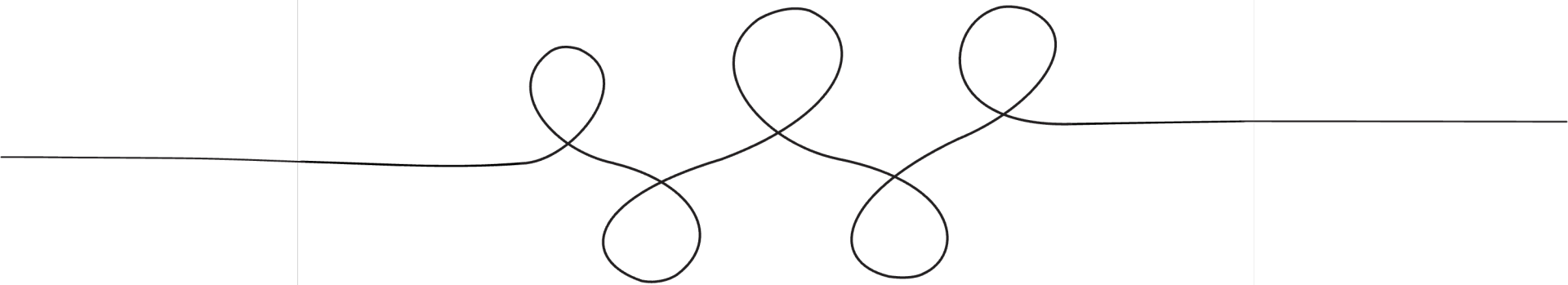
Atlas of AI Risks

Enhancing Public Understanding of AI Risks

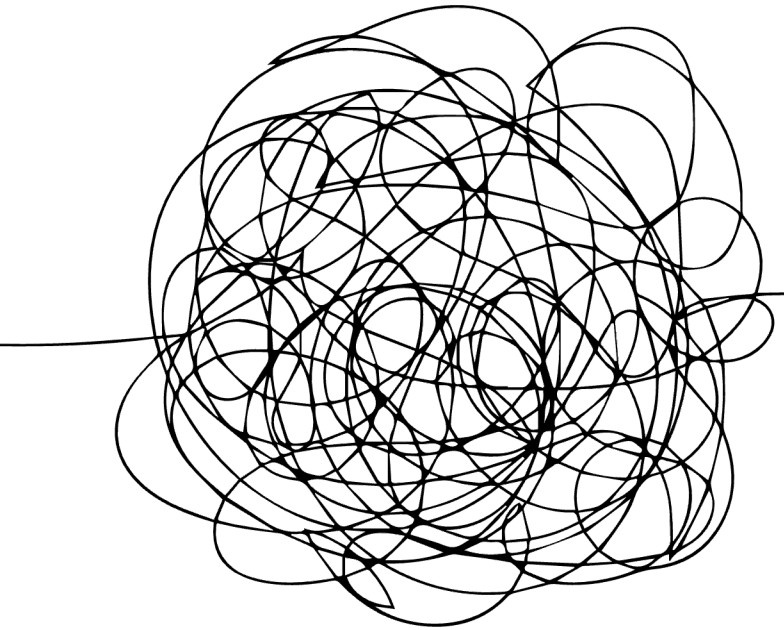
Edyta Bogucka, Sanja Šćepanović, Daniele Quercia

NOKIA BELL LABS
100 years innovating

Dealing with well-defined problems

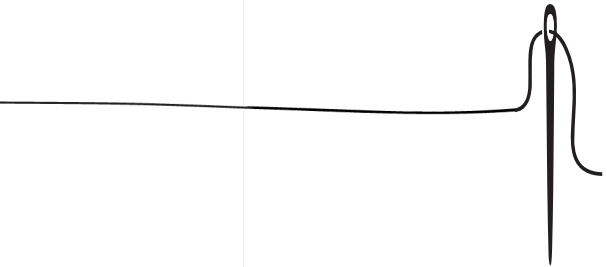


NOKIA BELL LABS



What about wicked problems?

