

unbias.wp.horizon.ac.uk

Funded through EPSRC:
UnBias: Empowering Users Against Algorithmic Biases for a Trusted Digital Economy (EP/N02785X/1)
Attribution-NonCommercial-ShareAlike 4.0 International
Licensed under Creative Commons :
Published by ProboScis 2018
Written, Edited & Designed by Giles Lane
Illustrations by Alice Angus

Unbias Fairness Toolkit Handbook

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

Published by ProboScis 2018

Attribution-NonCommercial-ShareAlike 4.0 International

Funded through EPSRC:

UnBias: Empowering Users Against Algorithmic Biases for a Trusted Digital Economy (EP/N02785X/1)

unbias.wp.horizon.ac.uk

What Happens Next

If your Trustscape is shared online, it may be responded to by people working in the technology industry, policymakers or others (see the next section on Stakeholder MetaMaps). The MetaMaps can also provide further opportunities for discussion and debate. Additionally, you might use the *Participant Value Perception* worksheet (p. 22) to evaluate any MetaMaps and the quality or value of their responses.

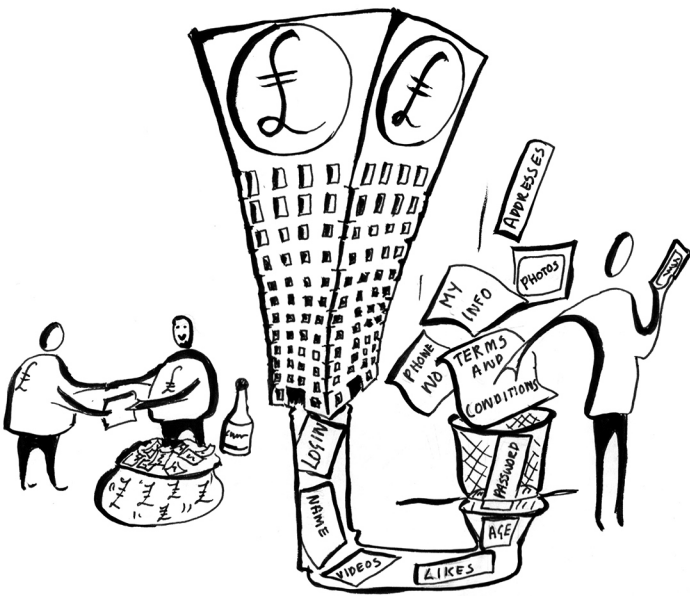
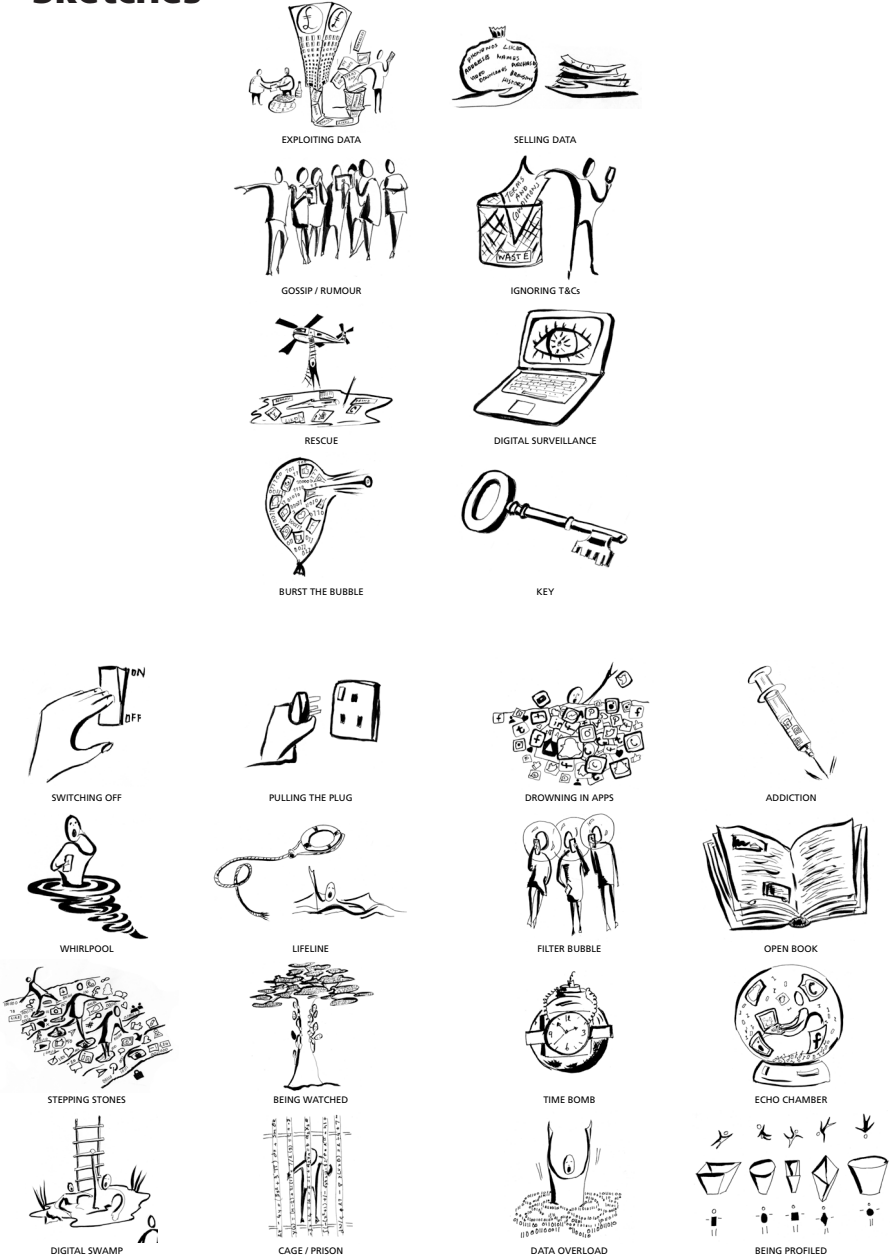
Check out Unbias' website and social media for other Trustscapes and use them to discuss issues in class or group settings.

