

**The
Alan Turing
Institute**



**Open source practices
underpinning the
SAFE-D principles**

Kirstie Whitaker

Pronouns: she/her

@TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.8381378>

A photograph of a forest path with sunlight streaming through the trees, creating a warm and bright atmosphere. The sun is visible in the upper right corner, casting rays of light across the scene. The trees are lush green, and the path is covered in fallen leaves and grass.

**“Sunlight is the
best disinfectant”**

Louis Brandeis

https://en.wiktionary.org/wiki/sunlight_is_the_best_disinfectant

@TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.8381378>

Open source practices already deliver safe and ethical AI*

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

Open source practices already deliver safe and ethical AI*

(*Depending on how you define open source!)

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

@TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.8381378>

Open source practices already deliver safe and ethical AI*

(* Depending on how you define open source!)

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

@TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.8381378>

A large field of glowing jack-o'-lanterns at night, with a semi-transparent text box in the center. The pumpkins are carved with various faces and are illuminated from within, creating a warm, orange glow. The background is dark, with some blue and purple lights visible in the distance.

“Given enough eyeballs, all bugs are shallow”

Eric Steven Raymond, *The Cathedral and the Bazaar*
Linus' Law: https://en.wikipedia.org/wiki/Linus%27s_law

@TuringWay @kirstie_
<https://doi.org/10.5281/zenodo.8381378>

Open source practices already deliver safe and ethical AI*

(* Depending on how you define open source!)

- **Safety and sustainability** Having a broad and engaged community involved throughout an AI development workflow keeps infrastructure more secure and keeps the purpose of the work aligned with the needs of the impacted stakeholders.
- Accountability
- Fairness
- Explainability
- Data stewardship

@TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.8381378>


Open source practices already deliver safe and ethical AI*

(* Depending on how you define open source!)

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

@TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.8381378>

The background is a dark-themed Git commit history. It shows a vertical timeline of commits with colored dots (blue, pink, green) and commit messages. Some visible commit messages include "feat: dynamic directive arguments for v-on, v-bind and custom directives", "perf: improve scoped slots change detection accuracy", and "fix: fix checkbox event edge case in Firefox". A semi-transparent grey box is overlaid on the center, containing white text.

“Anyone could say anything...nothing was official. And to emphasize the point, I labelled the notes ‘Request for Comments.’”

Open source practices already deliver safe and ethical AI*

(* Depending on how you define open source!)

- 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.

@TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.8381378>

Open source practices already deliver safe and ethical AI*

(* Depending on how you define open source!)

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

@TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.8381378>



**“You don’t need a
diversity committee if
you have a diverse
committee”**

Who is building open source AI?

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

Register on Eventbrite

Defining



Arielle Bennett

Programme Manager
The Alan Turing Institute



Mophat Okinyi

Union Representative
African Content Moderators Union



Marzieh Fadaee

Senior Research Scientist
Cohere for AI



Abinaya Mahendiran

CTO
Nunnari Labs



David Gray Widder

Postdoctoral Fellow
Cornell Tech



Jennifer Ding

Senior Researcher
The Alan Turing Institute



Open source practices already deliver safe and ethical AI*

(* Depending on how you define open source!)

- 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. The open source AI community can embed existing tools for bias reporting into an interoperable open source AI ecosystem.

@TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.8381378>

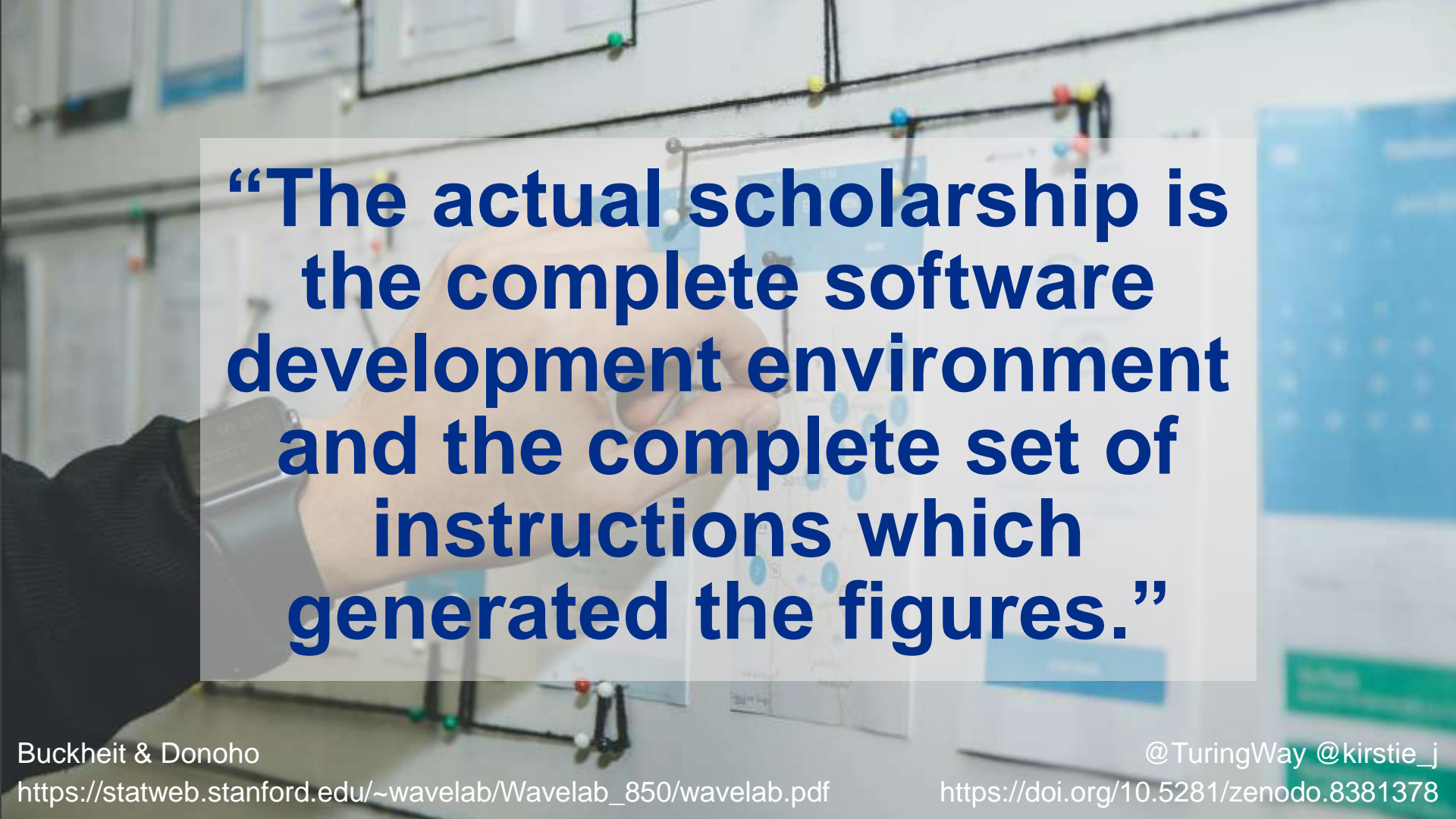
Open source practices already deliver safe and ethical AI*

(* Depending on how you define open source!)

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

@TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.8381378>

A hand in a dark sleeve points towards a whiteboard. The whiteboard is covered with various diagrams, charts, and notes, some of which are pinned with colorful pushpins. The background is slightly blurred, showing more of the whiteboard and some office equipment.

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

Open source practices already deliver safe and ethical AI*

(* Depending on how you define open source!)

- 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 source ways of working will enhance transparency and explainability across the AI ecosystem.

@TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.8381378>

Open source practices already deliver safe and ethical AI*

(* Depending on how you define open source!)

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

@TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.8381378>



**“Digital assets should
be findable, accessible,
interoperable, and
reusable”**

Open source practices already deliver safe and ethical AI*

(* Depending on how you define open source!)

- 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.

@TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.8381378>



Stop
Start
Continue

@TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.8381378>

Stop

- Thinking of ourselves as separate, disconnected, individual people and projects.
- Avoiding responsibilities around data governance, and crediting the people who are affected by the tools or models we build.

Start

- Building in accountability for decisions across an AI development workflow.
- Explicitly connecting data processing responsibilities with bias mitigation reporting frameworks through the tools we build.

Continue

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

@TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.8381378>



@TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.8381378>

<https://the-turing-way.netlify.com>

Operationalising the SAFE-D principles for Open Source AI

The
Alan Turing
Institute



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