NOKIA BELL LABS



Methodology to make broad public reflect on wicked problems

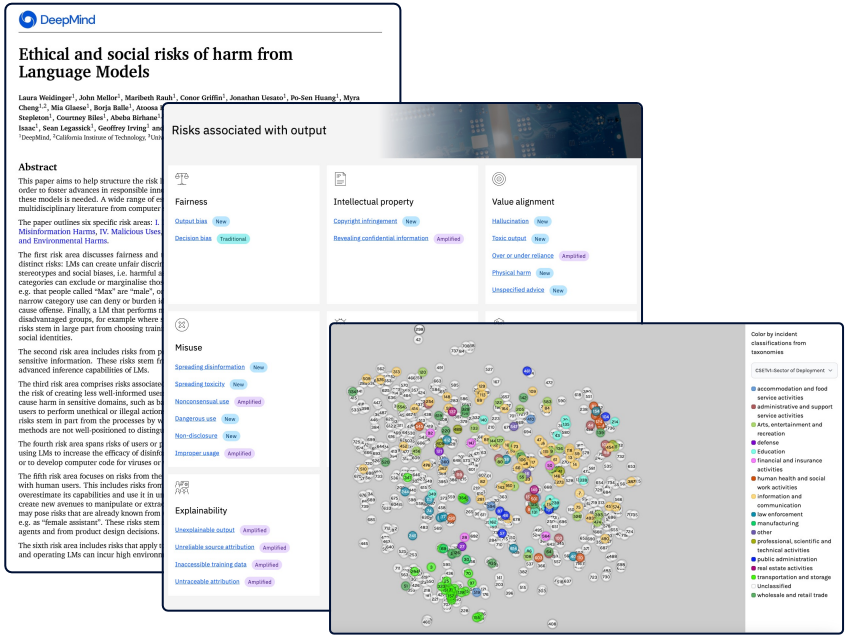
NOKIA BELL LABS

Two problems in building awareness tools about AI

sampling biases in facial recognition

security risks of chatbots

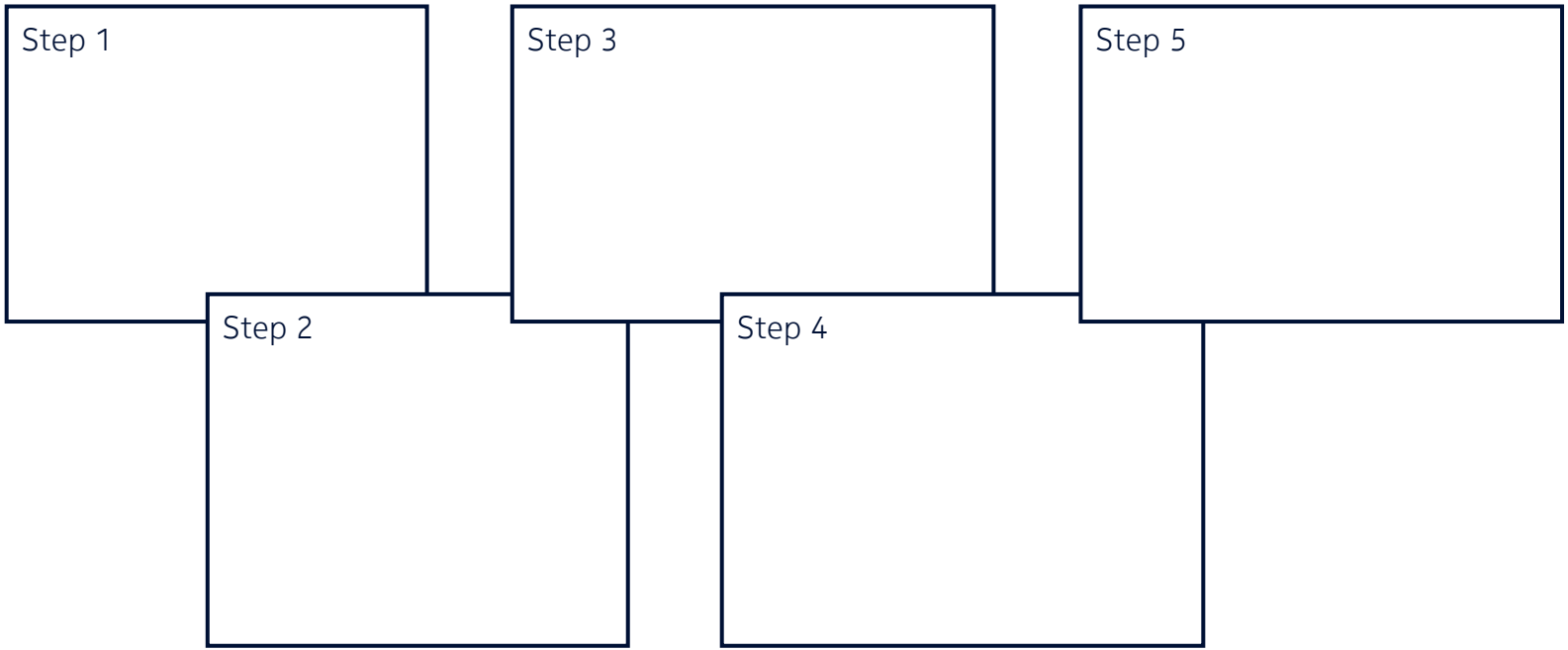
model vulnerabilities in image generators



Overfocusing on technological risks…

…and visualizing them for tech-savvy individuals

How did we do it?



How did we do it?

Step 1

Crowdsourcing design requirements for the tool
(N=40)

How did we do it?

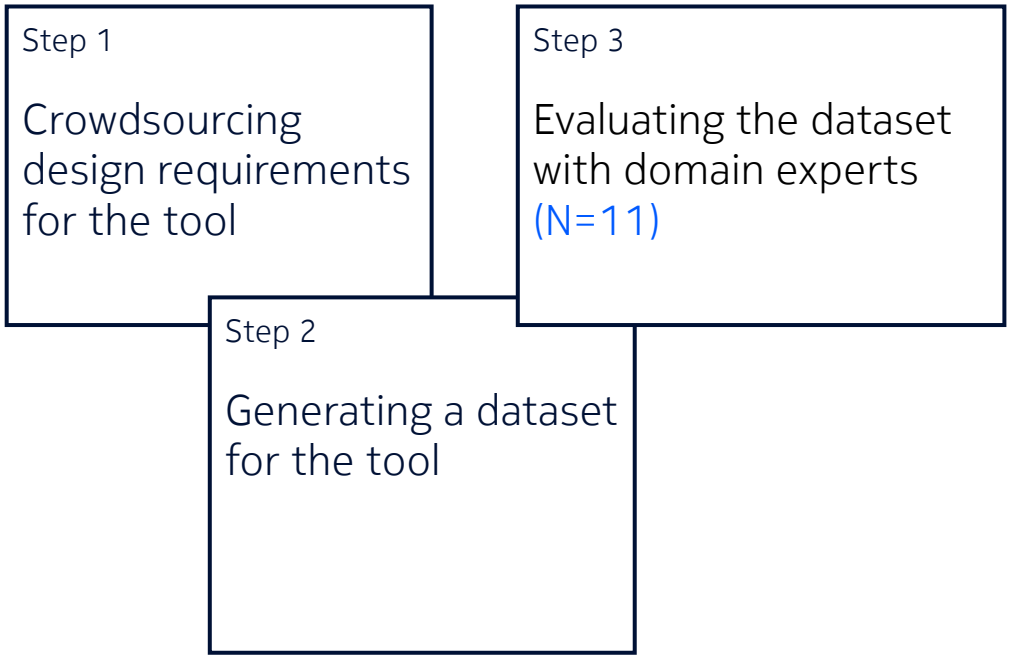
Step 1

Crowdsourcing design requirements for the tool

Step 2

Generating a dataset for the tool

How did we do it?



How did we do it?

Step 1

Crowdsourcing design requirements for the tool

Step 2

Generating a dataset for the tool

Step 3

Evaluating the dataset with domain experts

Step 4

Applying data visualization strategies to the dataset to build the tool

NOKIA BELL LABS

How did we do it?