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

informal	ignorance	handicuffs	genuine	formal	fiction	fair	enemy	echo chamber	disbelief	denial	convenience	conformity	confined	choice	central	belief	bandwagon	allowed	access
intervention	ignored	health	good	free	filter	fake news	escape	empowered	disruption	depth	danger	confusion	comfortable	claim	certain	bias	barriers	ambitious	accountable
justice	illegal	hidden	gossip	freedom	filter bubble	false	exploited	enabled	distracted	detached	deception	control	common	clear	certify	buying	behaviour	aware	acknowledged
knowledge	illness	hollow	hacked	friend	following	familiar	fact	encryption	dynamic	different	definitive	compromise	closed	change	cage	being followed	bad	attached	addition
unnecessary	uncommon	truth	tracked	surface	stress	static	shared	rumour	review	reliable	reality	protest	privacy	positive	open	meaningless	lies	lawful	
verifiable	unfair	unaware	transparency	surveillance	substantial	status quo	safety	risk	representation	recommend	regulations	profit	power	owned	negative	mistrust	magnetic	location	laws
virtual	unfamiliar	uncertain	true	suspicion	suggestive	stereotyped	solid	secret	reputation	ratings	personal	pressure	permission	normal	moderate	manipulation	locked up	legal	
well-being	unlawful	uncomfortable	trust	swamped	superficial	strange	selling	rules	restriction	released	rebellion	protection	prevented	personal	obscure	monitored	marginal	loopholes	license

Keywords

Sketches



FAIRNESS TOOLKIT HANDBOOK

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 such technologies often cause.

The Fairness Toolkit consists of three main parts:

- *Awareness Cards* – to help people to explore and think critically about the issues in an engaging way;
- *TrustScapes* – to help people express their perceptions and feelings about the issues in a shareable format;
- *MetaMaps* – to help stakeholders (industry, researchers, policy and decision-makers) respond to the issues raised in the TrustScapes and contribute to an evolving public civic debate.

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.

The UnBias project explores the user experience of algorithm 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.

INTRODUCTION



FAIRNESS TOOLKIT HANDBOOK

CONTENTS

Introduction	3
Awareness Cards	5
TrustScape	11
Stakeholder MetaMap	17
Value Perception	21
Resources	24
Acknowledgements	25

We welcome comments or feedback:

- has it been helpful in raising your awareness of how algorithms can be biased, trustworthy or unfair?
- has it inspired you to change your use of online systems and how much personal data you share?
- has it helped you form your own opinion about what you 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?
- do you feel your ideas have been listened to?
- has it helped you understand more about the experiences of young people, or users in general?

