unbias.wp.horizon.ac.uk

Economy (EP/N02785X/1) UnBias: Emancipating Users Against Algorithmic Biases for a Trusted Digital Funded through EPSRC:

> Attribution-NonCommercial-ShareAlike 4.0 International : snommo Saviteative Commons :

> > Published by Proboscis 2018

Illustrations by Alice Angus Written, Edited & Designed by Giles Lane

UnBias Fairness Toolkit Handbook

quality or value of their responses. Perception worksheet (p. 22) to evaluate any MetaMaps and the

and debate. Additionally, you might use the Participant Value MetaMaps can also provide further opportunities for discussion others (see the next section on Stakeholder MetaMaps). The people working in the technology industry, policymakers or If your TrustScape is shared online, it may be responded to by

What Happens Next

and use them to discuss issues in class or group settings. Check out UnBias' website and social media for other TrustScapes

Twitter) with the hashtag: #trustscape We will post it on UnBias' website and social media (Instagram &

| gniad-llaw | lsutriv | verifiable | huuecessau | kuomledge | justice | noitnevretni | lemrotni |
|----------------|--------------|----------------|-------------|----------------|---------------|--------------|--------------|
| lutweinu | reilimetnu | ristru | uowwooun | ssəulli | lsgəlli | ignored | ignorance |
| uncomfortable | uncertain | unaware | truth | wollod | uəppiq | health | sttuobned |
| trust | true | £ransparency | tracked | рэскөд | dissog | рооб | əuinuəb |
| pədwems | noisiqsus | surveillance | surface | bnəiri | mobeeri | free | femrof |
| lsioiheque | uoitseggus | leitnetedue | stress | gniwollof | filter bubble | filter | fiction |
| egnerts | stereotyped | oup sufets | static | familiar | eslef | take news | tair |
| brebnete | bilos | lsioos | spared | tost | betiolqxe | ədessə | euemy |
| 6uilles | secret | ysafety | rumour | eucıλbtion | pəlqeuə | embowered | есро срашрек |
| ınles | rubbish | risk | review | dynamic | distraction | disruption | feiledsib |
| restriction | reputation | representation | reliable | different | bedached | qebth | lsinab |
| released | regulations | кесошшеид | reality | avitinitab | deception | qsuðer | convenience |
| rebellion | ratings | paddud | protest | control | connection | noizutnoo | conformity |
| protection | tiforq | private | brivacy | compromise | common | comfortable | confined |
| brevented | bressure | bower | positive | closed | clear | mislo | choice |
| bersonal | permission | pəumo | uədo | срапде | certify | certain | central |
| opscnie | normal | avitegan | necessary | cede | puying | seid | feiled |
| perotinom | moderate | tsurtsim | ssəlgninsəm | bewolloł gnied | pehaviour | barriers | nogewbned |
| lenigrem | noiteluqinem | nagnetic | ssol | peq | aware | betemotus | attention |
| səloydool | lοcked up | noitsool | səil | behbette | gnxiety | snongidma | pəwolls |
| esneoil | legal | swel | lutwal | addiction | scknowledged | accountable | scess |

Keywords

FAIRNESS TOOLKIT HANDBOOK

** ABM**

UNSIAS



FAIRNESS TOOLKIT HANDBOOK

*RAIBAU

Sketches

























policy and decison-makers) respond to the issues raised in the TrustScapes and contribute to an evolving public civic debate.

- feelings about the issues in a shareable format;

 MetaMaps to help stakeholders (industry, researchers,
- critically about the issues in an engaging way;
 TrustScapes to help people express their perceptions and
 - The toolkit consists of three main parts:

 Awareness Cards to help people to explore and think

such technologies often cause.

The Fairness Toolkit is one of our project outputs aiming to promote awareness and stimulate a public civic dialogue about how algorithms shape online experiences and to reflect on possible changes that can address issues of online unfairness. The tools are not just for critical thinking, but for civic thinking and doing – supporting a more collective approach to imagining the future, in contrast to the individual atomising effect that

The UnBias project explores the user experience of algorithm design. We driven internet sites and the processes of algorithm design. We are particularly interested in circumstances in which algorithmic processes might (intentionally or unintentionally) produce biased or unfair outcomes – for instance in the form of helping fake content to spread on social media, producing search results that reinforce prejudiced attitudes, or the excessive personalisation of content and collection of personal data.