Step 1

Crowdsourcing design requirements for the tool

Step 2

Generating a dataset for the tool

Step 3

Evaluating the dataset with domain experts

Step 4

Applying data visualization strategies to the dataset to build the tool

Step 5

Evaluating the tool with members of the broader public (N=140)

Step 1: Crowdsourcing design requirements for the tool

THE PANOPTICON | OCT. 12, 2018

Here Is a List of Every Animal Humans Currently Monitor Using

[redacted]

By Mack DeGeurin

NOKIA BELL LABS

Step 1: Crowdsourcing design requirements for the tool

THE PANOPTICON | OCT. 12, 2018

Here Is a List of Every Animal Humans Currently Monitor Using

Facial Recognition Technology

By Mack DeGeurin

NOKIA BELL LABS

Step 1: Crowdsourcing design requirements for the tool

AI and Ethics (2021) 1:159–172
https://doi.org/10.1007/s43681-020-00014-3

ORIGINAL RESEARCH

Smile, you are being identified! Risks and measures for the use of facial recognition in (semi-)public spaces

Thiago Guimarães Moraes[1] · Eduarda Costa Almeida[1] · José Renato Laranjeira de Pereira[1]

Received: 27 July 2020 / Accepted: 10 September 2020 / Published online: 10 October 2020

Abstract
This article analyses the use of facial recognition technology (FRT) in (semi-)public spaces with a focus in the Brazilian context. Therefore, the operation of the FRT processing chain is addressed, as well as the juridical nature of the facial signature, focusing mainly in the Brazilian data protection framework. FRT has been used in everyday life for several purposes, such as security, digital ranking, targeted marketing and health protection. However, the indiscriminate use of FRT poses high risks to privacy and data protection. In this perspective, to avoid harms such as inaccuracy, normalisation of cyber-surveillance and lack of transparency, safeguards were identified to guarantee individual rights, such as soft law, oversight, international standards and regulatory sandboxes.

Keywords Facial recognition · Surveillance · Data protection · Safeguard · Sandbox

1 Introduction

Initially restricted to physical access control systems in chemical/radioactive laboratories, facial recognition technology (FRT) is being increasingly applied to identify individuals on *web* pages, photos, video recordings and in physical spaces. This has raised concern about the right to privacy of individuals being identified: who is surveilling? In what context? For which purposes? These questions are even more sensitive when facial recognition is used in (semi-) public spaces indiscriminately, without the establishment of a proper criteria to filter which personal data will be collected, and from whom.

Semi-public spaces are characterized by being freely accessible and having few usage restrictions. According to Peterson, these are places which, although belonging to private entities, are freely used in a shared way by different social groups [1]. Examples are shopping centres, supermarkets and libraries. Their protection should take account of private security regulations, while stricto sensu public spaces are protected by the legitimate interest of public security.[1]

Around the world, several cases involving the deployment of facial recognition systems have come to the attention of digital rights organizations and the general public. In the UK, the London Metropolitan Police—MET used two facial recognition cameras in one of the most crowded sites in London, the *King's Cross Central* [2]. The experiment lasted months, and the authorities had no concern to establish transparency and information mechanisms to passersby who had their data collected.

In the Brazilian context, facial recognition has already been used in carnival blocks in Rio de Janeiro and Salvador [3] and in a "smart/safe city" project in Campinas [4]. In June 2019, the Metropolitan Company of São Paulo opened a procurement for the implementation of FRT in three metro lines [5].

So far, the main purpose for the deployment of such technology in Brazil has been security [4], as facial recognition helps in the identification of individuals who have committed crimes or are about to commit it. Security is also a concern for shopping centres, supermarkets and other spaces, as

✉ Eduarda Costa Almeida
eduarda@lapin.org.br
Thiago Guimarães Moraes
thiago@lapin.org.br
José Renato Laranjeira de Pereira
joserenato@lapin.org.br

[1] Public Policy Division, LAPIN - Laboratory of Public Policy and Internet, Brasília, DF, Brazil

[1] In this article, we use the sentence "(semi-)public" to represent the two types of spaces, one of public nature and other of a private nature. Where it is necessary to distinguish between them, we use the term sensu stricto to identify the former.

AI and Ethics
https://doi.org/10.1007/s43681-023-00344-y

ORIGINAL RESEARCH

Socially responsible facial recognition of animals

Fred S. Roberts[1]

Received: 29 April 2023 / Accepted: 4 September 2023

Abstract
Automated identification of people using facial recognition algorithms, while of widespread potential use, has been criticized for being biased, unfair, discriminatory, or potentially harmful. Facial recognition algorithms to identify individual domesticated and wild non-human animals are increasingly used but there has been much less discussion of their potential dangers. This paper explores the ways in which such algorithms are used in farming and conservation, and discusses potential issues in such uses.

Keywords Facial recognition · Socially responsible algorithms · Precision livestock farming · Factory farming · Animal ethics · Wildlife populations · Camera traps

Abbreviations

ASF	African swine fever
BSE	Bovine spongiform encephalopathy
FACS	Facial action coding systems
FAUs	Facial action units
LINC	Lion identification network of collaborators
NIST	U.S. National Institute of Standards and Technology
NOAA	U.S. National Oceanic and Atmospheric Administration
PLF	Precision livestock farming

1 Introduction

Facial recognition algorithms show tremendous promise in applications, such as policing, medicine, and commerce [1]. However, automated identification of people using such algorithms has been shown to be biased, unfair, discriminatory, or potentially harmful [1], and these considerations have led to an emphasis on social responsibility of algorithms involving facial recognition of people (see, for example, [2]). Facial recognition algorithms are increasingly used with both domesticated and wild non-human animals (hereafter just referred to as animals), to aid in more efficient farming and in conservation of wild populations. However, there has been much less discussion of the potential dangers of using facial recognition algorithms for animals. This paper explores the applications of such algorithms and the social responsibility issues that arise.

Issues with facial recognition of humans have been well-documented. For instance, U.S. government tests find even top-performing facial recognition systems misidentify blacks at rates five to 10 times higher than they do whites [3]. The French company Idemia's algorithms scan millions of faces in uses by police in the US, Australia, and France, but a U.S. National Institute of Standards and Technology study showed that two of Idemia's algorithms were significantly more likely to mix up black women's faces than those of white women, or black or white men [3]. Buolamwini and Gebru [4] showed that in three commercial gender classification systems, darker-skinned females were misclassified with error rates up to 34.7%, where lighter-skinned males had a maximum error rate of 0.8%.

