

The Alan Turing Institute

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

Jennifer Ding | The Alan Turing Institute

jding@turing.ac.uk

What is “Open AI”?



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

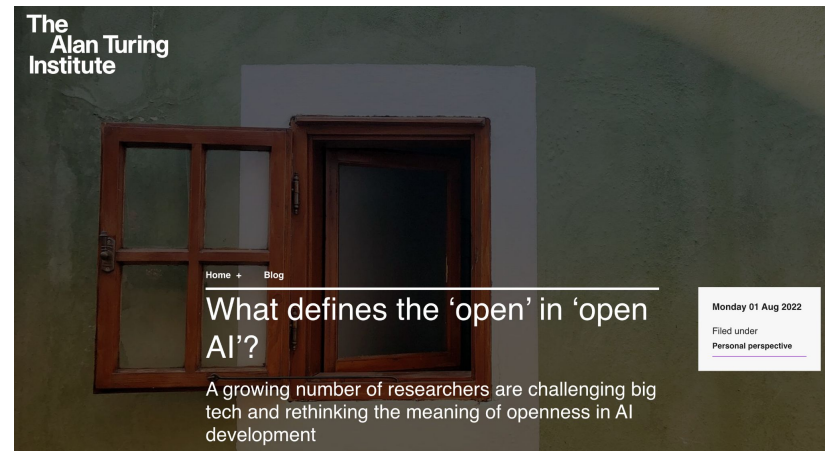
The Alan Turing Institute
asteele@turing.ac.uk

Temi Popo

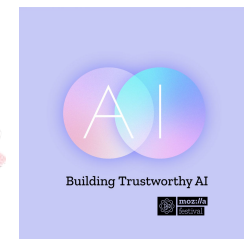
Mozilla Foundation
temi@mozillafoundation.org

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 additionalality 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.



BigScience



Overview

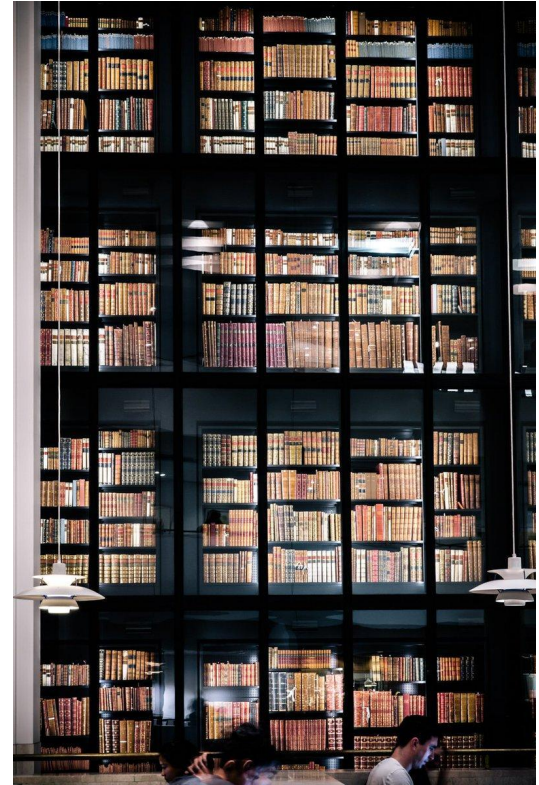
1. Open Science at The Alan Turing Institute
2. Approaches to Open Source in AI
3. The Open AI Collaborative
4. User Journeys through 3 Open AI Collaboratives

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

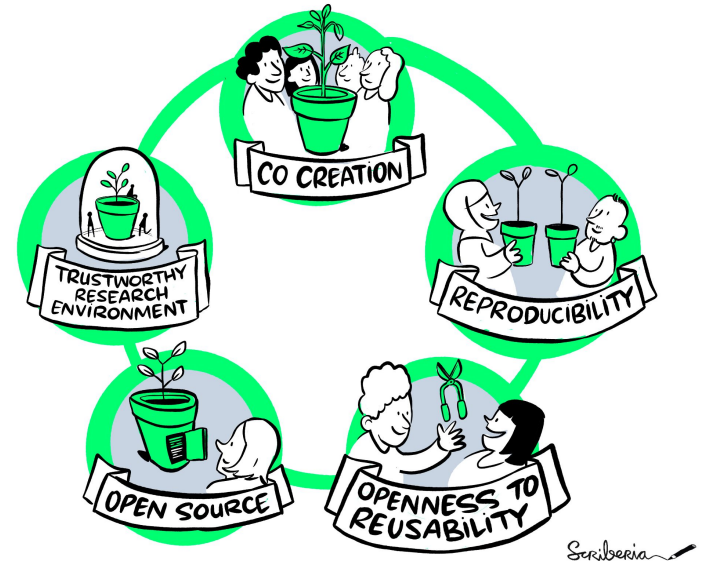
The Alan Turing Institute

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



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**



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

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.

