The Alan Turing Institute

Safe, ethical, and open research in the age of Al

Kirstie Whitaker Pronouns: she/her



configured ETs plants in collection series points of increases or control than explain control mannering series of the constant, making series of the control of the processing points receiving where the processing points receiving where the process are control or processing points or complete process

Research Programme Director for Tools, **Practices & Systems** The Alan **Turing Institute**



Research Programme Director for Tools, Practices & Systems The Alan **Turing Institute**





The Turing Way is an open source book project that involves and supports a diverse research community in ensuring that reproducible and ethical data science is accessible and comprehensible for everyone.

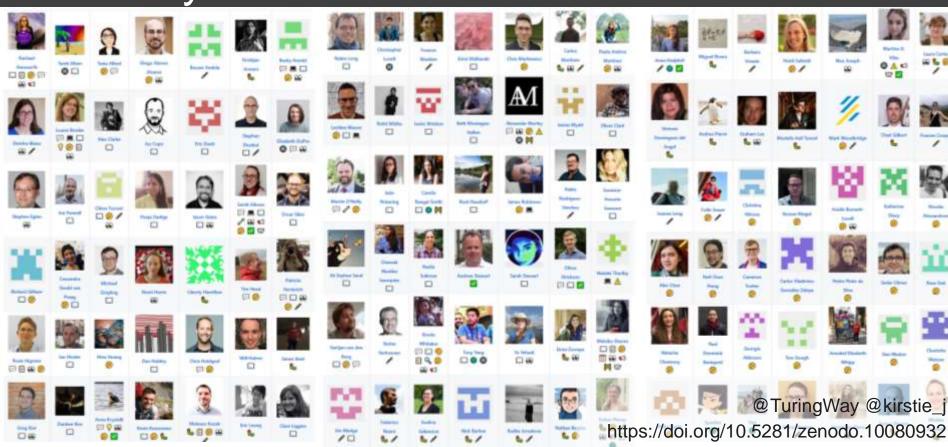


The Turing Way is:

- -a book
- –a community
- a globalcollaboration
- a whole tonne of work



Thank you to our 400+ contributors



Malvika Sharan

"No one can change research culture on their own. Scaling our community by empowering YOU to participate is how we will change the world."





https://the-turing-way.netlify.app/index.html https://github.com/the-turing-way/the-turing-way



Personal experience of:

- File drawer effect
- Lack of reproducibility
- Imposter syndrome around coding
- Lonely
- Wasted time
- Hypocrisy of academia



Personal experience of:

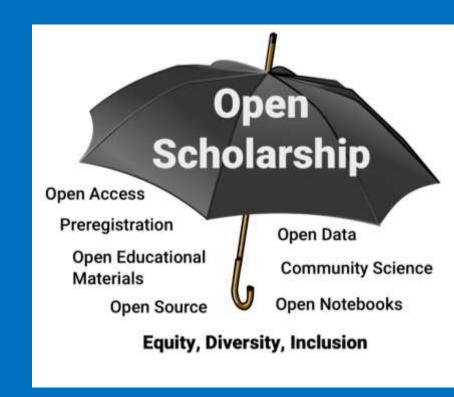
- File drawer effect
- Lack of reproducibility
- Imposter syndrome around coding
- Lonely
- Wasted time
- Hypocrisy of academia

#bropenscience is broken science

All elements of open scholarship should be in service of equity, diversity and inclusion.

Faster, more accessible

- → Better quality insights
- → A more just world



Whitaker & Guest, 2020.

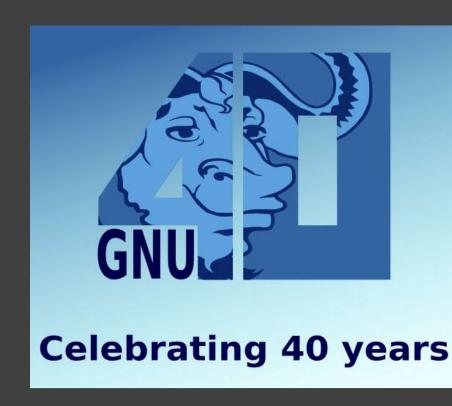
https://www.bps.org.uk/psychologist/bropenscience-broken-science https://hdl.handle.net/2066/226397