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

Send us feedback

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

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
Arstechnica – arstechnica.com
ProPublica – www.propublica.com
Bloomberg – www.bloomberg.com
IEEE Use Cases for P7003 Standard – sites.ieee.org/sagroups-7003
BBC Bitesize – www.bbc.com/education

Sources

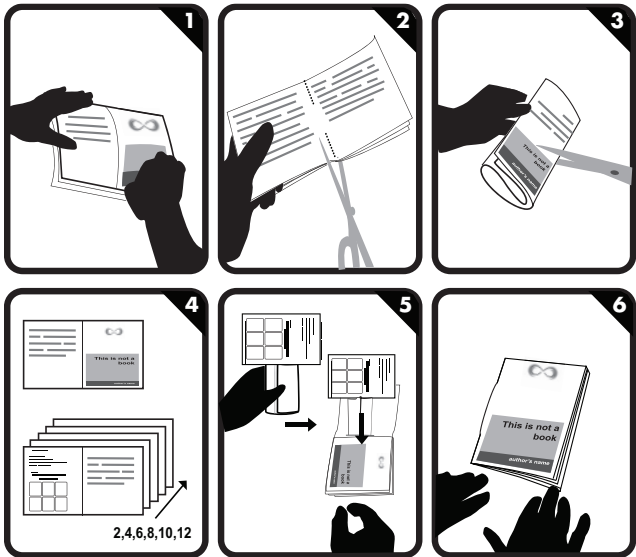
FAIRNESS TOOLKIT HANDBOOK



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

made with bookleteer.com from [proboscis](http://proboscis.org)



<http://bkltr.it/2laV46E>

The Deck

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

Example – 12 real life examples of algorithmic bias, unfairness and untrustworthiness.

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

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

Purpose

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

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

- What are the relationships between data, rights, values and other factors in algorithmic systems?
- How do we know when, where, how and why bias, unfairness or untrustworthy decisions and outcomes are taking place?

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

AWAWARENESS CARDS



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.

Download


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



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

Open Educational Resource


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



<https://ujf.wp.horizon.ac.uk>

UnBias Animation


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



<https://youtu.be/ovAvjDgaqWk>

UnBias Youtube Channel

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



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

RESOURCES



ACKNOWLEDGEMENTS

Toolkit Concept & Design : Giles Lane
Illustrations : Alice Angus
proboscis.org.uk

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

UnBias Team
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 Jirotko, 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.

- What decides the stories in our news feeds?
- Why do we trust the media?
- How can we detect bias in historical sources?
- Recognising how bias affects your daily life.

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

Topics of Investigation

Openness	Focus	Understanding
Inquisitiveness	Reflection	Concentration
Peacefulness	Enthusiasm	Passion
Diversity	Inspiration	Belief
Engagement	Education	Curiosity
Playfulness	Wisdom	Resourcefulness
Courage	Support	Challenge
Vulnerability	Strength	Optimism
Life Experience	Skills	Integrity
Presence	Creativity	Drive
Expertise	Confidence	Beauty
Listening	Youth	Talent
Approachability	Hope	Awareness
Empathy	Thoughtfulness	Belief
Kindness	Spark	Sensitivity
Fun	Faith	Honesty
Nerve	Vitality	Insight
Humour	Grace	Wit
Cheek	Determination	Tenacity
Vision	Intuition	Patience
Thoughtfulness	Attentiveness	Commitment

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 people or systems make decisions:

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

Further Reflections

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

How to Share

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).



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.

TRUSTSCAPE

Purpose



- 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 algorithms 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.

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 a Trustscape.

to these concerns and contribute to the public civic dialogue.

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

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

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

concerns and visions for trust and fairness.

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

Purpose

#TRUSTSCAPE #METAMAP

unbias.wp.horizon.ac.uk

DO YOU RECOGNISE HOW THE ISSUES ARE DESCRIBED?
WHAT INSIGHTS DOES IT OFFER YOU?

WHAT MOTIVATES YOU TO ENGAGE
WITH PEOPLE AND IDEAS LIKE THIS?

PRINT OUT AND PASTE A
COMPLETED #TRUSTSCAPE HERE

<small>This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License</small> 	<small>Created by Benjamin probascois.org.uk</small>
--	--

WHAT IS THE GAP BETWEEN YOUR UNDERSTANDING
OF THE ISSUES AND WHAT YOU WOULD
WHAT CAN BE LEARNED FROM THIS?

WHAT COULD BE THE COST OF IGNORING THESE VIEWS?
WHAT BENEFITS COULD BE MISSED OUT ON?

When complete, photo this sheet and email it to: unbias@ox.ac.uk
To provide a public/private dialogue, it will be placed on Unbias's website and social media
(Twitter & LinkedIn) using the hashtag #Unbias, #Trustscape & #Metamap.