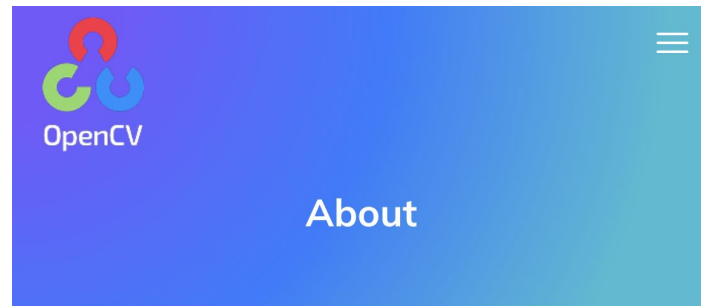
The project was started in 2007 by David Cournapeau as a Google Summer of Code project, and since then many volunteers have contributed. See the [About us](#) page for a list of core contributors.

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.

Open ML Licensing



The Turing Way

Search this book...

- Welcome
- [Guide for Reproducible Research](#)
- Overview
- Open Research
- Version Control
- [Licensing](#)
- Software Licenses
- Data Licenses
- [Machine Learning Model Licenses](#)
- Case Studies: Choosing an ML License
- License Compatibility
- Checklist
- Research Data Management
- Reproducible Environments
- BinderHub
- Code quality
- Code Testing
- Code Reviewing Process
- Reusable Code
- Continuous Integration (CI)
- Reproducible Research with Make
- Research Compendia
- Risk Assessment

- 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

Towards Openness Beyond Open Access

Open...

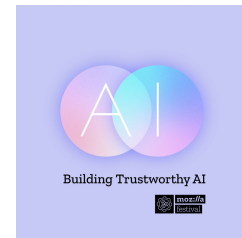
- as a way of collaboration
- for setting culture & norms
- beyond “available on Github”
- expanded by “responsible”?

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

What is an Open AI collaborative?

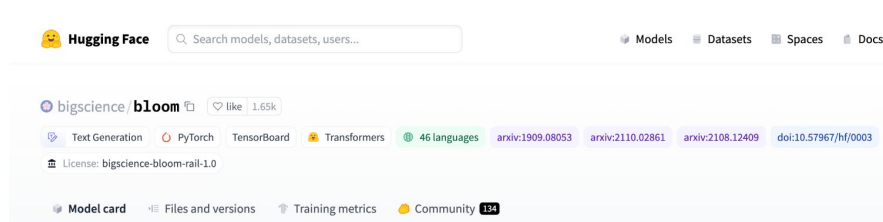
- Distributed, volunteer-led teams creating alternative AI development pathways grounded in “open” principles
- Focus on activities relevant for their respective communities:
 - building models (BigScience Workshop)
 - tools/ways of working (The Turing Way)
 - ecosystems (MozFest TAIWG)



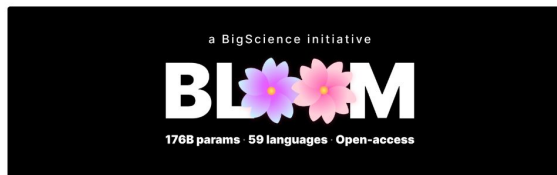
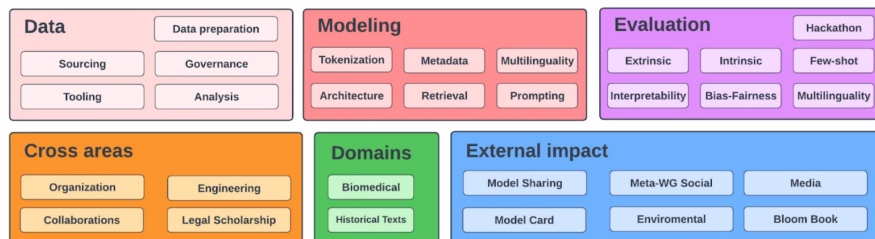
BigScience Workshop