Newspeak

"propagandistic language marked by euphemism, circumlocution, and the inversion of customary meanings"

Free software: "users have the freedom to run, copy, distribute, study, change and improve the software"

https://www.merriam-webster.com/dictionary/newspeak https://www.gnu.org/philosophy/free-sw.en.html https://opensource.org/osd



Newspeak

"propagandistic language marked by euphemism, circumlocution, and the inversion of customary meanings"

Open source: "Open source doesn't just mean access to the source code"



https://www.merriam-webster.com/dictionary/newspeak https://www.gnu.org/philosophy/free-sw.en.html https://opensource.org/osd

LLaMA2 is not open

"Meta using the term 'open source' for this is positively misleading: There is no source to be seen, the training data is entirely undocumented, and beyond the glossy charts the technical documentation is really rather poor." - Mark Dingemanse



https://spectrum.ieee.org/open-source-llm-not-open Liesenfeld, Lopez & Dingemanse, 2023; https://doi.org/10.1145/3571884.3604316



https://opening-up-chatgpt.github.io Liesenfeld, Lopez & Dingemanse, 2023

https://doi.org/10.5281/zenodo.10080932

"Publicly" available data is not open

Creators are being harmed through this technology:

- Authors Guild are suing OpenAl.
- Double strike over fair compensation: SAG-AFTRA and Writers Guild of America.

https://authorsguild.org/news/ag-and-authors-file-class-action-suit-against-openai

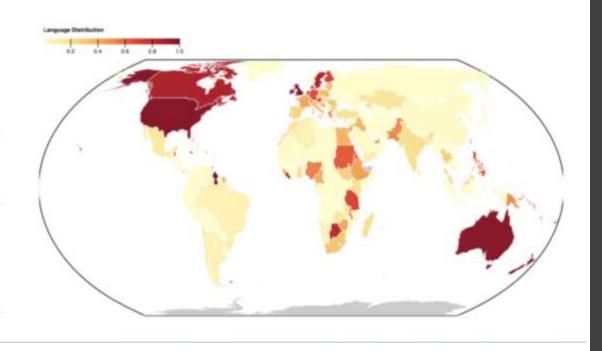
https://www.vox.com/culture/2023/7/13/23793828/sag-aftra-strike-wga-hollywood



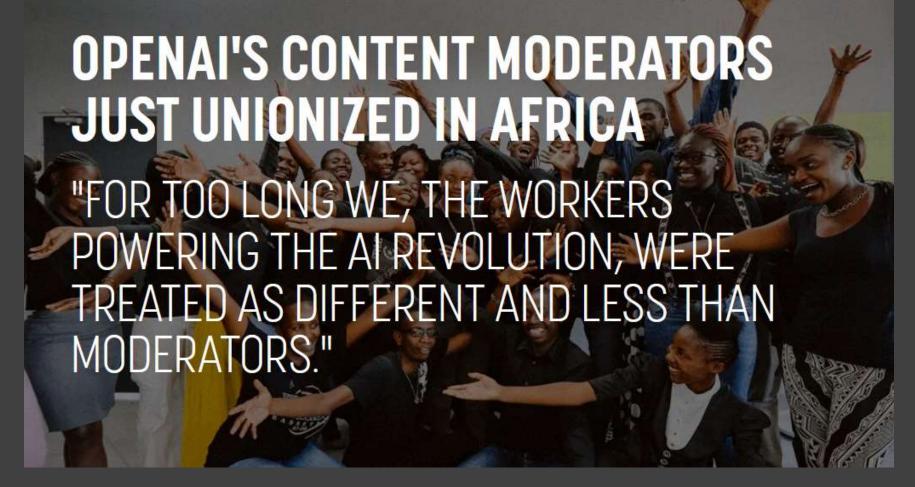
Data Provenance for Al

A massive audit of 1800+ Al text datasets, tracing their provenance and composition from origin to creation.