WHAT OPPORTUNITIES DOES THIS SUGGEST? DOES IT INDICATE ANY POTENTIAL SOLUTIONS OR NEW DIRECTIONS?
COULD YOU IMPLEMENT ANY INSIGHTS FROM THIS?



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 TrustScale 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.

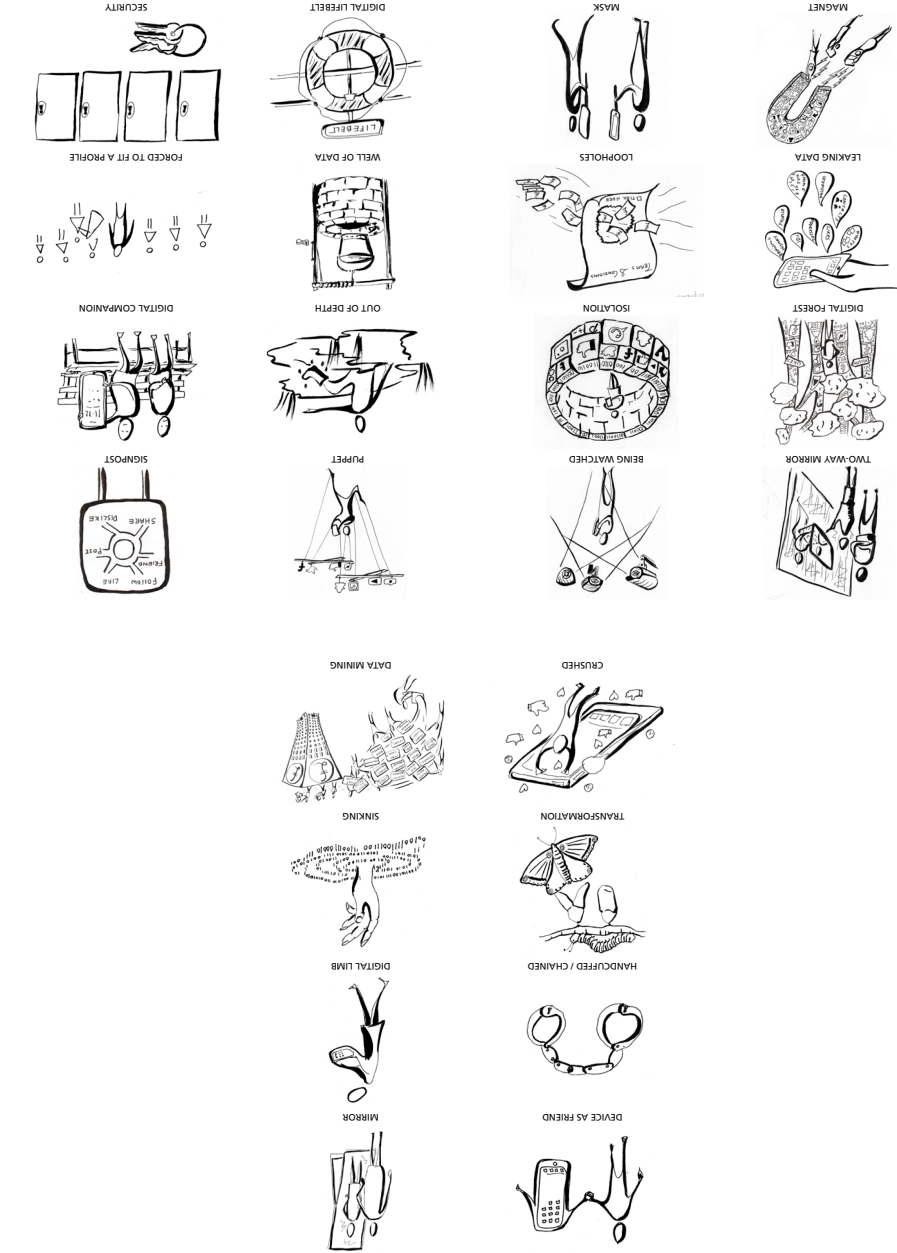
STAKEHOLDER METAMAP

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.