Increasingly, our everyday interactions in the modern world are mediated and filtered through algorithms: analysing our behaviours, inferring our preferences and channeling the choices we are offered. These algorithms are embedded both within the personal computers and devices we use, as well as within the larger infrastructures that manage and facilitate everyday life.

INTRODUCTION

25



FAIRNESS TOOLKIT HANDBOOK

CONTENTS

Acknowledgements

| Introduction | 3 |
|---------------------|----|
| Awareness Cards | 5 |
| TrustScape | 11 |
| Stakeholder MetaMap | 17 |
| Value Perception | 21 |
| Resources | 24 |
| | |

Email us at : unbias@cs.ox.ac.uk

young people, or users in general?

- has it helped you understand more about the experiences of
 - do you feel your ideas have been listened to?
- would like to see happen to make algorithms and the online systems that depend on them fair, trustworthy and unbiased?

 has it helped you articulate and express your ideas?
 - how much personal data you share?

 has it helped you form your own opinion about what you
- algorithms can be biased, untrustworthy or unfair?

 has it inspired you to change your use of online systems and
 - has it been helfpul in raising your awareness of how

We welcome comments or feedback:

Send us feedback

Citizens Advice – www.citizensadvice.org.uk
Equality & Human Rights Commission – www.equalityhumanrights.com
UK Information Commissioner – www.ico.org.uk
National Health Service – www.nhs.uk
Health & Safety Executive – www.hse.gov.uk
The Guardian – www.theguardian.com
Gizmodo – www.gizmodo.com
Gizmodo – www.gizmodo.com
Arstechnica – arstechnica.com
ProPublica – www.propublica.com
Bloomberg – www.bloomberg.com
Bloomberg – www.bloomberg.com
Bloomberg – www.bloomberg.com
Bloomberg – www.bloomberg.com

Content for the cards was sourced from many different research studies, sites and media organisations, including:

Sources

EVIRNESS TOOLKIT HANDBOOK



UnBias Fairness Toolkit Handbook

Giles Lane

2018-08-31 & CC BY-NC-SA

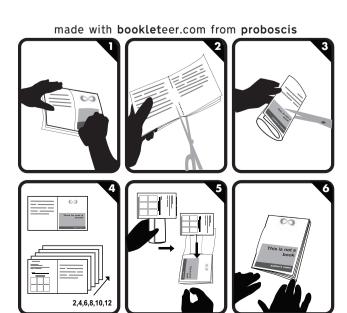
Published by Proboscis

A project of the Horizon Digital Economy Institute

(University of Nottingham), University of Oxford,

University of Edinburgh & Proboscis

unbias.wp.horizon.ac.uk





https://tinyurl.com/our-future-internet



responses to the Toolkit on our moderated channel: Post your own videos inspired by the animation, or your

UnBias Youtube Channel

https://youtu.be/ovAvjDgaqWk



the issues at its heart: A short animation by Scriberia introduces the UnBias project and

UnBias Animation

回读自 表情報 https://uyj.wp.horizon.ac.uk 回读的



resource to accompany their Youth Jury programme: University of Nottingham have created an open educational Our colleagues at the Horizon Digital Economy Institute,

Open Educational Resource

https://unbias.wp.horizon.ac.uk/fairness-toolkit



compressed Zip archive here: The Toolkit is available to download as individual files or in a

Download



FAIRNESS TOOLKIT HANDBOOK



ACKNOWLEDGEMENTS

Toolkit Concept & Design : Giles Lane Illustrations : Alice Angus

Designing the toolkit was a participatory process involving the following people, to whom many thanks are offered:

UnBias Team

proboscis.org.uk

Horizon Digital Economy Institute, University of Nottingham Professor Derek McAuley, Dr Elvira Perez Vallejos, Dr Ansgar Koene, Dr Liz Dowthwaite, Dr Virginia Portillo, Monica Cano Gomez & Dr Helen Creswick.

Department of Computer Science, University of Oxford Professor Marina Jirotka, Dr Helena Webb, Dr Menisha Patel, Paula Fiddi & Ross Gales.

