRANTIES
LIABILITY CONTRACT TORT LIABLE CLAIMS RESTRICTION MERCHANTABILITY. NOW HERE'S
THE REAL LICENSE:

0. Darknet is public domain.
1. Do whatever you want with it.
2. Stop emailing me about it!
```

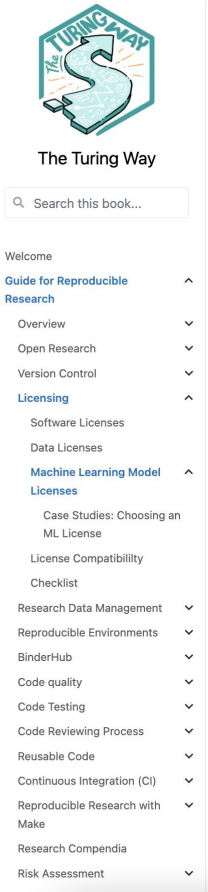


OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. Being an Apache 2 licensed product, OpenCV makes it easy for businesses to utilize and modify the code.

@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

Open ML Licensing

- Started with repurposing of MIT & Apache 2.0 licenses…
- Closed or bespoke open licenses for models today



Examples of ML models and their licenses

The table below showcases several well-known examples of ML models in the fields of NLP, vision, and multimodal generatives. The aim of it is to inform the reader on the licensing options chosen by each of the projects which sometimes differ from one another. Licensing differences might stem from business models, research purposes or ethics-informed community values. Each license carries licensor's values and a message from the former to potential users on how the model should be used.

Model	Model License	Description	Link to License
GPT-2	MIT License + generated output disclaimer	Permissive open source license	Link to license
GPT-3	Exclusive	Licensed to Microsoft	News Article
YOLO	YOLO License	Public domain license	Link to license
DALLE-pytorch	MIT License	Permissive open source license	Link to license
Stable Diffusion	CreativeML Open RAIL-M	Open & Responsible AI License (RAIL) created by Stability.ai and adapted from the BLOOM RAIL license, including use restrictions (see attachment A)	Link to license
OPT	OPT-175B License	Meta restrictive license enabling use of the model weights for research purposes while establishing a set of use restrictions, which could be considered a RAIL	Link to license
BigScience	BigScience OpenRAIL-M	Open & Responsible AI License (RAIL) created by BigScience and adapted from the BLOOM RAIL license, including use restrictions (see attachment A)	Link to license
Tsinghua University	GLM-130B license	Restrictive license enabling use of the model weights for research purposes	Link to license

@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

Overview

1. Open Science at The Alan Turing Institute
2. Approaches to Open Source in AI
3. ***The Open AI Collaborative***
4. User Journeys through 3 Open AI Collaboratives

@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

Towards Openness Beyond Open Access: User Journeys through 3 Open AI Collaboratives

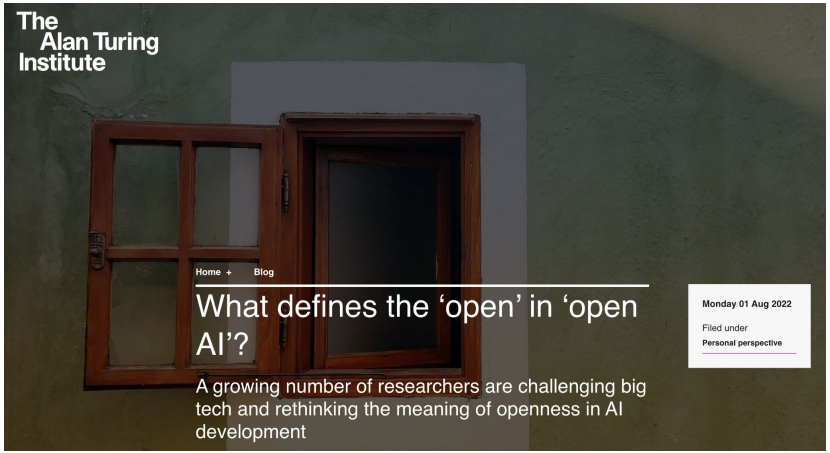
Jennifer Ding
The Alan Turing Institute
jding@turing.ac.uk

Christopher Akiki
Leipzig University
christopher.akiki@uni-leipzig.de

Yacine Jernite
Hugging Face
yacine@huggingface.co

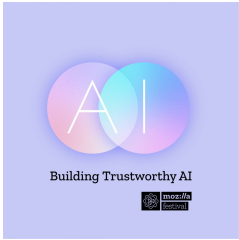
Anne Lee Steele
The Alan Turing Institute
asteele@turing.ac.uk

Temi Popo
Mozilla Foundation
temi@mozillafoundation.org



Abstract

Open Artificial Intelligence (Open AI) collaboratives offer alternative pathways for how AI can be developed beyond well-resourced technology companies and who can be a part of the process. To understand how and why they work and what additionality they bring to the landscape, we focus on three such communities, each focused on a different kind of activity around AI: building *models* (BigScience workshop), *tools/ways of working* (The Turing Way), and *ecosystems* (Mozilla Festival's Building Trustworthy AI Working Group). First, we document the community structures that facilitate these distributed, volunteer-led teams, comparing the collaboration styles that drive each group towards their specific goals. Through interviews with community leaders, we map user journeys for how members discover, join, contribute, and participate. Ultimately, this paper aims to highlight the diversity of AI work and workers that have come forth through these collaborations and how they offer a broader practice of openness to the AI space.

