Who am I (why am I here, and why do I care about responsible tech)?

Tanya Johnson (She/Her) · 1st

Chief Product Officer at Auror

Today I want to talk about

- Where we're at with technology
- Why this is so hard to get right
- How we can build products that don't harm vulnerable people
- The way we approach this at Auror, and the things you can take action on

Regrettable tech

Racially biased optic sensors

Chukwuemeka Afigbo
@nke_ise · Follow

If you have ever had a problem grasping the importance of diversity in tech and its impact on society, watch this video

Continue watching on Twitter

0:15 / 0:45

9:48 PM · Aug 16, 2017

Amazon's sexist AI recruiting tool scrapped

Apple Sued Over Racial Bias Of Apple Watch

AI tool turns an Asian MIT student white for a professional headshot

Courtesy of Rona Wang

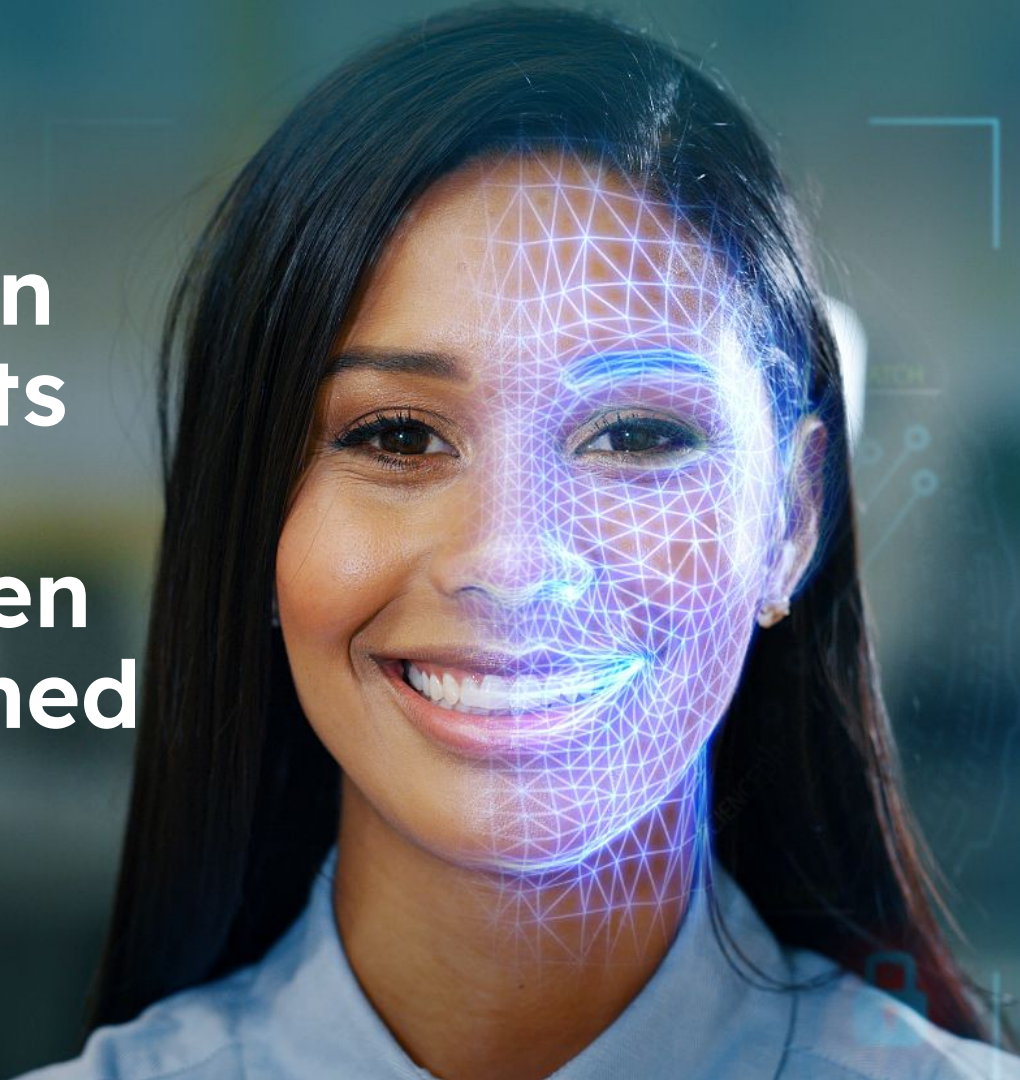# Speech recognition software leading to new forms of discrimination