The race to train language models on vast, diverse, and inconsistently documented datasets has raised pressing concerns about the legal and ethical risks for practitioners, and the models they train. To remedy these dubious practices in data transparency and understanding, we convene a multi-disciplinary effort between legal and machine learning experts to systematically audit and trace



"Frequent miscategorization of licenses on widely used dataset hosting sites, with license omission of 68%+ and error rates of 50%+."



Who is building open source AI?

21 September, 14:00 - 15:30 UTC+1 Register on Eventbrite



Arielle Bennett
Programme Manager
The Alan Turing Institute



Mophat Okinyi
Union Representative
African Content Moderators Union



Marzieh Fadaee Senior Research Scientist Cohere for Al



Abinaya Mahendiran cro Nunnari Labs



David Gray Widder

Postdoctoral Fellow

Cornell Tech



Jennifer Ding Senior Researcher The Alan Turing Institute





Open Source Initiative Deep Dive: Defining Open Source Al





Prof David Leslie
Director of Ethics and
Responsible Research
and Innovation



Victoria Kwan
Corporate Governance
Research Ethics
Manager



Dr Kirstie Whitaker
Director of Tools,
Practices and Systems
Research Programme

Operationalising the SAFE-D principles for Open Source Al

https://opensource.org/deepdive/webinars/operationalising-the-safe-d-principles-for-open-source-ai



- Open Science and society
- Research integrity and publishing ethics
- Al and scholarly communication
- Open repositories
- New horizons in metrics and research assessment



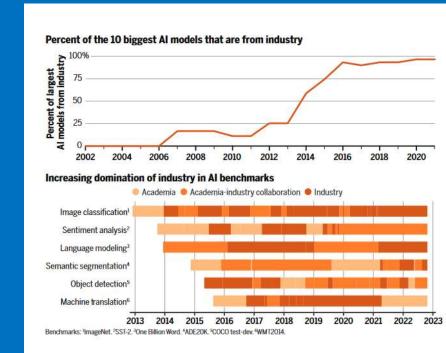


- Open Science and society
- Research integrity and publishing ethics
- Al and scholarly communication
- Open repositories
- New horizons in metrics and research assessment



Industry is dominating the Al conversation

- 70% of people with a PhD in Al are employed in industry
- Costs prohibitively high for academic researchers
- Very little incentive to work on safe & equitable AI for the public good



Ahmed, Wahed & Thompson, 2023 https://doi.org/10.1126/science.ade2420





@TuringWay @kirstie_j



Our UK university network



University College London University of Aberdeen University of Bath University of Birmingham University of Bradford University of Bristol University of Cambridge University of Dundee University of East Anglia University of Edinburgh University of Essex University of Exeter University of Glasgow University of Greenwich University of Hull University of Leeds University of Lincoln University of Liverpool University of Manchester University of Nottingham University of Oxford University of Plymouth University of Portsmouth University of Reading University of Salford University of Sheffield University of Southampton University of Strathclyde University of Surrey University of Sussex University of Warwick University of Westminster

University of York@TuringWay @kirstie_j

https://doi.org/10.5281/zenodo.10080932

https://www.turing.ac.uk/turing-university-network

Our international network



Our partners and collaborators

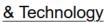




Home > Business and industry > Science and innovation > Artificial intelligence > AI Safety Summit 2023: The Bletchley Declaration



<u>Department for</u> Science, Innovation





Foreign, Commonwealth





Prime Minister's Office, 10

Downing Street

Policy paper

The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023

Published 1 November 2023

https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration

Business and industry > Science and innovation > Artificial intelligence > AI Safety Summit 2023: The Bletchley Declaration

Artificial Intelligence (AI) presents enormous global

Depart opportunities: it has the potential to transform and enhance

& Tech human wellbeing, peace and prosperity. To realise this, we affirm that, for the good of all, Al should be designed, developed, deployed, and used, in a manner that is safe, in

Posuch a way as to be human-centric, trustworthy and responsible. We welcome the international community's efforts so far to cooperate on AI to promote inclusive economic growth, sustainable development and innovation,

to protect human rights and fundamental freedoms, and to Publifoster public trust and confidence in Al systems to fully

realise their potential.