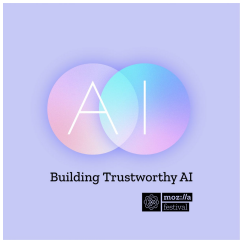


@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

What is an Open AI collaborative?

- Distributed, volunteer-led teams creating alternative AI development pathways grounded in “open” principles

- Focus on activities relevant for their respective communities:
  - building models (BigScience Workshop)
  - tools/ways of working (The Turing Way)
  - ecosystems (MozFest TAIWG)

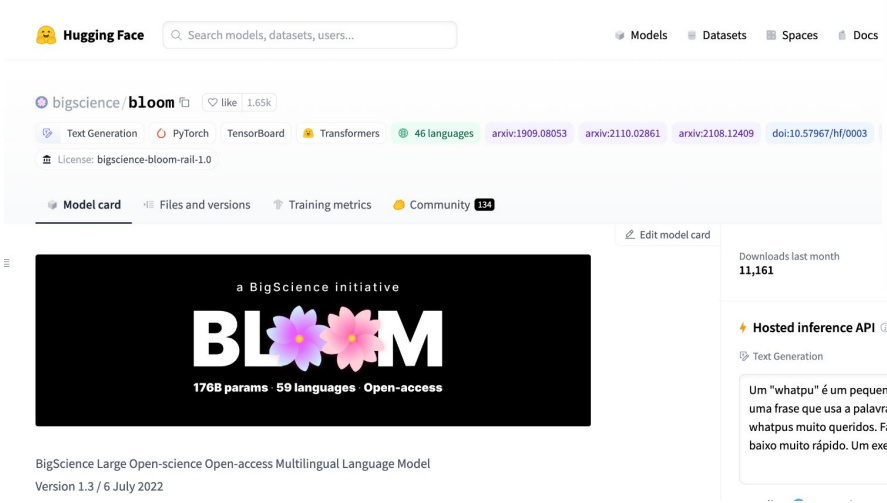


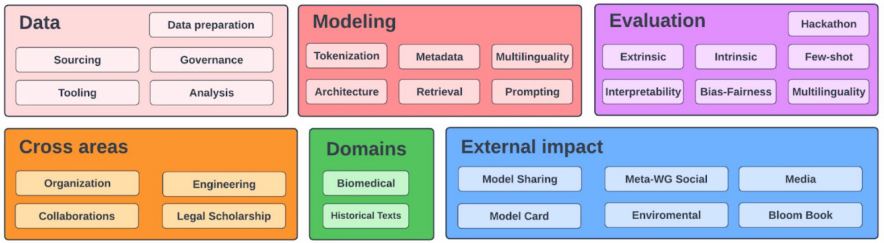
@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

BigScience Workshop

"Open in the way people can understand what you're talking about and can try it themselves…"

BigScience Workshop Working Groups

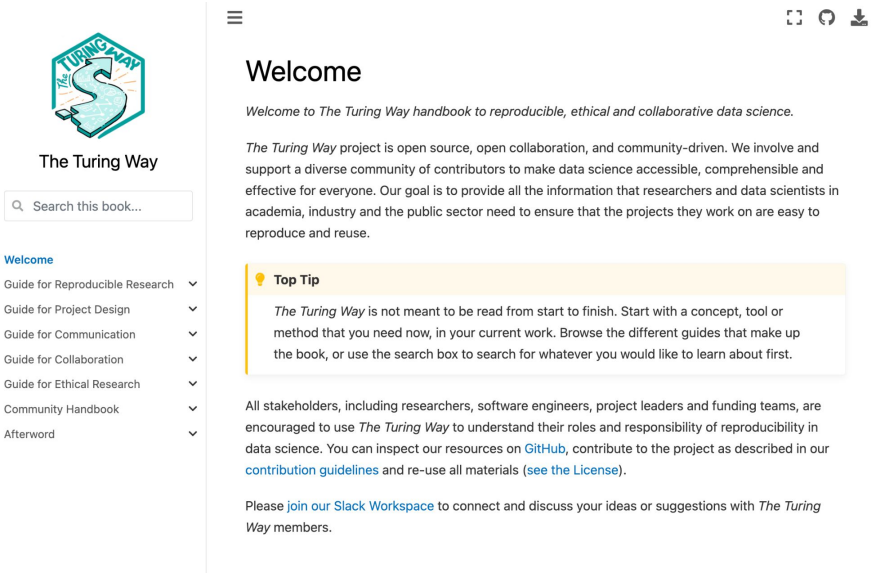


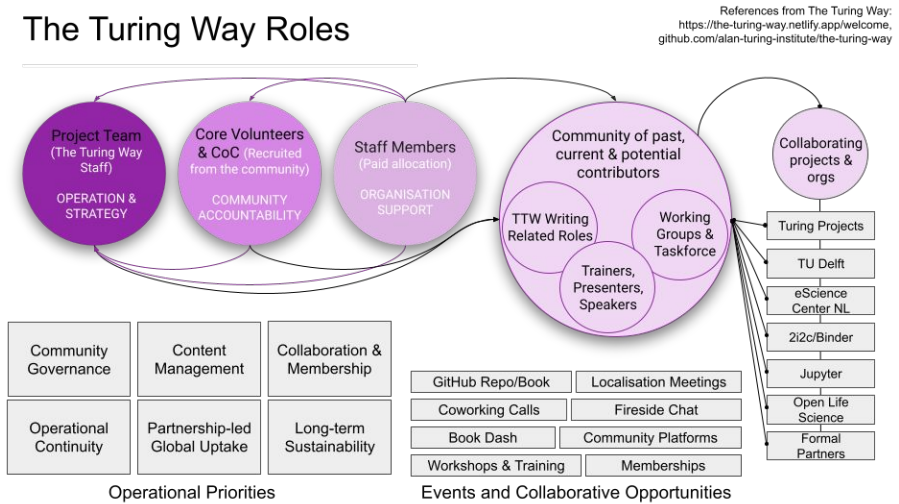


@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

The Turing Way

“Open Science is a buffet…”

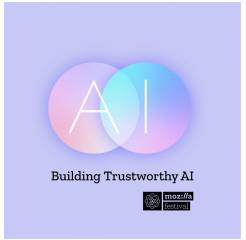




@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

MozFest TAIWG

“Accessible and transparent to all, and with ownership shared by all…”



Building Trustworthy AI Working Group

These projects have been invited to the MozFest TAI Working Group for AI Builders where they will develop tools and technology that promote Trustworthy AI. Working group projects will be showcased at MozFest 2023.

- **Kwanele App Pilot**
- **System to Filter Out Unwanted Content from Incoming Social Media Data**
- **Bountiful Futures**

Trustworthy AI Community Experiences in Mozilla Hubs

These projects have been invited to build virtual worlds that promote Trustworthy AI in Mozilla Hubs. TAI community experiences will be available for exploration at MozFest 2023.