Amazon's "Rekognition" mistakenly identified 28 members of the U.S. Congress (disproportionately people of color) as criminals [5, 6]. Leslie [7] describes a variety of examples, where use of facial recognition algorithms has led to problems, e.g., in faulty face recognition algorithms leading to arrests or denial of passport photos for dark-skinned people. As Cavazos, et al. [8] observe, "Nearly all of the face recognition algorithms studied over the past 30 years show some performance differences as a function of the race of the face." A U.S. National Institute of Standards and Technology (NIST) study [9] tested 189 face recognition algorithms and found a wide range of accuracy. It found that for "one-to-one

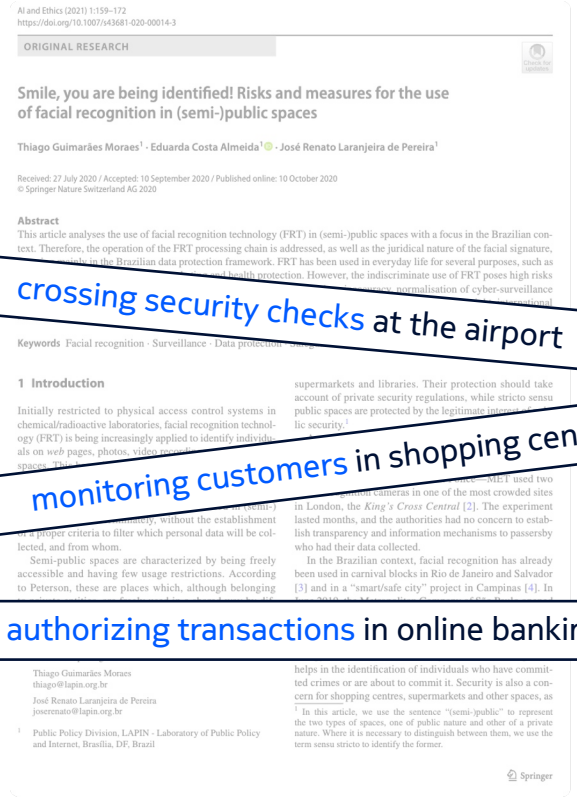
✉ Fred S. Roberts
froberts@dimacs.rutgers.edu

[1] DIMACS Center, Rutgers University, Piscataway, NJ, USA

Published online: 28 September 2023

NOKIA BELL LABS

Step 1: Crowdsourcing design requirements for the tool



AI and Ethics (2021) 1:159–172
https://doi.org/10.1007/s43681-020-00014-3

ORIGINAL RESEARCH

Smile, you are being identified! Risks and measures for the use of facial recognition in (semi-)public spaces

Thiago Guimarães Moraes[1] · Eduarda Costa Almeida[1] · José Renato Laranjeira de Pereira[1]

Received: 27 July 2020 / Accepted: 10 September 2020 / Published online: 10 October 2020
© Springer Nature Switzerland AG 2020

Abstract
This article analyses the use of facial recognition technology (FRT) in (semi-)public spaces with a focus in the Brazilian context. Therefore, the operation of the FRT processing chain is addressed, as well as the juridical nature of the facial signature, [illegible] in the Brazilian data protection framework. FRT has been used in everyday life for several purposes, such as [illegible] and health protection. However, the indiscriminate use of FRT poses high risks [illegible] normalisation of cyber-surveillance [illegible] informational

Keywords Facial recognition · Surveillance · Data protection · [illegible]

1 Introduction

Initially restricted to physical access control systems in chemical/radioactive laboratories, facial recognition technology (FRT) is being increasingly applied to identify individuals on *web* pages, photos, video records [illegible] spaces. This [illegible] [illegible] (semi-) [illegible] without the establishment of a proper criteria to filter which personal data will be collected, and from whom.

Semi-public spaces are characterized by being freely accessible and having few usage restrictions. According to Peterson, these are places which, although belonging

supermarkets and libraries. Their protection should take account of private security regulations, while stricto sensu public spaces are protected by the legitimate interest of [illegible] lic security.[1]

[illegible] cameras in one of the most crowded sites in London, the *King's Cross Central* [2]. The experiment lasted months, and the authorities had no concern to establish transparency and information mechanisms to passersby who had their data collected.

In the Brazilian context, facial recognition has already been used in carnival blocks in Rio de Janeiro and Salvador [3] and in a "smart/safe city" project in Campinas [4]. In [illegible] helps in the identification of individuals who have committed crimes or are about to commit it. Security is also a concern for shopping centres, supermarkets and other spaces, as

[1] In this article, we use the sentence "(semi-)public" to represent the two types of spaces, one of public nature and other of a private nature. Where it is necessary to distinguish between them, we use the term sensu stricto to identify the former.

Thiago Guimarães Moraes
thiago@lapin.org.br

José Renato Laranjeira de Pereira
joserenato@lapin.org.br

[1] Public Policy Division, LAPIN - Laboratory of Public Policy and Internet, Brasília, DF, Brazil

Springer

crossing security checks at the airport

monitoring customers in shopping centers

authorizing transactions in online banking

AI and Ethics
https://doi.org/10.1007/s43681-023-00344-y

ORIGINAL RESEARCH

Socially responsible facial recognition of animals

Fred S. Roberts[1]

Received: 29 April 2023 / Accepted: 4 September 2023
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

Abstract
Automated identification of people using facial recognition algorithms, while of widespread potential use, has been criticized for being biased, unfair, discriminatory, or potentially harmful. Facial recognition algorithms to identify individual domesticated and wild non-human animals are increasingly used but there has been much less discussion of their potential dangers. This paper explores the ways in which such algorithms are used in farming and conservation, and discusses potential [illegible] such uses.

[illegible] Factory farming · Animal [illegible]

Abbreviations

ASF	African swine fever
BSE	Bovine spongiform encephalopathy
FACS	Facial action coding systems
FAUs	Facial action units
LINC	Lion identification network of collaborators
NIST	U.S. National Institute of Standards and Technology
NOAA	U.S. National Oceanic and Atmospheric Administration
PLF	[illegible]

paper explores the applications of such algorithms [illegible] social responsibility issues that arise.

Issues with facial recognition of humans have been well-documented. For instance, U.S. government tests find even top-performing facial recognition systems misidentify blacks at rates five to 10 times higher than they do whites [3]. The French company Idemia's algorithms scan millions of faces [illegible] Buolamwini and Gebru [4] showed that in three commercial gender classification systems, darker-skinned females were misclassified with error rates up to 34.7%, where lighter-skinned males had a maximum error rate of 0.8%.