Gender diverse people missing cervical screenings due to health software setup

Most digital assistants don't respond well to questions about violence and health

Many facial-recognition software products not accurately identifying women and darker-skinned people

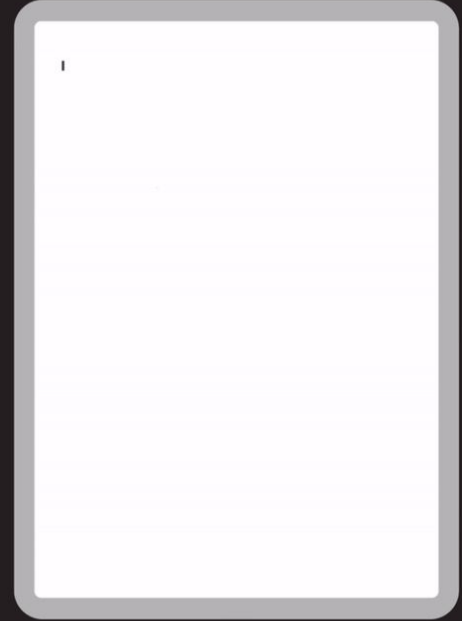# Why is this so hard to get right?

So how can we build products that don't harm vulnerable people?

Auror empowers the retail community to **report crime, reduce loss and harm,** and **make stores safer.**

424

violent or abusive incidents per day
(+9% on the previous year)*

$150B

lost to **customer theft****

4x

more likely a repeat offender will be
aggressive and cause a safety event

*Source ANZ crime study 2022 ($4.2b /365 days)   **US Chamber of Commerce 2023 - March 2023

Denver

London

Melbourne

Auckland

150-200%

Increase in fuel retail loss

You have three seconds

Pick a person you think is going to steal fuel from your petrol station

I immediately regret this decision.

Reducing crime *and* bias

License plate accuracy **HIGH**

ABC123

Monday 21 January 2019 11:05am
Pump 8, Camera 1

Summary VEHICLE OF INTEREST AGGRESSIVE

3	1	1	1	2
Drive offs	Customer charge error	Pay at pump error	Unable to pay	Mobile app payment error