- **AI-musement Park**
- **Public Engagement in AI: An Around-the-World Tour by AI Future Lab**
- **Algorithmic Oppression: Online Representation of Reproductive Rights**
- **A Game Jam on Tackling Misinformation and Disinformation**

@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

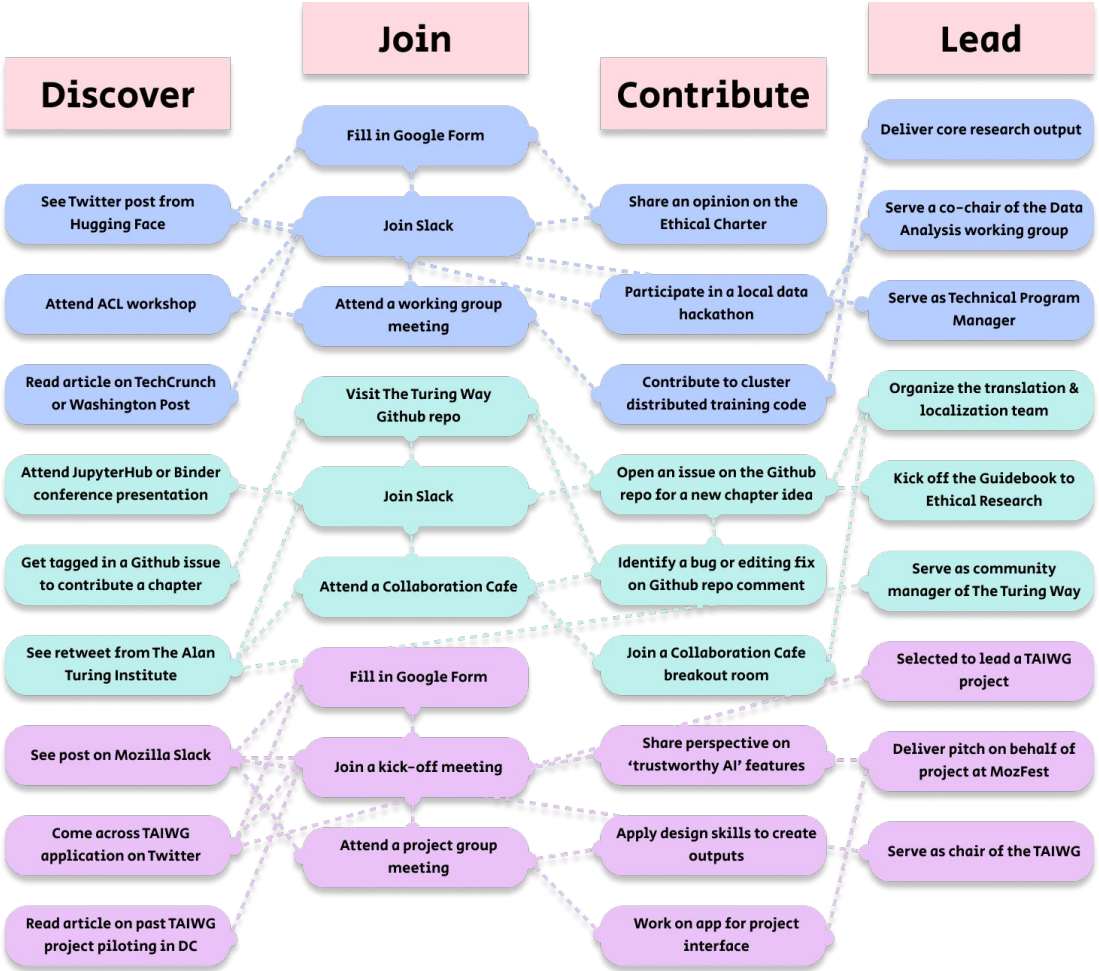
Overview

1. Open Science at The Alan Turing Institute
2. Approaches to Open Source in AI
3. The Open AI Collaborative
4. ***User Journeys through 3 Open AI Collaboratives***

@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

User journeys through open AI communities

BigScience



@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

User journeys through open AI communities

Discover

BigScience

See Twitter post from Hugging Face

Attend ACL workshop

Read article on TechCrunch or Washington Post

The Turing Way

Attend JupyterHub or Binder conference presentation

Get tagged in a Github issue to contribute a chapter

See retweet from The Alan Turing Institute

Building Trustworthy AI

See post on Mozilla Slack

Come across TAIWG application on Twitter

Read article on past TAIWG project piloting in DC

Join

Fill in Google Form

Join Slack

Attend a working group meeting

Visit The Turing Way Github repo

Join Slack

Attend a Collaboration Cafe

Fill in Google Form

Join a kick-off meeting

Attend a project group meeting

Contribute

Share an opinion on the Ethical Charter

Participate in a local data hackathon

Contribute to cluster distributed training code

Open an issue on the Github repo for a new chapter idea

Identify a bug or editing fix on Github repo comment

Join a Collaboration Cafe breakout room

Share perspective on 'trustworthy AI' features

Apply design skills to create outputs

Work on app for project interface

Lead

Deliver core research output

Serve a co-chair of the Data Analysis working group

Serve as Technical Program Manager

Organize the translation & localization team

Kick off the Guidebook to Ethical Research

Serve as community manager of The Turing Way

Selected to lead a TAIWG project

Deliver pitch on behalf of project at MozFest

Serve as chair of the TAIWG

@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

User journeys through open AI communities

BigScience

Join

- Fill in Google Form
- Join Slack
- Attend a working group meeting
- Visit The Turing Way Github repo
- Join Slack
- Attend a Collaboration Cafe
- Fill in Google Form
- Join a kick-off meeting
- Attend a project group meeting

Discover

- See Twitter post from Hugging Face
- Attend ACL workshop
- Read article on TechCrunch or Washington Post
- Attend JupyterHub or Binder conference presentation
- Get tagged in a Github issue to contribute a chapter
- See retweet from The Alan Turing Institute
- See post on Mozilla Slack
- Come across TAIWG application on Twitter
- Read article on past TAIWG project piloting in DC

Contribute

- Share an opinion on the Ethical Charter
- Participate in a local data hackathon
- Contribute to cluster distributed training code
- Open an issue on the Github repo for a new chapter idea
- Identify a bug or editing fix on Github repo comment
- Join a Collaboration Cafe breakout room
- Share perspective on 'trustworthy AI' features
- Apply design skills to create outputs
- Work on app for project interface

Lead

- Deliver core research output
- Serve a co-chair of the Data Analysis working group
- Serve as Technical Program Manager
- Organize the translation & localization team
- Kick off the Guidebook to Ethical Research
- Serve as community manager of The Turing Way
- Selected to lead a TAIWG project
- Deliver pitch on behalf of project at MozFest