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

BigScience Workshop Working Groups



The screenshot shows the Hugging Face interface for the 'bloom' model. At the top, there is a search bar and navigation tabs for Models, Datasets, Spaces, and Docs. The model name 'bloom' is highlighted, with a like count of 1.65k. Below this, there are tags for 'Text Generation', 'PyTorch', 'TensorBoard', 'Transformers', '46 languages', and several arXiv and DOI links. A license 'bigsience-bloom-rail-1.0' is listed. At the bottom of the header, there are tabs for 'Model card', 'Files and versions', 'Training metrics', and 'Community' (with 134 members).



BigScience Large Open-science Open-access Multilingual Language Model
Version 1.3 / 6 July 2022

Downloads last month
11,161

Hosted inference API

Text Generation

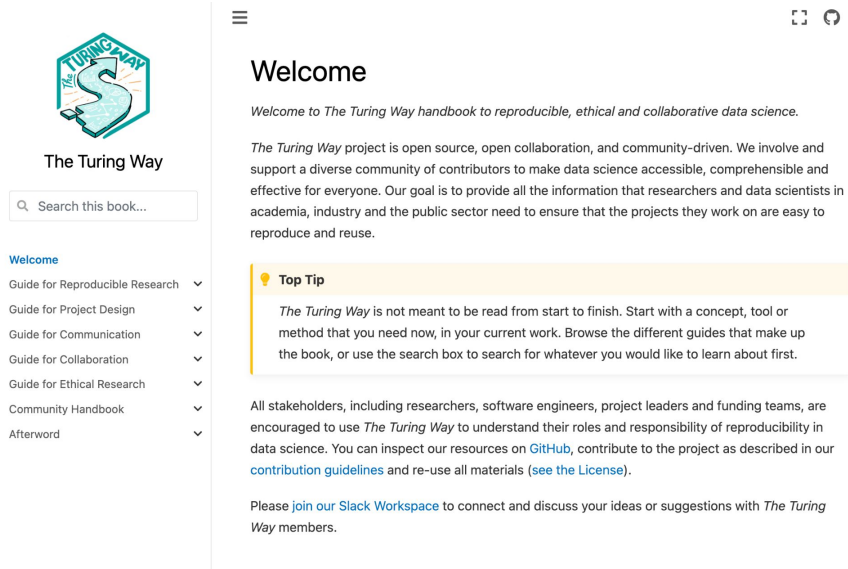
Groups

Examples

Um "whatpu" é um pequeno animal peludo nativo da Tanzânia. Um exemplo de uma frase que usa a palavra whatpu é: Estávamos a viajar por África e vimos uns whatpus muito queridos. Fazer um "farduddle" significa saltar para cima e para baixo muito rápido. Um exemplo de uma frase que usa a palavra farduddle é:

The Turing Way

“Open Science is a buffet...”



The screenshot shows the 'Welcome' page of the Turing Way handbook. It features the Turing Way logo, a search bar, and a list of navigation links. The main content includes a welcome message, a top tip, and information about the project's goals and stakeholder involvement.

Welcome

Welcome to *The Turing Way* handbook to reproducible, ethical and collaborative data science.

The Turing Way project is open source, open collaboration, and community-driven. We involve and support a diverse community of contributors to make data science accessible, comprehensible and effective for everyone. Our goal is to provide all the information that researchers and data scientists in academia, industry and the public sector need to ensure that the projects they work on are easy to reproduce and reuse.