https://www.gov.uk/government/publications/ ai-safety-summit-2023-the-bletchley-declaration @TuringWay @kirstie i

Business and industry > Science and innovation > Artificial intelligence > AI Safety Summit 2023: The Bletchley Declaration



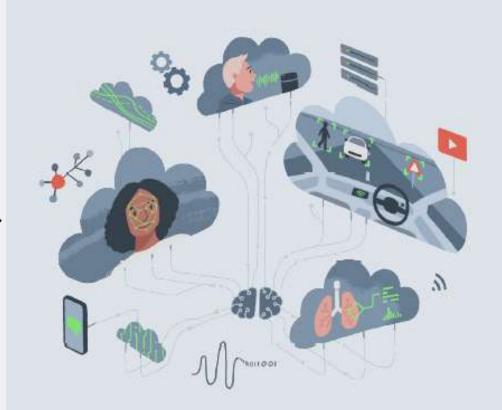
Depart Scienc In the context of our cooperation, and to inform action at the <u>8 Tech</u> national and international levels, our agenda for addressing

frontier AI risk will focus on:

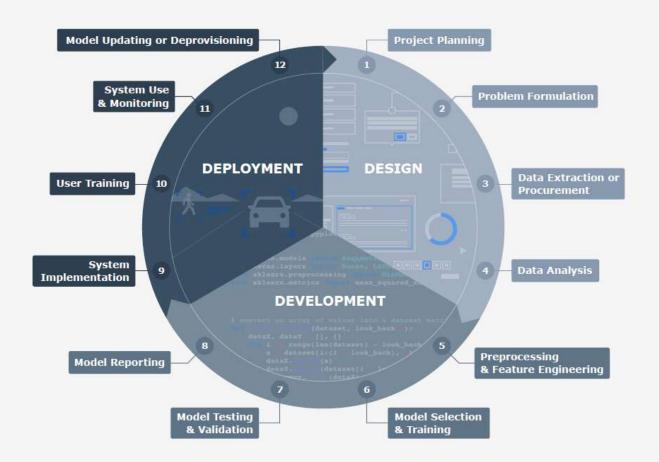
identifying AI safety risks of shared concern, building a shared scientific and evidence-based understanding of these risks, and sustaining that understanding as capabilities continue to increase, in the context of a wider global approach to understanding the impact of Al in our societies.

https://www.gov.uk/government/publications/ ai-safety-summit-2023-the-bletchley-declaration

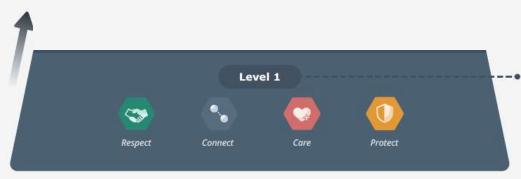
"Al system" means any computational system (or a combination of such systems) that uses methods derived from statistics or other mathematical techniques to carry out tasks that are commonly associated with, or would otherwise require, human intelligence and that either assists or replaces the judgment of human decision-makers in carrying out those tasks.



https://www.turing.ac.uk/research/research-projects/ai-ethics-and-governance-practice



https://www.turing.ac.uk/research/research-projects/ai-ethics-and-governance-practice



The Sum Values

These values support, underwrite, and motivate responsible Al projects and provide criteria for assessing their potential social and ethical impacts.

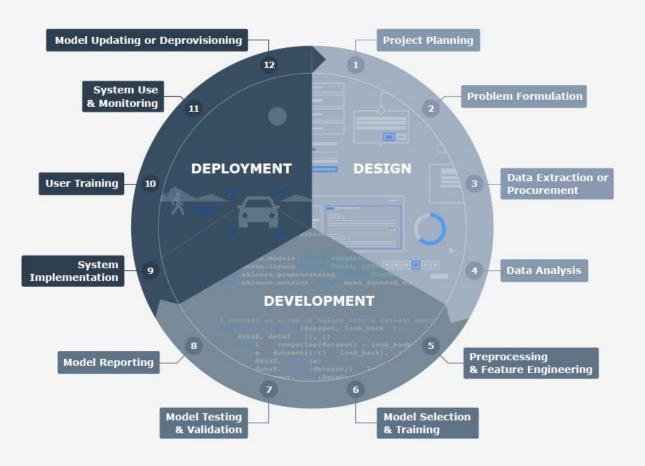
https://www.turing.ac.uk/research/research-projects/ai-ethics-and-governance-practice



https://www.turing.ac.uk/research/research-projects/ai-ethics-and-governance-practice



https://www.turing.ac.uk/research/research-projects/ai-ethics-and-governance-practice