Amazon's "Rekognition" mistakenly identified 28 members of the U.S. Congress (disproportionately people of color) as criminals [5, 6]. Leslie [7] describes a variety of examples, where use of facial recognition algorithms has led to problems, e.g., in faulty face recognition algorithms leading to arrests or denial of passport photos for dark-skinned people. As Cavazos, et al. [8] observe, "Nearly all of the face recognition algorithms studied over the past 30 years show some performance differences as a function of the race of the face." A U.S. National Institute of Standards and Technology (NIST) study [9] tested 189 face recognition algorithms and found a wide range of accuracy. It found that for "one-to-one

Facial recognition algorithms show tremendous promise in applications, such as policing, medicine, and commerce [1]. However, automated identification of people using such algorithms has been shown to be biased, unfair, discriminatory, or potentially harmful [1], and these considerations have led to an emphasis on social responsibility of algorithms involving facial recognition of people (see, for example, [2]). Facial recognition algorithms are increasingly used with both domesticated and wild non-human animals (hereafter just referred to as animals), to aid in more efficient farming and in conservation of wild populations. However,

Fred S. Roberts
froberts@dimacs.rutgers.edu

[1] DIMACS Center, Rutgers University, Piscataway, NJ, USA

Published online: 28 September 2023

Springer

identifying polar bears based on whisker spot patterns

tracking migration of humpback whales

NOKIA BELL LABS

Step 1: Crowdsourcing design requirements for the tool

40 individuals from the general public

matching US census in sex and ethnicity

contacted on Prolific

Task 1

Write an email to regulators in which you list specific uses of facial recognition and request either their ban or further adoption

Task 2

How can we best present the risks of 30 facial recognition uses in an interactive tool?

NOKIA BELL LABS

Step 1: Crowdsourcing design requirements for the tool

Requirement 1

Multiple uses

Learn about a variety of uses

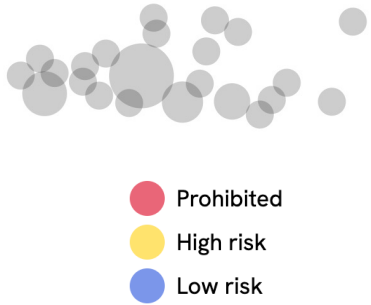
NOKIA BELL LABS

Step 1: Crowdsourcing design requirements for the tool

Requirement 1

Multiple uses

Learn about a variety of uses



Requirement 2

Structured uses

Categorize uses for better understanding

Step 1: Crowdsourcing design requirements for the tool

Requirement 1

Multiple uses

Learn about a variety of uses

Requirement 2

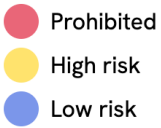
Structured uses

Categorize uses for better understanding

Requirement 3

Balanced assessment of uses

Present each use not only with its risks but also benefits and mitigation strategies



BENEFITS IF USED RESPONSIBLY

→ Enhances efficiency of welfare case evaluations, improving service delivery

NOKIA BELL LABS

Step 1: Crowdsourcing design requirements for the tool

Requirement 1

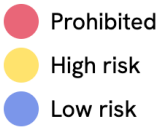
Multiple uses

Learn about a variety of uses

Requirement 2

Structured uses

Categorize uses for better understanding



Requirement 3

Balanced assessment of uses

Present each use not only with its risks
but also benefits and mitigation strategies

Requirement 4

Broad appeal

Make the uses, risks, benefits, and mitigations relevant to members of the broader public

NOKIA BELL LABS

Step 1: Crowdsourcing design requirements for the tool

Requirement 1

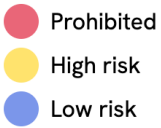
Multiple uses

Learn about a variety of uses

Requirement 2

Structured uses

Categorize uses for better understanding



Requirement 3

Balanced assessment of uses

Present each use not only with its risks but also benefits and mitigation strategies

Requirement 4

Broad appeal

Make the uses, risks, benefits, and mitigations relevant to members of the broader public

Requirement 5

Engaging exploration

Engage users in exploring the uses, risks, benefits, and mitigation strategies

NOKIA BELL LABS

Dataset generated through crowdsourcing (N=23)

High-risk AI is on everyone's radar

3 (13%)
low risk uses

20 (87%)
high risk uses

18 risks
8 benefits
9 mitigations

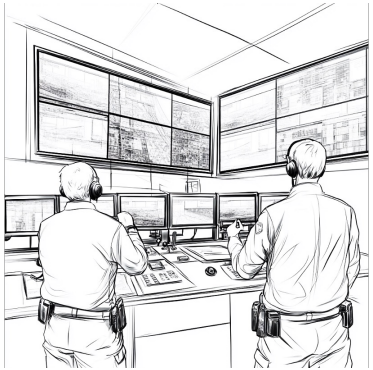
Dataset generated through crowdsourcing (N=23)

People narrow AI's role to law enforcement

Identifying suspects for crime prevention

High risk

N=5



Apprehending individuals on the run

High risk

N=4

15 (65%)
high risk uses

18 risks
8 benefits
9 mitigations

Dataset generated through crowdsourcing (N=23)

Few see AI's positive applications

Unlocking devices

Low risk

N=8

3 (13%)
low risk uses

18 risks
8 benefits
9 mitigations

NOKIA BELL LABS

Step 2: Generating uses

Prompt 1

Use generation

Domain: Urban planning

AI User: Urban planners

AI Subject: Citizens

Capability: Counting faces in public spaces…

Purpose: …for…

Monitoring pedestrian activity for city planning

Herdel et al. 2024, Constantinides et al. 2024, Bogucka et al. 2024

Step 2: Identifying the risks

Prompt 2

Risk assessment

The use is **high risk** because it can be applied for biometric identification in public spaces by a public authority

Monitoring pedestrian activity for city planning

High risk

Herdel et al. 2024, Constantinides et al. 2024, Bogucka et al. 2024

NOKIA BELL LABS

Step 2: Identifying the benefits

Prompt 3

Benefit assessment

The system can advance **sustainable communities** (SDG 11) and help identify **new areas for public services** (SDG 9)

Monitoring pedestrian activity for city planning

High risk

Herdel et al. 2024, Constantinides et al. 2024, Bogucka et al. 2024

NOKIA BELL LABS

Step 2: Identifying mitigations understandable to all

regardless of technical knowledge

Prompt 4

Mitigation generation

→ Teach the AI using a wide range of examples

→ Regularly check how data is being managed to ensure it's safe and secure

→ Include a way for people to give feedback so the system can get better over time