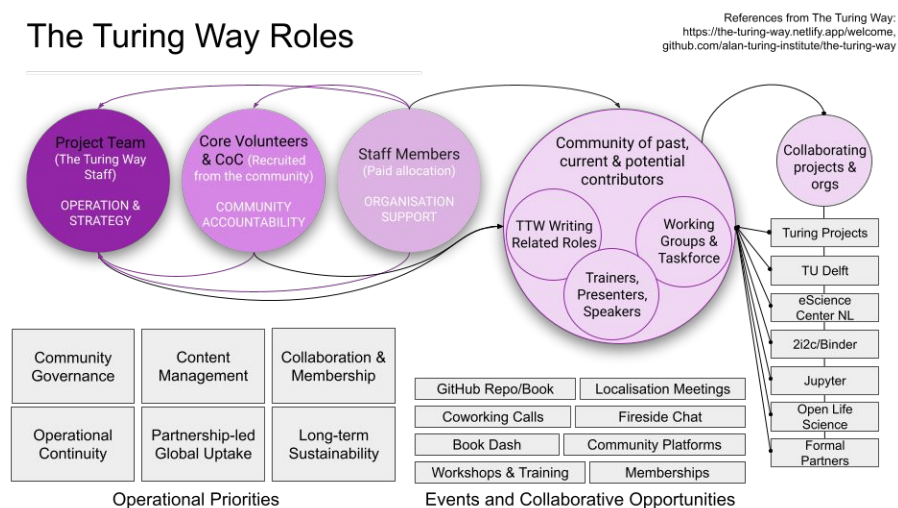
Top Tip

The Turing Way is not meant to be read from start to finish. Start with a concept, tool or method that you need now, in your current work. Browse the different guides that make up the book, or use the search box to search for whatever you would like to learn about first.

All stakeholders, including researchers, software engineers, project leaders and funding teams, are encouraged to use *The Turing Way* to understand their roles and responsibility of reproducibility in data science. You can inspect our resources on [GitHub](#), contribute to the project as described in our [contribution guidelines](#) and re-use all materials ([see the License](#)).

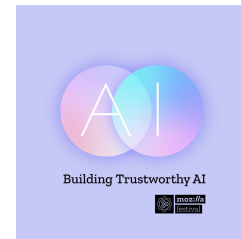
Please [join our Slack Workspace](#) to connect and discuss your ideas or suggestions with *The Turing Way* members.

The Turing Way Roles



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

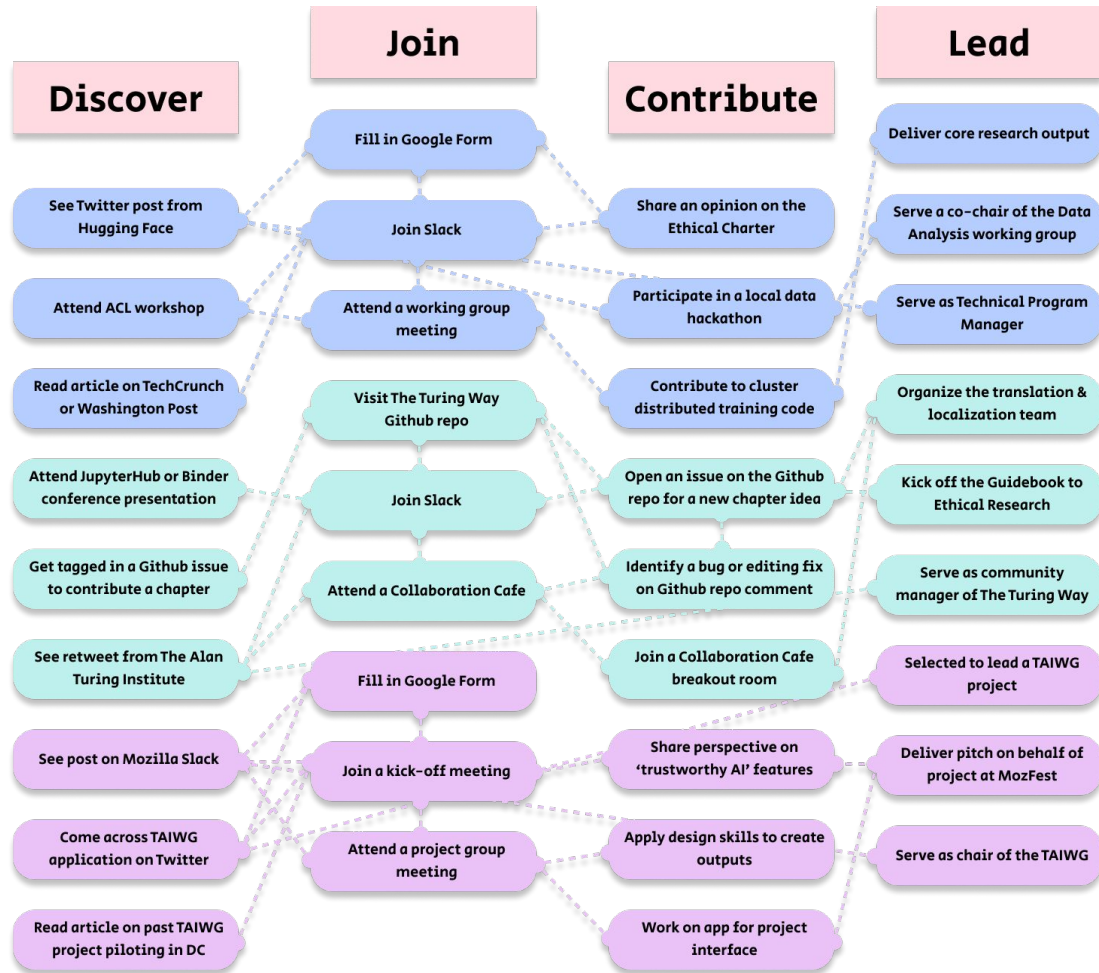
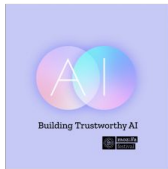
These projects have been invited to build virtual worlds that promote Trustworthy AI in Mozilla Hubs. TAI community experiences will be available for exploration at [MozFest 2023](#).