- Open Science and society
- Research integrity and publishing ethics
- Al and scholarly communication
- Open repositories
- New horizons in metrics and research assessment

https://www.turing.ac.uk/research/research-projects/ai-ethics-and-governance-practice



- Safety and sustainability
- Accountability
- Fairness
- Explainability
- Data stewardship

(*We just have a little more work to do to promote and normalise them!)

- Safety and sustainability
- Accountability
- Fairness
- Explainability
- Data stewardship

- Safety and sustainability
- Accountability
- Fairness
- Explainability
- Data stewardship



- Safety and sustainability
- Accountability
- Fairness
- Explainability
- Data stewardship

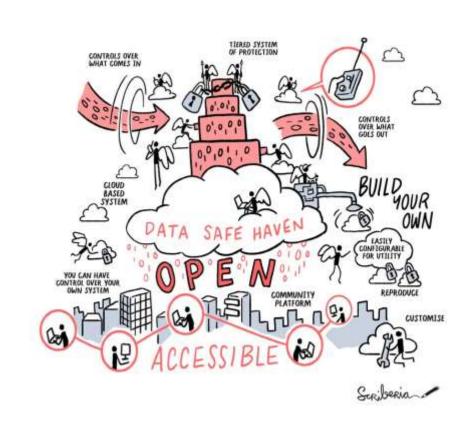
Having a broad and engaged community involved throughout an Al development workflow keeps infrastructure more secure and keeps the purpose of the work aligned with the needs of the impacted stakeholders.

Our 'North Star'

To remove barriers to working safely and effectively with sensitive data,

by promoting and demonstrating a culture of open, community-led development

of interoperable foundational infrastructure and governance.



https://github.com/alan-turing-institute/data-safe-haven https://data-safe-haven.readthedocs.io/en/latest

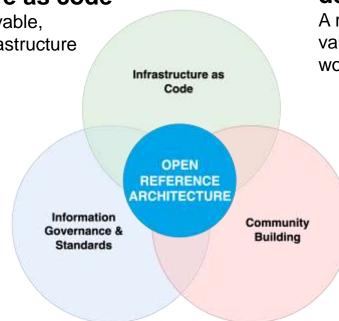
Project pillars

Infrastructure as code

Building a deployable, reproducible infrastructure

Information governance & standards

An accompanying IG framework that is standardised, auditable, accessible and practical



A Turing production deployment

A real-world application and validation of the research workstreams

TRESA - a Turing production deployment

Community building

An open-first community that empowers contributions from all stakeholders

https://github.com/alan-turing-institute/data-safe-haven https://data-safe-haven.readthedocs.io/en/latest

- Safety and sustainability
- Accountability
- Fairness
- Explainability
- Data stewardship

anach seed water revaled dynamic anguerates a v2.6.0-bets.2 build; release 2.6.0-bets 2 build: build 2.6.0-beta.2 feat: dynamic directive arguments for v-on, v-bind and custom origin/dynamic-directive-arguments | feat: dynamic args for customers perf: improve scoped slots change detection accuracy (#9371) packages "Anyone could say anything was a function of the state of the D STC test official. And to emphasize types the point, labelled the notes estinucis' Request for Comments."" fix: fix checkbox event edge case in Figure 1 .gitignore test: fix tests in IE/Edge BACKERS:md Stephen Crocker, Request For Comments (RFC's) History