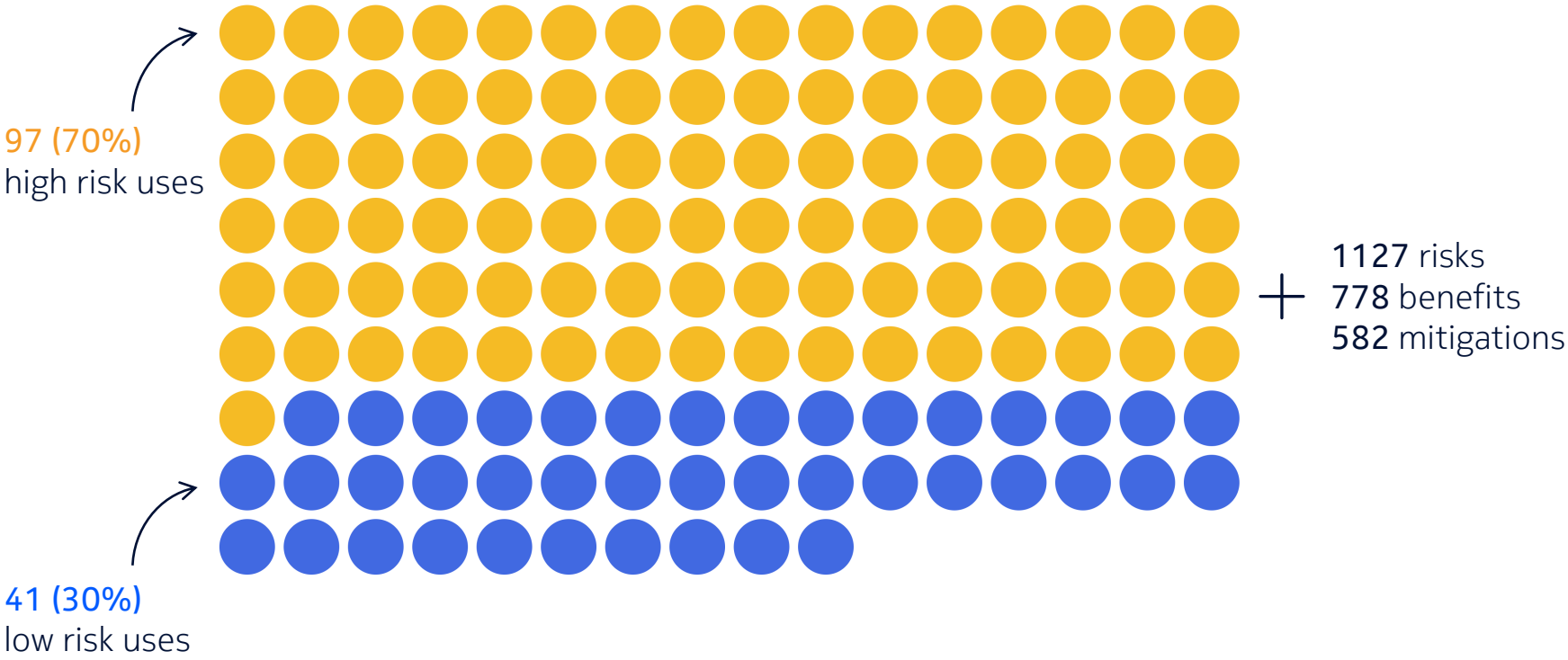
Monitoring pedestrian activity for city planning

High risk

Herdel et al. 2024, Constantinides et al. 2024, Bogucka et al. 2024

NOKIA BELL LABS

Dataset generated by the LLM



NOKIA BELL LABS

Step 3: Evaluating the dataset with domain experts

Study with 3 AI compliance experts

Use #67

The AI system intended to be used by border control officers and immigration officials for identity verification by matching faces to passport or ID photos.

High risk use

Justification:
High Risk due to its use in sensitive identity verification in migration and border control, as specified in EU AI Act Annex III, 7(da).

→ EU AI Act

Q1 How probable do you find this use? ⓘ

- ◯ Existing
- ◯ Upcoming
- ◯ Unrealistic

Q2 Do you agree with the use risk classification? ⓘ

- ◯ Yes
- ◯ No

Q3 Do you agree with the use risk justification?

- ◯ Yes
- ◯ No

Q4 Please explain your reasoning about the use risk classification and justification.

NOKIA BELL LABS

Step 3: Evaluating the dataset with domain experts

Study with 3 AI compliance experts

Use #67

The AI system intended to be used by border control officers and immigration officials for identity verification by matching faces to passport or ID photos.

High risk use

Justification:
High Risk due to its use in sensitive identity verification in migration and border control, as specified in EU AI Act Annex III, 7(da).

Q1 How probable do you find this use? ⓘ

- ◯ Existing
- ◯ Upcoming
- ◯ Unrealistic

Q2 Do you agree with the use risk classification? ⓘ

- ◯ Yes
- ◯ No

Q3 Do yo use risk

- ◯ Yes
- ◯ No

Step 3: Evaluating the dataset with domain experts

Study with 3 AI compliance experts

Q3 Do you agree with the use risk justification?

- ◯ Yes
- ◯ No

Q4 Please explain your reasoning about the use risk classification and justification.

NOKIA BELL LABS

Step 3: Evaluating the dataset with domain experts

Study with 8 researchers and developers

Use #106

Monitoring pedestrian traffic for city planning purposes

The AI system with the capability of counting and tracking faces in public spaces, intended to be used by urban planners and impacting city residents.

Q1 Please check all the potential benefits you think are incorrect ⓘ

- ☐ Estimates pedestrian traffic for effective city planning.
- ☐ Identifies areas for new businesses or public services.
- ☐ Promotes gender equality in urban planning decisions.
- ☐ Improves security through better urban planning and resource allocation.
- ☐ Enhances quality of life and privacy in the city.
- ☐ Promotes social, economic, and political inclusion of all city residents.

Q3 Please check all the potential risks and their mitigations below you think are incorrect

- ☐ **R1: Infringes on the privacy of individuals by tracking faces in public spaces.**
  - ☐ Anonymize the data collected.
  - ☐ Ensure the data is used solely for the purpose of city planning.
  - ☐ Implement strict data handling and privacy policies.
- ☐ **R2: Reduces individuals to data points, infringing on their right to be recognized as a person before the law.**
  - ☐ Ensure the data collected is not used to identify individuals.
  - ☐ Use the data solely to understand pedestrian traffic patterns.
  - ☐ Implement strict data handling and privacy policies.

Q4 Please leave any comments about the reasons why you checked these benefits, risks, and/or mitigations.