- **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**

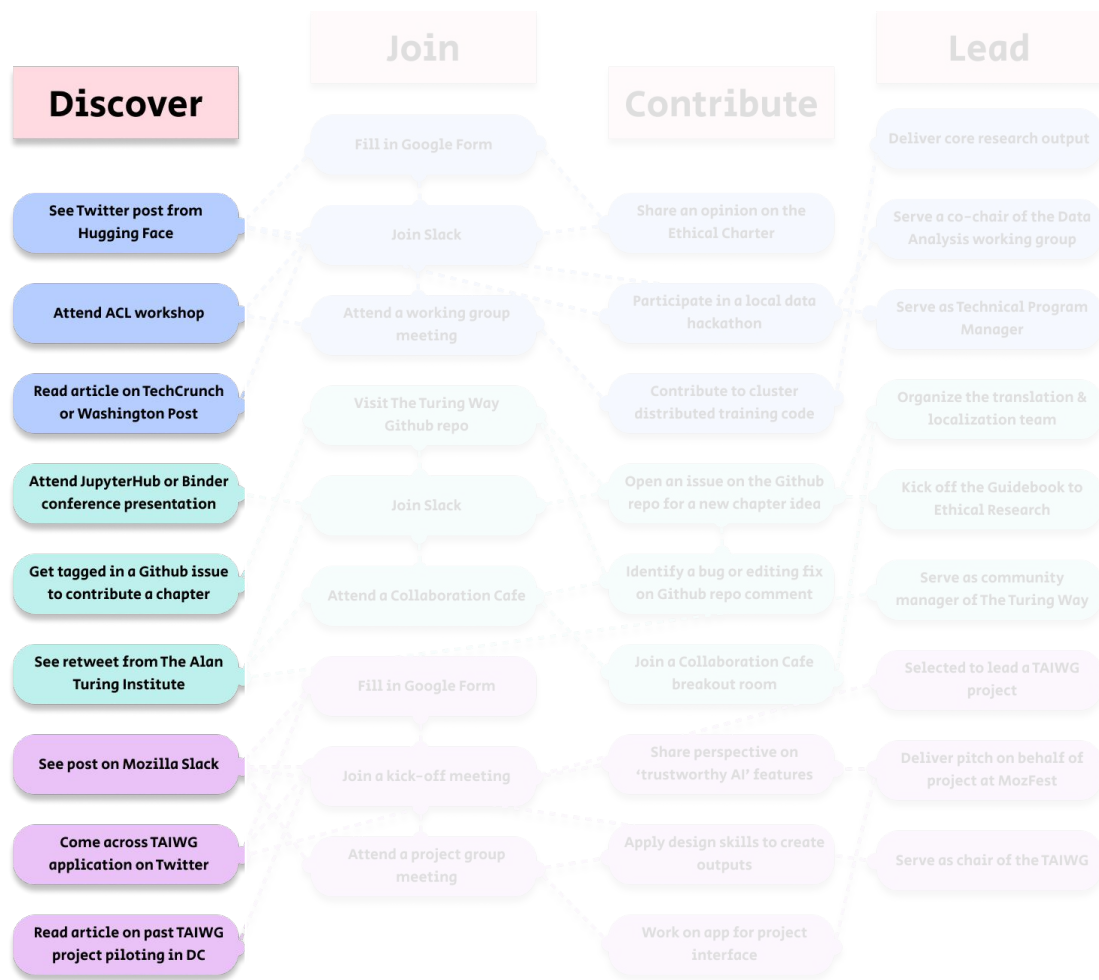
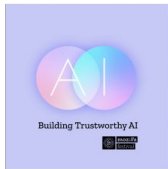
Overview