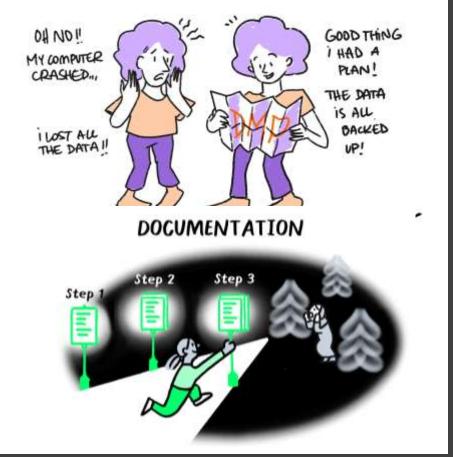
https://www.livinginternet.com/i/ia_rfc_invent.htm Chickle, Upstable Still @TuringWay @kirstie i

https://doi.org/10.5281/zenodo.10080932

- Safety and sustainability
- Accountability
- Fairness
- Explainability
- Data stewardship

Providing documentation covering how decisions are made, how changes can be suggested through requests for comment, and how concerns can be raised for example through codes of conduct – facilitates accountability through the AI development workflow.

TRACK PROJECT HISTORY draft doc. draftdoe -final decfinal-final Seriberia 6



https://the-turing-way.netlify.app/project-design/data-governance/data-gov-ml https://the-turing-way.netlify.app/reproducible-research/vcs https://the-turing-way.netlify.app/project-design/project-design https://

- Safety and sustainability
- Accountability
- Fairness
- Explainability
- Data stewardship

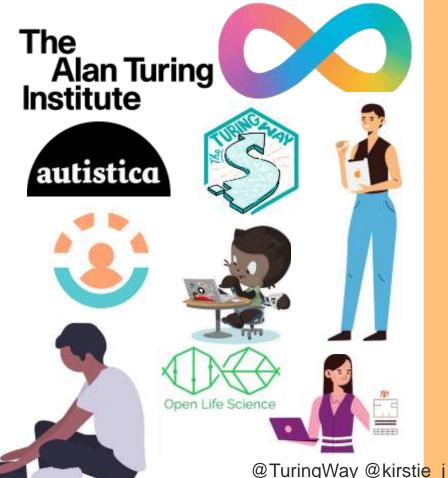


About AutSPACEs

A **research collaboration** between The Turing, Autistica, & Open Humans

A **community** of autistic people, their supporters, researchers and open source developers

A **citizen science** platform where people can share their experiences of sensory processing.



https://doi.org/10.5281/zenodo.10080932

https://doi.org/10.5281/zenodo.8164375



Collect qualitative data to improve our understanding of sensory processing in people's daily lives

Share people's
stories and
adaptive
techniques with
others who have
similar experiences

Educate
neurotypical
people to better
support their
friends, family and
colleagues

Advise
organisations on
how they can
design and adapt
spaces to improve
people's lives

@TuringWay @kirstie_j

Moderation Goals







Susanna Fantoni

A welcoming, inclusive, and safe online space for autistic people

A space where autistic people can share challenging experiences

Decisions
which balance
the complex
needs of the
community

- Safety and sustainability
- Accountability
- Fairness
- Explainability
- Data stewardship

Open source AI should clearly communicate how the AI model and workflow are considering equity and justice. Scholarly communications experts can connect existing tools for bias reporting into an interoperable open source AI ecosystem.

- Safety and sustainability
- Accountability
- Fairness
- Explainability
- Data stewardship

"The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures."

- Safety and sustainability
- Accountability
- Fairness
- Explainability
- Data stewardship

The ability to interrogate all aspects of a workflow is the foundation of the definition of open source: access to the source code, data, and documentation. Current open scholarship ways of working will enhance transparency and explainability across the Al ecosystem.





27 Universities

12 NHS Trusts

& charities, city councils, and public involvement organisations.

https://www.turing.ac.uk/research/research-projects/ai-multiple-long-term-conditions-research-support-facility Slide by Sophia Batchelor: https://doi.org/10.5281/zenodo.8082165

[NEW CHAPTER] - Planning Inclusive and Accessible Events #3307

BrainonSilicon opened this issue on Oct 2 · 11 comments

This chapter is separated into different subsections:

- Chapter Landing page: has been reviewed and merged into the chapter's WIP branch.
 - Next review will be when the full chapter PR is made.
- The core chapter content: currently WIP is separated into different subsections
 - Planning and Logistics: \$\frac{1}{3}\$ Added Inclusive Planning and Logistics subsection #3315
 - Location: \$\frac{1}{2}\$ Added Inclusive Events Location subsection #3318
 - o Communication: \$\ Added Inclusive Events Communications Subsection #3327
 - o Schedule and Agenda Planning: \$\ Added Inclusive Events Schedule and Agenda Subsection #3337
 - Networking and Social Events
 - Technology and Hybrid Considerations
 - Content Accessibility Guidance
 - Facilitation and Moderation
- · A checklist summarizing the content:
 - o link to branch, link to file
- A page of resources:
 - o link to branch, link to file

- Safety and sustainability
- Accountability
- Fairness
- Explainability
- Data stewardship



Wilkinson, et al; The FAIR Guiding Principles for scientific data manageme and stewardship. https://doi.org/10.1038/sdata.2016.18

https://doi.org/10.5281/zenodo.10080932

@TuringWay @kirstie i

- Safety and sustainability
- Accountability
- Fairness
- Explainability
- Data stewardship

Metadata standards allow for interoperability even where data can not be made openly available. Documentation, accountability, and transparency around how data is gathered and used are the building blocks of open source AI.



Stop

- Focussing too narrowly on research outputs.
- Reinventing the wheel for each individual institution or domain.
- Avoiding the wicked problem of deployment challenges.

Start

- Promoting our solutions for accountability in an Al development workflow.
- Explicitly connecting data processing responsibilities with bias mitigation reporting frameworks in a socio-technical framework.

Continue

- Working transparently.
- Acknowledging diverse communities.
- Documenting data governance decisions.
- Supporting
 maintenance and
 infrastructure work.
- Scaling access to knowledge worldwide.

@TuringWay @kirstie_j

https://doi.org/10.5281/zenodo.10080932

"Opening science involves opening up the processes of creating, evaluating, sharing, exploring and storing scientific knowledge, practices and perspectives."

"All scientific disciplines and aspects of scholarly practices, including basic and applied sciences, natural and social sciences and the humanities, are included."

UNESCO, Understanding open science, 2022. https://doi.org/10.54677/UTCD9302



@TuringWay @kirstie_j https://doi.org/10.5281/zenodo.10080932

Thank you

- TPS Senior Researchers:
 Malvika Sharan, Aida Mehonic,
 Alden Conner, Jennifer Ding,
 Bastian Greshake Tzovaras &
 Christopher Burr
- Research Community
 Managers: Emma Karoune,
 Vicky Hellon, Anne Lee Steele,
 Sophia Batchelor, Eirini Zormpa,
 Gabin Kayumbi, Cassandra
 Gould Van Praag
- GC Research Ethics Manager: Victoria Kwan

- Research Application
 Managers: Hari Sood, Shakir
 Laher, Kalle Westerling, Sophie
 Arana & Francisco Gomez
 Medina
- Programme management:
 Arielle Bennett, Dave Chapman,
 Davide Sarmiento Perez, Batool
 Almarzouq, Alexandra Araujo
 Alvarez, Sydney Ambrose
- Research associates and fellows: Georgia Aitkenhead,
 Giulia Tomba, Priscilla Canizares
 @TuringWay @kirstie_j

https://doi.org/10.5281/zenodo.10080932

Please join us in The Turing Way!

- Book: https://the-turing-way.netlify.com
- Newsletter: https://tinyletter.com/TuringWay
- GitHub: https://github.com/alan-turing-institute/the-turing-way
- Slack: https://tinyurl.com/jointuringwayslack
- Book dash community share outs on Friday 17 November
 - https://www.eventbrite.co.uk/e/the-turing-way-book-dash-community-share-outs-and-celebrations-nov-2023-tickets-735393400457
- Original artwork by Scriberia: https://doi.org/10.5281/zenodo.3332807
- This work was supported by The UKRI Strategic Priorities Fund under the EPSRC Grant
 EP/T001569/1, particularly the "Tools, Practices and Systems" theme within that grant, and by The
 Alan Turing Institute under the EPSRC grant EP/N510129/1.
 @TuringWay @kirstie_j