NOKIA BELL LABS

Step 3: Evaluating the dataset with domain experts

Study with 8 researchers and developers

Use #106

Monitoring pedestrian traffic for city planning purposes

The AI system with the capability of counting and tracking faces in public spaces, intended to be used by urban planners and impacting city residents.

Q1 Please check all the potential benefits you think are incorrect ⓘ

- ☐ Estimates pedestrian traffic for effective city planning.
- ☐ Identifies areas for new businesses or public services.
- ☐ Promotes gender equality in urban planning decisions.
- ☐ Improves security through better urban planning and resource allocation.
- ☐ Enhances quality of life and privacy in the city.

Step 3: Evaluating the dataset with domain experts

Study with 8 researchers and developers

Q3 Please check all the potential risks and their mitigations below you think are incorrect

- [ ] **R1: Infringes on the privacy of individuals by tracking faces in public spaces.**
  - [ ] Anonymize the data collected.
  - [ ] Ensure the data is used solely for the purpose of city planning.
  - [ ] Implement strict data handling and privacy policies.
- [ ] **R2: Reduces individuals to data points, infringing on their right to be recognized as a person before the law.**
  - [ ] Ensure the data collected is not used to identify individuals.
  - [ ] Use the data solely to understand pedestrian traffic patterns.

Q4 Please leave any comments about the reasons why you checked these benefits, risks, and/or mitigations.

NOKIA BELL LABS

Step 3: Evaluating the dataset with domain experts

Good agreement on risk classification (EU AI Act)

91% agreement between experts and LLM classification on

→ widely-adopted uses: e.g., accessing devices

→ already regulated domains: e.g., medical assistance

NOKIA BELL LABS

Step 3: Evaluating the dataset with domain experts

Good agreement on risk classification (EU AI Act)

91% agreement between experts and LLM classification on

→ widely-adopted uses: e.g., accessing devices

→ already regulated domains: e.g., medical assistance

9% disagreement arose when the LLM downplayed risks in:

→ privacy: e.g., location tracking for safety,

→ vulnerable groups: e.g., elderly health monitoring,

→ emerging tech: e.g., VR, with evolving ethical boundaries

Monitoring elderly health and activity

High risk

Step 3: Evaluating the dataset with domain experts

Good agreement on risks, benefits and mitigations

93% correct risks

82% correct benefits

95% correct mitigations

Managing access to climate-sensitive areas

High risk

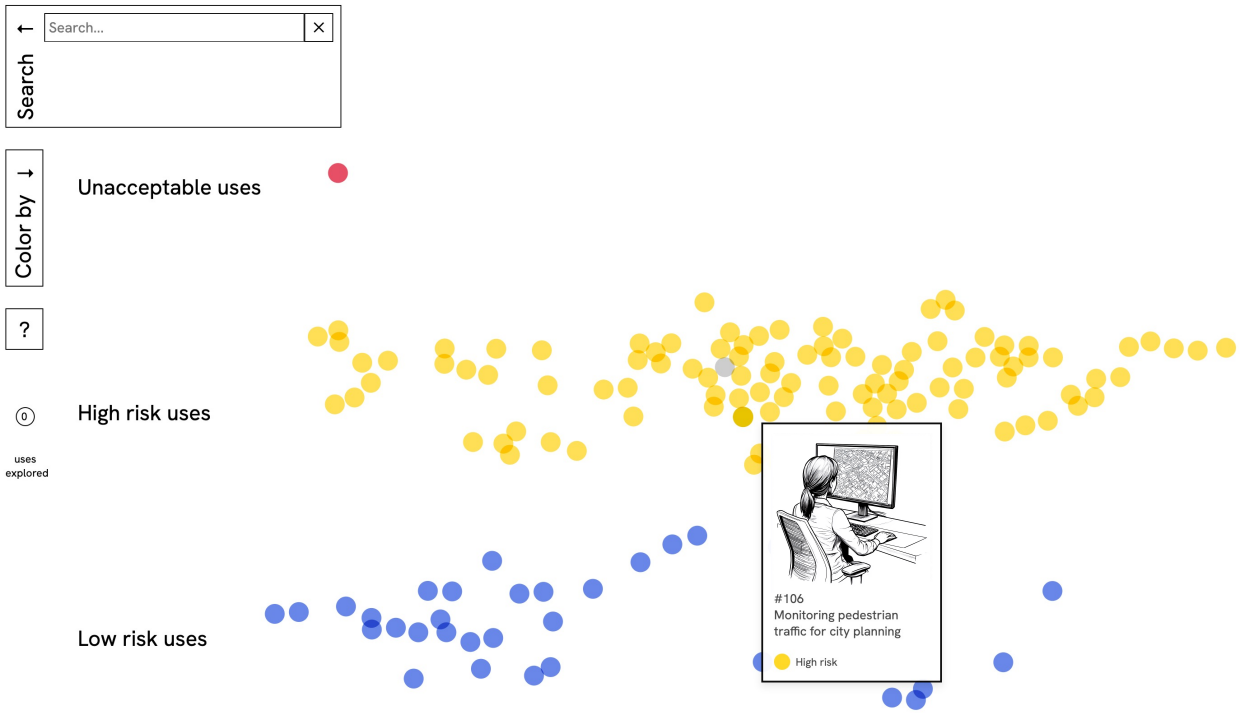
Disagreements due to LLM's techno-optimism clashing with

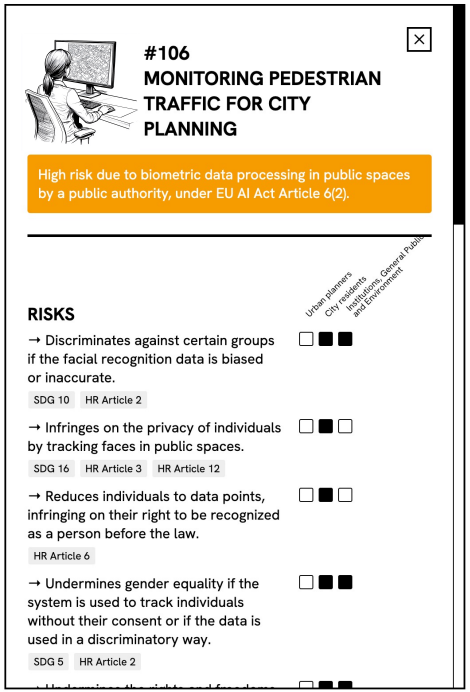
→ human realism: e.g., gates for managing access to climate-sensitive areas