School of Informatics, University of Edinburgh Dr Michael Rovatsos, Dr Sofia Ceppi & Dr Alan Davoust.

Co-Design Participants:

Staff & Students at Sir John Lawes School, Harpenden Staff & Students at *Elizabeth Garret Anderson School*, Islington Mel Jewell, Community EET Worker at Oxfordshire County **Council** and members of the **go_girl:code** and create group

Special thanks to theatremaker and improv expert Alex Murdoch for devising the Awareness Cards Exercises.

trust, prejudice, unfairness and discrimination operate. Exercise – 8 dynamic group activities for exploring how bias,

The consequences of its decisions? and other factors which influence and effect an algorithm and Process – 8 "Be the Algorithm" cards: what are the inputs, steps

and untrustworthiness.

Example - 12 real life examples of algorithmic bias, unfairness

The deck has eight different suits, or types, of cards:

The Deck

any way that suits you or your group best. There are no right or wrong ways to use this deck, use them in

or untrustworthy decisions and outcomes are taking place? • How do we know when, where, how and why bias, unfairness

other factors in algorithmic systems? What are the relationships between data, rights, values and but as communities and societies too.

we do in everyday life. They affect us not just as individuals, interview, who gets a bank loan, and many other things that dictating our news feeds, internet searches, who gets a job Algorithms are everywhere, affecting our lives and privacy,

in algorithmic decisions. racial and gender profiling and misinformation often hidden and activities to inspire new ideas which can tackle the bias, that they have on our lives. Use the cards to start conversations exploring how decisions are made by algorithms and the impact The deck of Awareness Cards is a civic thinking and doing tool for

Purpose

AWARENESS CARDS



FAIRNESS TOOLKIT HANDBOOK

It also includes additional materials:

- Value Perception Worksheets for assessing and evaluating the value of using the toolkit;
- Sketch Sheets illustrations by artist, Alice Angus, which are derived from our co-design workshops with young people visualising many of their concerns and feelings about the internet, data and algorithms;
- Keyword Sheets keywords also derived from our co-design workshops with young people.

Co-Design Workshop Findings

The toolkit has been co-created with young people and stakeholders whose input has greatly informed the design process. We conducted a series of workshops with young people aged twelve to seventeen in two schools and with a group of sixteen to twenty-two year olds in a community centre to understand more about how aware they are of the issues, how relevant to their own lives they perceive them to be and, what they thought should be done.

Our observations revealed a consistent fatalism and lack of a sense of agency linked to the highly personalised nature of young people's interactions with the technologies. This fostered a sense of isolation and inability to effect change, coupled with a very limited sense of their rights and how the law already protects them in their interactions with service providers and big companies. They often feel that their voice is neither heard nor listened to, even when they are the targets of some of the most aggressive marketing techniques.

Many of the participants were surprised both at the scale of algorithmic systems and how much their data and behaviours are already tracked across the internet by government and industry. What might start with a degree of apathy and lack of concern always ended with intense debate and with a sense that their generation will be the next to shape and determine the future. This toolkit should help focus and amplify such debates.



EVIRNESS TOOLKIT HANDBOOK





What are algorithms? Put simply, an algorithm is a process or list of rules to follow in order to complete a task: such as solving a problem, doing a calculation or making a decision. The cards are designed to help people develop key critical skills in analysing and interpreting how and why decision making takes place, as well as the impacts

their effects might be.

Use the cards to help people develop critical thinking skills. They help analyse how and why decisions are made: what role algorithms and data sources play in automated systems; where and how bias, unfairness and untrustworthy occur; and what

Critical Thinking Skills

fairness and trust.

The **Process** and **Exercise** cards suggest a range of activities for participants to explore the issues in different ways: from the analytical to the experiential, the logical to the discursive, the solitary to the social. Participants can draw upon their own experiences, or on stories and reports they have encountered in the media. The Exercises provide oportunities for role play, im the media. The Description and shared experiences of bias, improvisation, collaboration and shared experiences of bias,

further stimulus and inspiration.

The Data, Rights, Values and Factors cards can be used as specific discussion topics that encompass wider themes and issues affecting society. They can also be used with the Process and Exercise cards to add depth to the discussion and to provide

The **Example** cards are ideal for introducing the issues of bias, trust and fairness in algorithmic systems with real world examples.

different combinations.