View details   Add event

Balance owed $257.52

Responsible tech and AI principles

⚖ Fairness and equality

◉ Transparency, explainability and accountability

🔒 Reliability, security and privacy
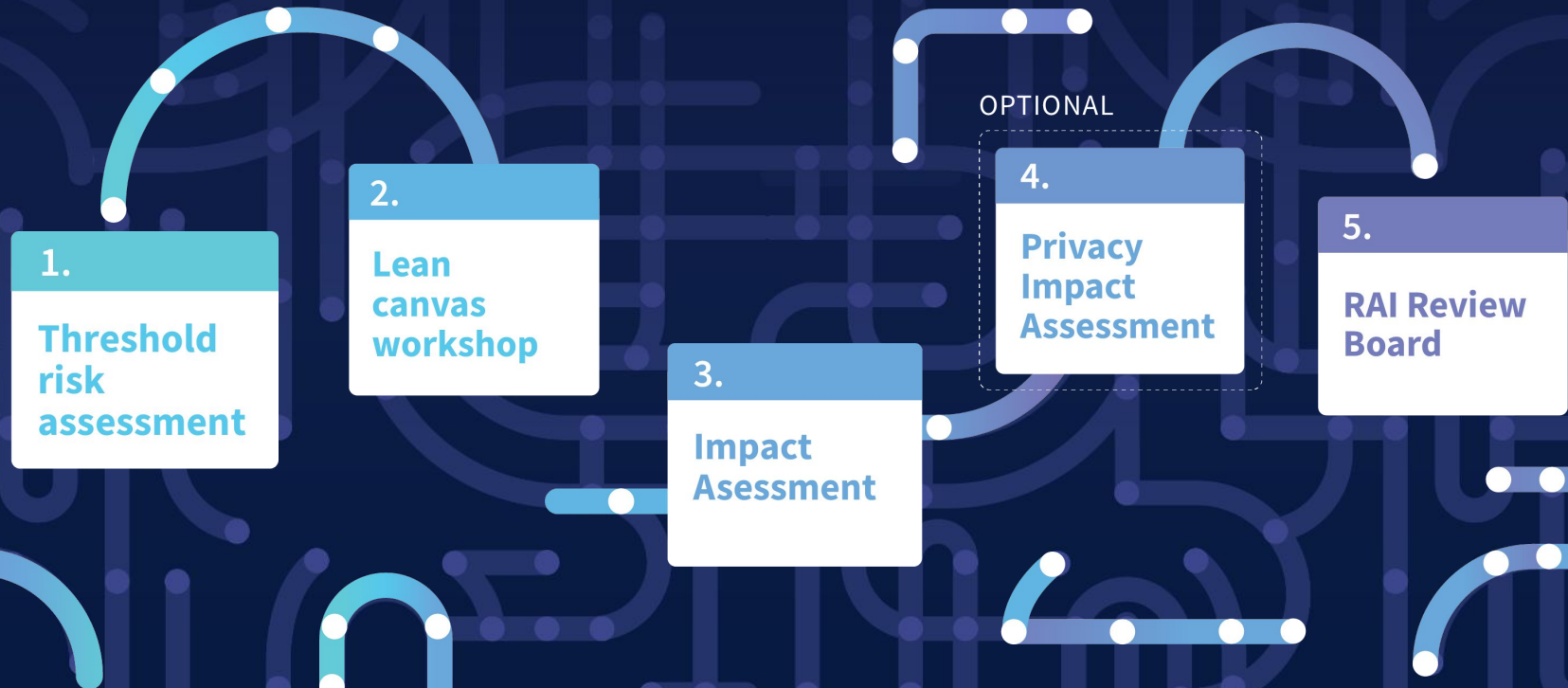
🧑‍💻 Human control and oversight

🤲 Community benefit

Responsible tech and AI process

1.
Threshold risk assessment

2.
Lean canvas workshop

3.
Impact Asessment

OPTIONAL

4.
Privacy Impact Assessment

5.
RAI Review Board

Responsible tech workshop

Problem 1
🤔
What is the problem we're trying to solve, or the opportunity we see?

People 2
👨‍👩‍👧‍👦
Who would use it? Who would be impacted by it? How would it be used by each? Do we need to account for arelationship of care with this solution? Eg. parent, teacher, lawyer, fiduciary responsibility.

Solution Options 3
🤖
What is the idea? How will it solve the problem? What type of technology might be used in this solution?

Value Proposition 4
⭐
Who does this help? How does it help our partners / users?

Commercial Opportunities 5
💵
How could we commercialise this?

Positive Impacts 6
😍
Describe the positive impacts this solution might have on Auror, our partners / users, or the wider community

Negative Impacts and Potential Risks 7
⚠️
What negative impacts, harms, or risks might we encounter with this solution and what vulnerable communities may be impacted directly or indirectly by this solution?

Mitigations 8
🚦
What are the mitigations we could take to reduce negative impacts, harm, or risk?

Impact assessment (IA)

- **High level information about the feature/change**

- **Ethical considerations**

  - Fairness & equality
  - Transparency, explainability and accountability
  - Reliability, security and privacy
  - Human control and oversight
  - Community benefit

- **Impact and harm analysis**

4. Impact & harm analysis

Consider the best and worst-case scenarios that could arise from use of the proposed system after it has been deployed. Detail any expected hurdles or challenges to overcome.

When thinking about affected stakeholders, consider both direct stakeholders (such as customers, users, and partners) and indirect stakeholders (such as data subjects, marginalised groups, regulators, civil society and the general public).