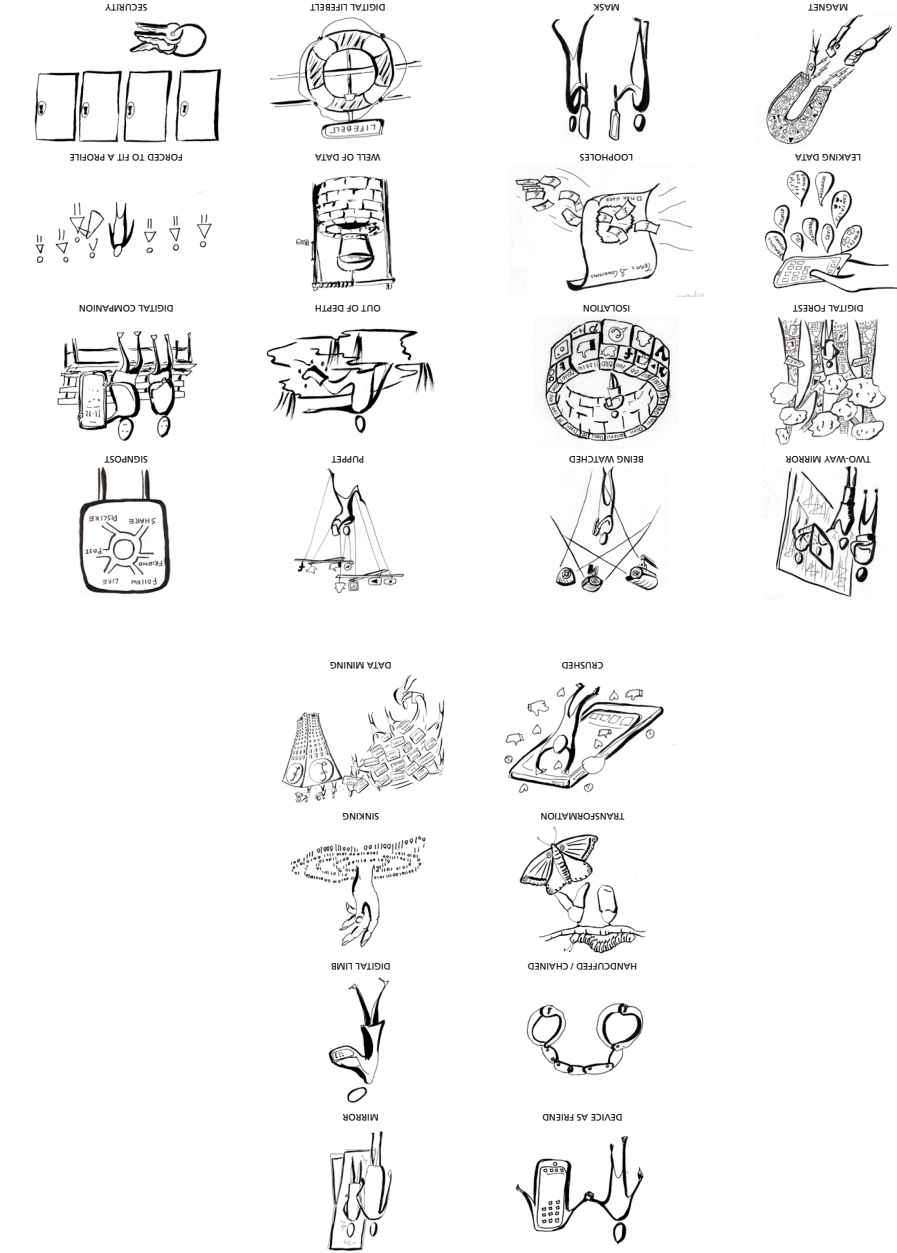
STAKEHOLDER METAMAP




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.



 #TRUSTSCAPE	unbias.wp.horizon.ac.uk	 This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License	 Cover & Design: proboscis.org.uk
<p>When completed, take a photo of this sheet & email it to: unbias@cs.ox.ac.uk</p> <p>Use this worksheet to participate and illustrate your experiences and concerns about Ecosystems. Think & Draw in using online systems. Use the word and image stickers (or draw your own pictures) and write in descriptions in the boxes below.</p>		<p>HOW DO YOU THINK THESE ISSUES ARE BEING ADDRESSED BY COMPANIES AND AUTHORITIES?</p>	
<p>DESCRIBE AN EXPERIENCE OF ONLINE BIAS, UNFAIRNESS OR UNTRUSTWORTHINESS YOU HAVE HAD OR ARE CONCERNED ABOUT</p>		<p>IDEALLY, WHAT WOULD YOU LIKE TO SEE DONE?</p>	
<p>ILLUSTRATE WHAT IS IMPORTANT TO YOU ABOUT THIS EXPERIENCE</p>			

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.: 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.

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.

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 affect young people today.

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.

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 need it.

How to Share

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

How to Use

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.: 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.

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.

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 affect young people today.

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.

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 need it.

How to Share

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