There are no specific rules to use the cards: you can use the whole deck or pick individual suits or cards and use them in

Getting Started



FAIRNESS TOOLKIT HANDBOOK

CALUES ## CAL

Data – 12 cards describing types of personal data we commonly share across numerous platforms and services.

Values – 10 cards with over one hundred values that typically motivate people in everyday life.

Rights – 8 cards summarising key rights upheld by United Kingdom law: human rights, disability rights, equality rights, consumer rights, employment rights, data protection rights, health and safety and patient rights.

Factors – 4 cards describing common factors which affect decision-making – whether human or machine-automated – bias & prejudice, discrimination, trust and fairness.

Glossary – a simple explanation of an algorithm is. An additional card describes Artificial Intelligence and Machine Learning.

Recognising how bias affects your daily life.

How can we detect bias in historical sources?

Why do we trust the media?

the curriculum:What decides the stories in our news feeds?

Teachers and educators can use the cards for debates and across activities in classes or workshops in different subjects and across

noitseitsevnl to soidoT

| Commitment | ssənəvitnəttA | ssəulnɨthguodT |
|-----------------|----------------|-----------------|
| Patience | noitiutal | noisi∨ |
| Tenacity | Determination | Сһеек |
| τί₩ | Grace | Humour |
| thgisnl | Vitality | Иегvе |
| Honesty | AtisA | un∃ |
| Sensitivity | рвак | Kindness |
| Pellef | ssəulnɨtdguodT | Empathy |
| Awareness | норе | Approachability |
| Talent | дэпод | Qninətsid |
| Beauty | Sonfidence | Expertise |
| Drive | Creativity | Presence |
| Integrity | Skills | Life Experience |
| msimitqO | Strength | Vulnerability |
| Challenge | hodduð | Courage |
| Resourcefulness | mobsiVV | Playfulness |
| Curiosity | Education | Engagement |
| Pelief | Inspiration | Diversity |
| noissa9 | Enthusiasm | Peacefulness |
| Concentration | Reflection | ssənəvitisiupnl |
| Understanding | Focus | ssəuuədO |

beoble or systems make decisions:

Personal qualities
People are complex and often have contradictory beliefs and values. Below are some qualities to consider when you are using the cards; combine them to discuss the ways they might influence



FAIRNESS TOOLKIT HANDBOOK



and consequences it can have.

Why is this important?

Decisions about our lives are made by systems that are increasingly being automated. The cards can help build awareness of where and how such decisions occur. They also help to develop critical thinking capacities about whether or not the decisions may be biased, and what factors might motivate any discriminatory or unfair outcomes.

Risk and trust

Risk and trust are inseparable and intertwined. The more we are at risk of vulnerability, the more important trust becomes. Many of the Exercises rely on and explore trust between participants. They also offer a space of self-reflection about how we trust ourselves, and how we perceive our world-view: as safe or unsafe, hostile or friendly, happy or sad, open or closed etc.

The **Stakeholder Value Perception** worksheet (p. 23) can also be used as an evaluation tool to assess what value you think participation in the process has brought you.

Complete as many MetaMaps as you wish: we aim for the public civic dialogue to evolve and expand over time.

Further Reflections

Check out UnBias' website and social media for other MetaMaps to see how peers are also contributing to the dialogue.

Photograph your completed MetaMap and email it to: unbias@cs.ox.ac.uk We will post it on UnBias' website and social media (Instagram & Twitter) with the hashtags: #metamap & #trustscape

How to Share

EAIRNESS TOOLKIT HANDBOOK



VALUE PERCEPTION

Purpose

The value perception worksheets have two purposes:

- 1 as a prospective tool both participants and stakeholders to imagine the potential value of participating in the UnBias public civic dialogue.
- 2 as an evaluation tool for reflecting on what participation has actually resulted in, for instance for participants to evaluate any responses to the TrustScapes in Stakeholders' MetaMaps.

Both worksheets use a matrix of hexagons to map out motivations, opportunities, costs, values and other factors in using the TrustScapes and MetaMaps.