Use the risk assessment table to indicate the expected likelihood and severity of each harm once the proposed mitigations are in place.

RESPONSIBLE AI RISK MATRIX

		Impact (of harm to affected person/group)		
		LOW Minor impact	**MED** Some impacts felt	**HIGH** Significant impact
Likelihood (of harm occurring)	**LOW** Unlikely to occur in most circumstances	**LOW** Unlikely to happen/ low frequency with minimal impact	**LOW** Unlikely to happen but moderate harm will occur	**MED** Low frequency but would cause significant harm
	MED Could occur in some circumstances	**LOW** Could happen but minimal impact	**MED** Could happen with moderate harm occurring	**HIGH** Could happen causing significant harm
	HIGH High chance of occurring in many circumstances	**MED** High frequency but low impact	**HIGH** High frequency with moderate harm occurring	**HIGH** High frequency causing significant harm

RISK DECISION-MAKING KEY

LOW Risk manageable through existing or proposed internal controls & mitigants	**MED** Requires additional risk treatment actions	**HIGH** Unacceptable
OK to proceed	Proceed only once appropriate mitigants are in place	Place project on hold pending investigation of further potential mitigants

Designed and developed by Auror Limited. Auror's Responsible Tech and AI Impact Assessment is licensed for your use under a Creative Commons Attribution 4.0 International License

Impact assessment (IA)

- **Data questions (for both training and production data)**

    - Where does the data come from (i.e. what is the source? Which country does it originate from?

    - What and who is represented in the data? What and who is not represented?

    - When and how was it collected? By who? And for what purpose?

1. Project information

This section of the IA requests background information for readers who may not be familiar with your project . It includes high-level context questions to help inform our thinking on legal and ethical considerations and potential harms in later sections of the document.

2. Data

A significant number of the questions in the IA are about data. Data powers many technologies and AI systems so we need to understand what training and production data we will be using and how accurate and reliable it is. You can use the table below to collect information you'll then use when completing the IA.

	Training data	Production data
Where does the data come from (i.e. what is the source? Which country does it originate from?		
What and who is represented in the data? What and who is not represented?		
When was it collected?		
How was it collected?		
Who collected it?		
Why/for what purpose was it collected?		

Designed and developed by Auror Limited. Auror's Responsible Tech and AI Impact Assessment is licensed for your use under a Creative Commons Attribution 4.0 International License

Review and go/no-go

So what were the key things we learned

Diverse perspectives drive innovation

Ethical considerations are non-negotiable

Continuous learning and adaptation

Now, take action

auror.co/responsible-ai

Responsible Tech and AI Impact Assessment (IA) guide

Problem
What is the problem we're trying to solve, or the opportunity we see?

People
Who would use it? Who would be impacted by it? How would it be used by each? Do we need to account for a relationship of care with this solution? Eg. parent, teacher, lawyer, fiduciary responsibility.

Solution Options
What is the idea? How will it solve the problem? What type of technology might be used in this solution?

Negative Impacts and potential risks
What negative impacts, harms, or risks might we encounter with this solution and what vulnerable communities may be impacted directly or indirectly by this solution?

Miti...
What...
negati...

Responsible Tech and AI Lean Canvas

Auror.

Further reading

Try

- Foundations of Humane Technology course
- Ethical OS - Risk mitigation checklist
- Consequence Scanning - an agile practice for innovators

Read

- Shouting Zeros and Ones: Digital Technology, Ethics, and Policy in New Zealand
- Data protection laws across the world
- New Zealand's Privacy Act 2020 and Privacy Principles
- The Ethical Algorithm: The Science of Socially Aware Algorithm Design
- Weapons of Math Destruction

- Hello World: Being Human in the Age of Algorithms
- The Alignment Problem: Machine Learning and Human Values
- Invisible Women: Data Bias in a World Designed for Men
- Queer Data: Using Gender, Sex and Sexuality Data for Action
- Algorithms of Oppression: How Search Engines Reinforce Racism
- Data Feminism

Thank you!

P.S. We're hiring at Auror

auror.co/careers