- Serve as chair of the TAIWG

@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

User journeys through open AI communities

BigScience

Discover	Join	Contribute	Lead
See Twitter post from Hugging Face	Fill in Google Form	Share an opinion on the Ethical Charter	Deliver core research output
Attend ACL workshop	Join Slack	Participate in a local data hackathon	Serve a co-chair of the Data Analysis working group
Read article on TechCrunch or Washington Post	Attend a working group meeting	Contribute to cluster distributed training code	Serve as Technical Program Manager
Attend JupyterHub or Binder conference presentation	Visit The Turing Way Github repo	Open an issue on the Github repo for a new chapter idea	Organize the translation & localization team
Get tagged in a Github issue to contribute a chapter	Join Slack	Identify a bug or editing fix on Github repo comment	Kick off the Guidebook to Ethical Research
See retweet from The Alan Turing Institute	Attend a Collaboration Cafe	Join a Collaboration Cafe breakout room	Serve as community manager of The Turing Way
See post on Mozilla Slack	Fill in Google Form	Share perspective on ‘trustworthy AI’ features	Selected to lead a TAIWG project
Come across TAIWG application on Twitter	Join a kick-off meeting	Apply design skills to create outputs	Deliver pitch on behalf of project at MozFest
Read article on past TAIWG project piloting in DC	Attend a project group meeting	Work on app for project interface	Serve as chair of the TAIWG

@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

User journeys through open AI communities

BigScience

Discover	Join	Contribute	Lead
See Twitter post from Hugging Face	Fill in Google Form	Share an opinion on the Ethical Charter	Deliver core research output
Attend ACL workshop	Join Slack	Participate in a local data hackathon	Serve a co-chair of the Data Analysis working group
Read article on TechCrunch or Washington Post	Attend a working group meeting	Contribute to cluster distributed training code	Serve as Technical Program Manager
Attend JupyterHub or Binder conference presentation	Visit The Turing Way Github repo	Open an issue on the Github repo for a new chapter idea	Organize the translation & localization team
Get tagged in a Github issue to contribute a chapter	Join Slack	Identify a bug or editing fix on Github repo comment	Kick off the Guidebook to Ethical Research
See retweet from The Alan Turing Institute	Attend a Collaboration Cafe	Join a Collaboration Cafe breakout room	Serve as community manager of The Turing Way
See post on Mozilla Slack	Fill in Google Form	Share perspective on 'trustworthy AI' features	Selected to lead a TAIWG project
Come across TAIWG application on Twitter	Join a kick-off meeting	Apply design skills to create outputs	Deliver pitch on behalf of project at MozFest
Read article on past TAIWG project piloting in DC	Attend a project group meeting	Work on app for project interface	Serve as chair of the TAIWG

@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

Towards Openness Beyond Open Access

Open…

- for diverse, global participation in AI
- to empower those outside the tech world
- create new forms of AI and AI production

@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884

The Alan Turing Institute

Thank you!

Read more:

- Open AI Philosophy: https://www.turing.ac.uk/blog/what-defines-open-open-ai
- Neurips Workshop Paper: https://openreview.net/pdf?id=slU-5h8rrCz
- ML Licensing in The Turing Way: https://the-turing-way.netlify.app/reproducible-research/licensing/licensing-ml.html

Continue the conversation:

- Email jding@turing.ac.uk

@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.7308884