How to Use

Print out a copy at the size most useful for your purposes, e.g.: for individuals – A3 size

for a small group – A2 size

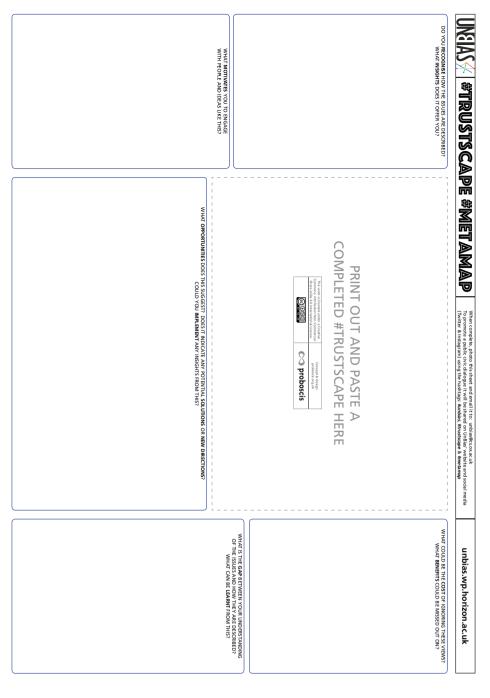
for larger groups – A1 poster size.

Bigger sizes can be printed cheaply at local copy shops.

The Value Perception sheets are intended as personal or group reflection tools, not as contributions to the public civic dialogue.

However, if you do wish to share a completed – and anonymous – worksheet, please email it to us: unbias@cs.ox.ac.uk We will post it on UnBias' website and social media (Instagram & Twitter).

21



FAIRNESS TOOLKIT HANDBOOK



How to Use

MetaMaps can be completed by an individual or by a group. Print out a copy at the size most useful for your purposes, e.g.:

for individuals – A3 size

for a small group - A2 size

for larger groups - A1 poster size.

Bigger sizes can be printed cheaply at local copy shops.

Select a TrustScape from those shared online via the UnBias social media (Instagram, Twitter & Facebook). Print it out and stick it onto the MetaMap in the box provided:

for A3 MetaMaps – print TrustScape at A5 size

for A2 MetaMaps – print TrustScape at A4 size

for A1 MetaMaps – print TrustScape at A3 size

Responding to a TrustScape

Bear in mind that, although anonymous, each TrustScape is the work of a person or group who are sharing their perceptions and feelings about crucial issues affecting our society. Try to respond in both an informative way and an aspirational way.

Please remember to be respectful about the young people's perceptions and feelings – your response will be a contribution to a public civic dialogue which they will be part of.

Sketches & Keywords

The toolkit also includes sketches and keywords you may wish to use in your own response. Print them out on label sheets and stick them in the boxes or use them as inspiration for your own words and drawings.

Keep Your MetaMap Anonymous

MetaMaps are designed to be shared online so please *do not* add any personal details or information that could identify you. But don't worry if you do – we'll anonymise any we spot that need it.

a TrustScape.

If you are using the TrustScapes in a class or group setting, you might like to use the **Participant Value Perception** worksheet (p. 22) to assess what you hope to get out of completing a TrustScape and participating in the UnBias public civic dialogue. This can be a useful exercise in helping to decide what topics and themes you are most concerned about and wish to articulate on themes you are most concerned about and wish to articulate on

Completed TrustScapes form the focus of the MetaMap tool (p. 17) which allows other people, such as stakeholders in the ICT industry, policymakers, regulators and researchers to respond to these concerns and contribute to the public civic dialogue.

internet could or should be.

TrustScapes form the first element in a public civic dialogue initiated by the UnBias project to amplify the voices of users (especially young people) and engage with stakeholders in industry, research, policy and regulation. They are designed to capture, in a dynamic and visual way, how people feel about the current situation and their dreams and ideals for what the

The TrustScape is a worksheet for people, especially young people, to visualise their perceptions and feelings about algorithmic bias, data protection and online safety and what they would like to see done to make the online world fair and trustworthy. Building on the Awareness Cards, the TrustScapes offer people a civic thinking tool to articulate and share their concerns and visions for trust and fairness.