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***

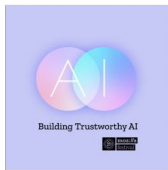
User journeys through open AI communities



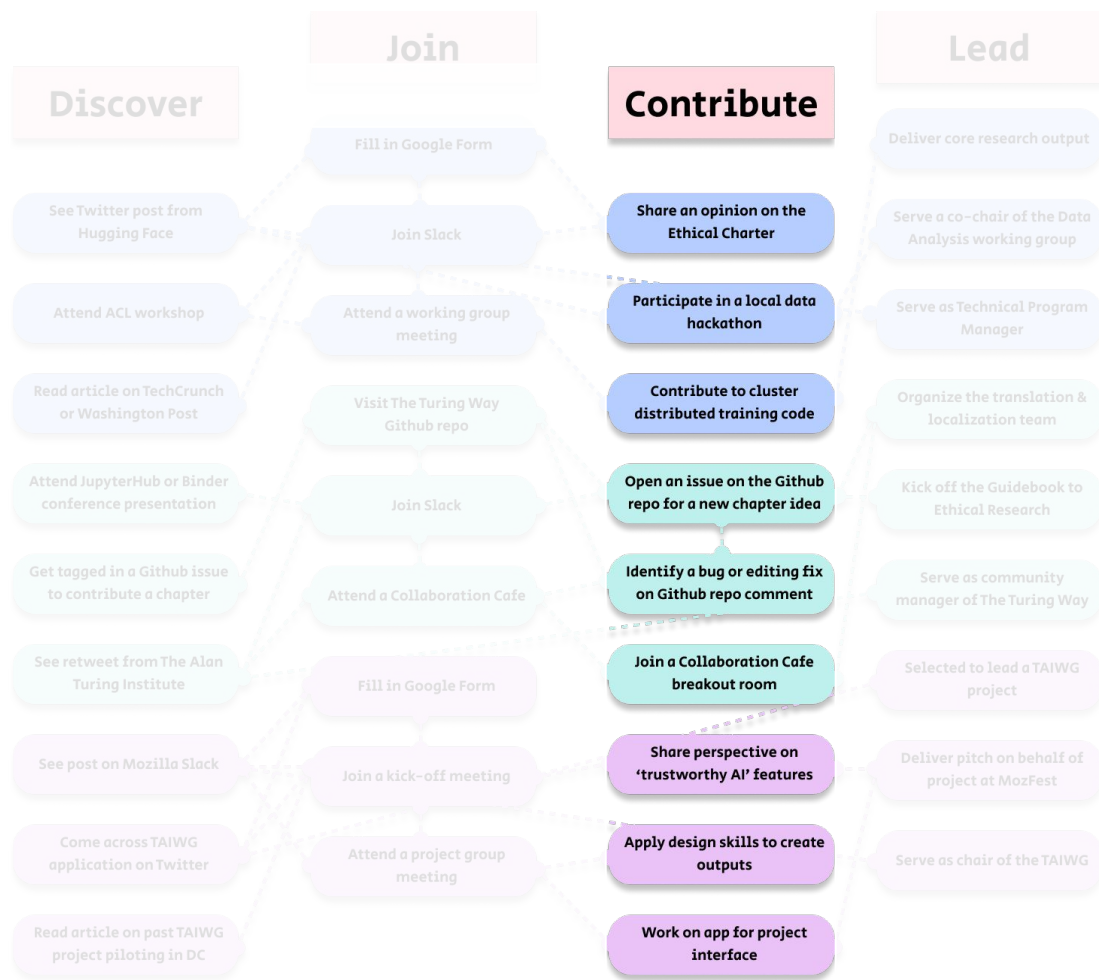
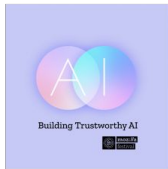
User journeys through open AI communities



