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

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

@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

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

Eric Steven Raymond, The Cathedral and the Bazaar

Linus’ Law: https://en.wikipedia.org/wiki/Linus%27s_law

@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

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.

@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

Stephen Crocker, Request For Comments (RFC's) History
https://www.livinginternet.com/i/ia_rfc_invent.htm

@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

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

Teon Brooks, https://teonbrooks.com

@TuringWay @kirstie_j

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

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 AI

Abinaya Mahendiran
CTO
Nunnari Labs

David Gray Widder
Postdoctoral Fellow
Cornell Tech

Jennifer Ding
Senior Researcher
The Alan Turing Institute

Hosted by:

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.

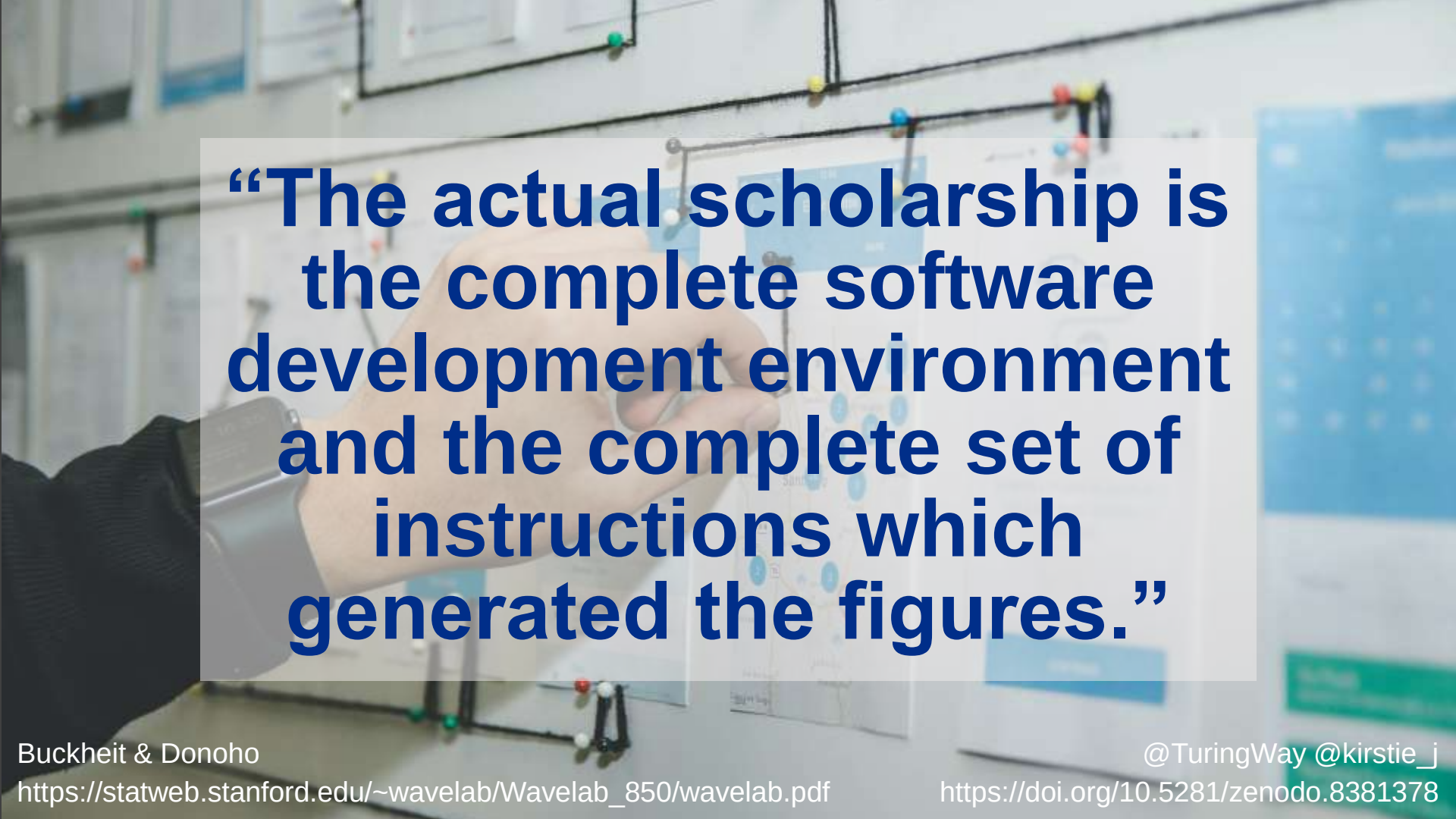
@TuringWay @kirstie_j
https://doi.org/10.5281/zenodo.8381378

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



Buckheit & Donoho
https://statweb.stanford.edu/~wavelab/Wavelab_850/wavelab.pdf

@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

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

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

@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

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

@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

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

Operationalising the SAFE-D principles for Open Source AI

The Alan Turing Institute

open source initiative®

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