Purpose

TRUSTSCAPE



FAIRNESS TOOLKIT HANDBOOK

- Acknowledging your own biases.
- Investigating the consequences of your own bias.
- How do you know if you have acted fairly?
- What makes us trust social media?
- How are agorithms created and used in computer science?
- Addressing bias when writing computer programmes.
- Are search engines trustworthy?
- How can we discern facts from opinions?
- How can you collect, analyse and represent data to prevent bias?
- Can you prevent bias in statistics?
- How do you identify information sources that are reliable and trustworthy?
- How does bias affect the use of data and graphs in different subjects?

Games & Play

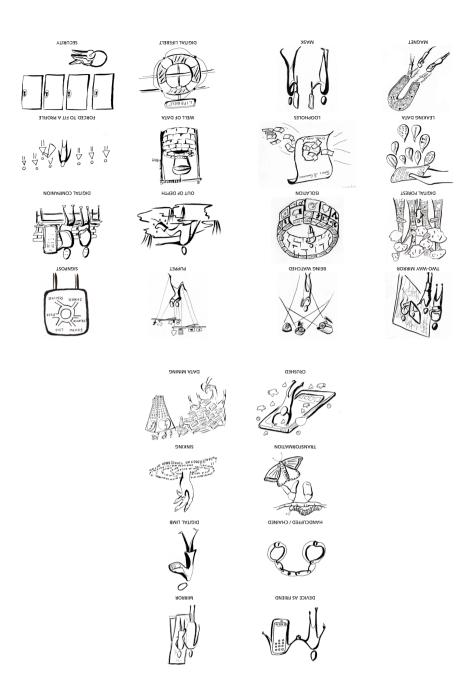
Adapt games you already know as a way of familiarising yourself and others with the issues. For instance:

HAPPY FAMILIES: Shuffle the deck and deal them out amongst the players. Players then take turns to pick a card from the player next to them. When a player has 4 cards from any of the same suits they can lay them down. The aim is to have as many suits of 4 as possible.

TRUMPS: Pick an Example Card place it face up. Select the Data, Rights, Factors and Values Cards, shuffle them and deal them out among the players. Players then take turns to add their cards on top explaining why they are relevant to the Example. The winner will have put down the most relevant cards per example.

GUESS WHO?: Use the Data, Rights, Factors, Example and Values cards. Shuffle them and deal each player a card. Players then take turns in asking each other one question to try to work out which card each player has. The correct guess wins the card, the aim being to win as many cards as possible.

1



EVIENESS TOOLKIT HANDBOOK



STAKEHOLDER METAMAP

Purpose

The MetaMap is a worksheet for stakeholders in the ICT industry, policymaking, regulation, public sector and research to respond to the TrustScapes. By selecting and incorporating a TrustScape from those shared online, stakeholders can respond to the participants' perceptions.

MetaMaps will also be captured and shared online via UnBias social media to enhance the public civic dialogue, demonstrating the value of participation to young people in having their voice listened and replied to.

MetaMaps offer an anonymous means for stakeholders to participate in a public civic dialogue about issues that are of fundamental importance to our society. They provide a civic thinking tool that adds value to the voices and concerns of users represented in the TrustScapes and which validates their participation in a responsible, civic process.

Photograph your completed TrustScape and email it to: unbias@cs.ox.ac.uk

How to Share

.ti bəən

Keep Your TrustScape Anonymous TrustScapes are designed to be shared online so please **do not** add any personal details or information that could identify you. But don't worry if you do – we'll anonymise any we spot that

But they are not definitive: there will be many others which affect specific groups of people and reflect different kinds of bias than we have identified here. TrustScapes are an opportunity to articulate and share these issues, and seek responses from stakeholders about what can be done about them.

affect young people today.

These sketches and keywords emerged from our co-design workshops with young people, as well as wider research by the UnBias team. They illustrate and describe many of the contemporary themes about our digital, networked world that

Themes, Sketches & Keywords

The toolkit includes lots of sketches and keywords to inspire you.

Print them out on label sheets and stick them in the boxes or use them as inspiration for your own words and drawings.

Bigger sizes can be printed cheaply at local copy shops.

for larger groups – A1 poster size.

for a small group – A2 size

for individuals – A3 size

TrustScapes can be completed by an individual or by a group. Print out a copy at the size most useful for your purposes, e.g.:

How to Use



FAIRNESS TOOLKIT HANDBOOK