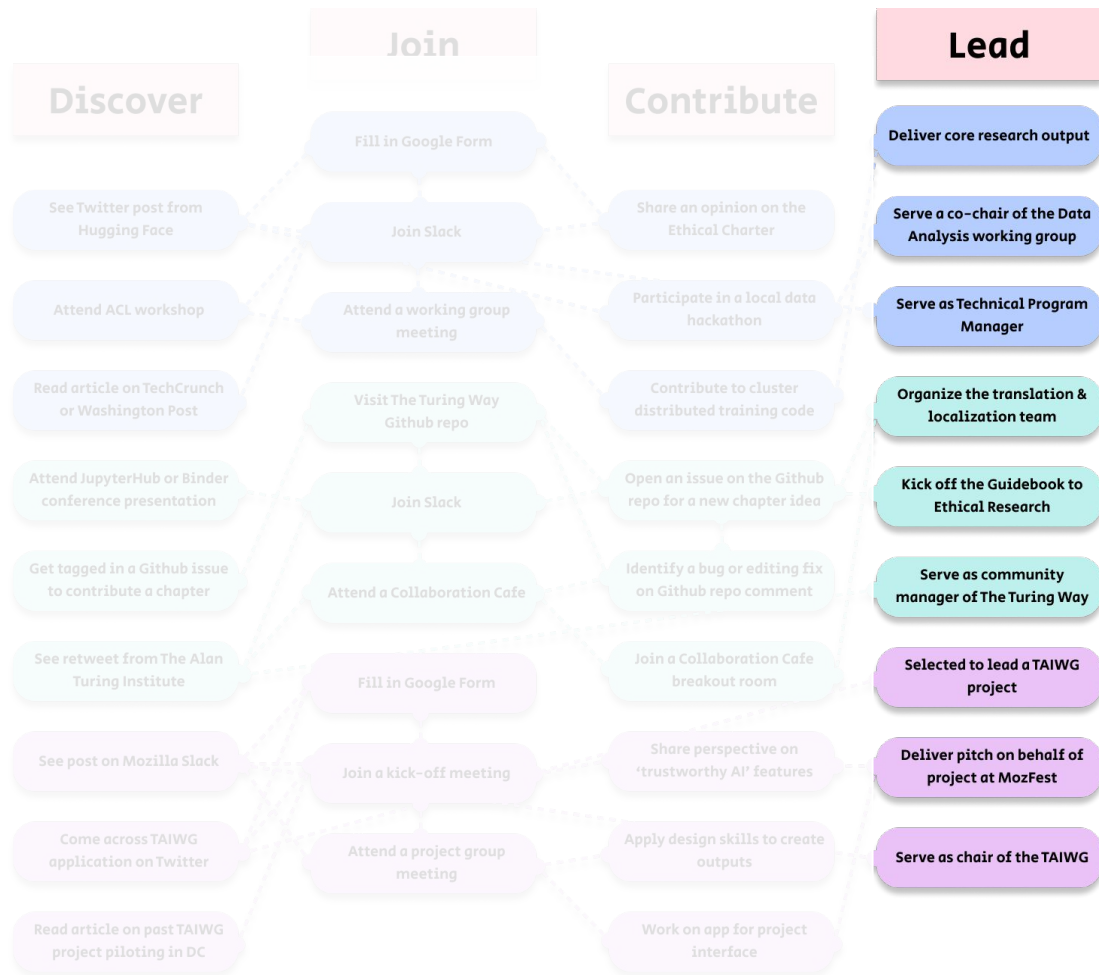
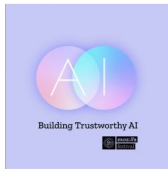
User journeys through open AI communities



User journeys through open AI communities



User journeys through open AI communities



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

The Alan Turing Institute

Thank you for tuning in!

Read more:

- Open AI Philosophy: <https://www.turing.ac.uk/blog/what-defines-open-open-ai>
- Neurips Workshop Paper: <https://openreview.net/pdf?id=sIU-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