→ digital divide awareness: e.g., telemedicine apps for rural communities

NOKIA BELL LABS

Step 4: Applying data visualization strategies



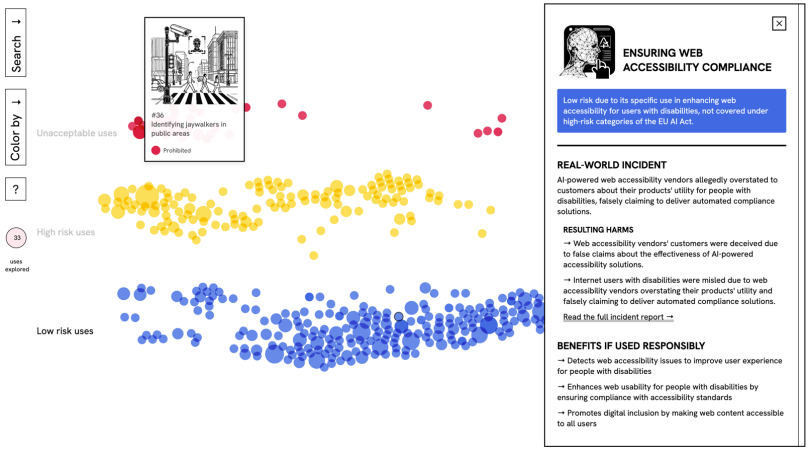


https://social-dynamics.net/atlas

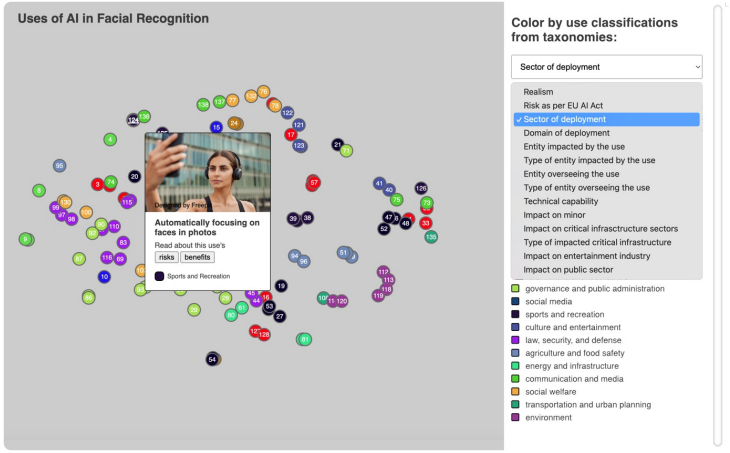
NOKIA BELL LABS

Step 5: Evaluating the tool with members of the broader public

Comparing it to state-of-the-art visualizations



vs



Atlas

Baseline: Spatial view of AI Incident Database

Step 5: Evaluating the tool with members of the broader public

140 individuals
from the general public

matching US census in sex,
age, and ethnicity

contacted on Prolific

Task 1

Explore the visualization

Task 2

Write an email to regulators in which you list specific uses of facial recognition and request either their ban or further adoption

Task 3

Rate the visualization for its usefulness for the task, usability, and aesthetics

The Atlas raised more awareness of the wicked problem…

The visualization helped me to understand
both the risks and benefits of facial recognition

...was more useful...

The visualization helped me to understand both the risks and benefits of facial recognition

The visualization was usable for the task

...more engaging to use...

The visualization kept eyes on it longer

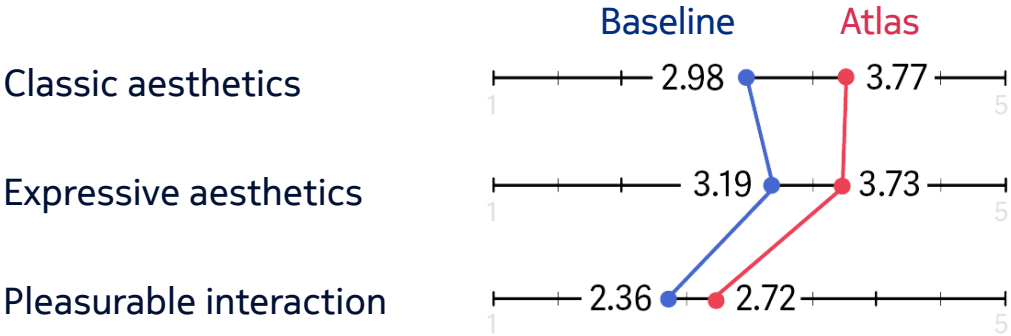
Baseline		6 min 28 sec
Atlas		10 min 17 sec

...more engaging to use...

The visualization kept eyes on it longer

Baseline		6 min 28 sec
Atlas		10 min 17 sec

The visualization visually amplified the seriousness of AI risks



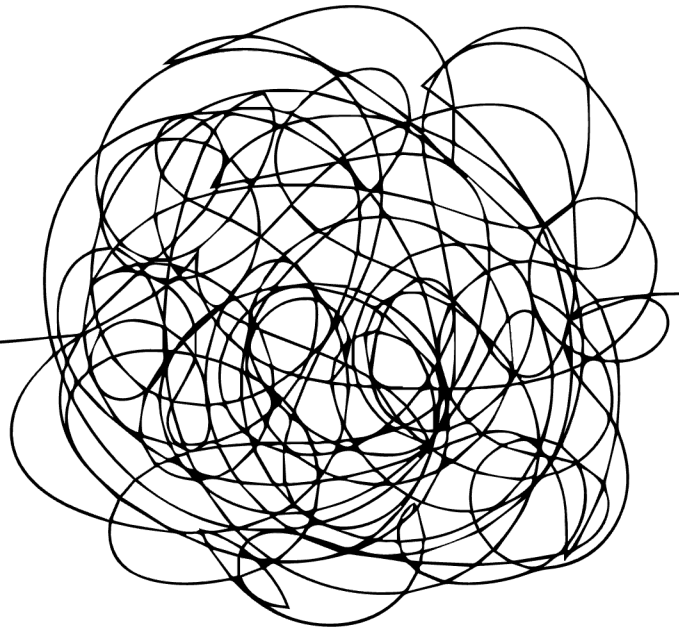
NOKIA BELL LABS

... and gave more personal insights

I think seeing the risks laid our plainly made me realize **I shouldn't be so casual in my acceptance** of facial recognition usage

It brought up a vast array of issues and problems **I could have never thought of on my own**, personally.

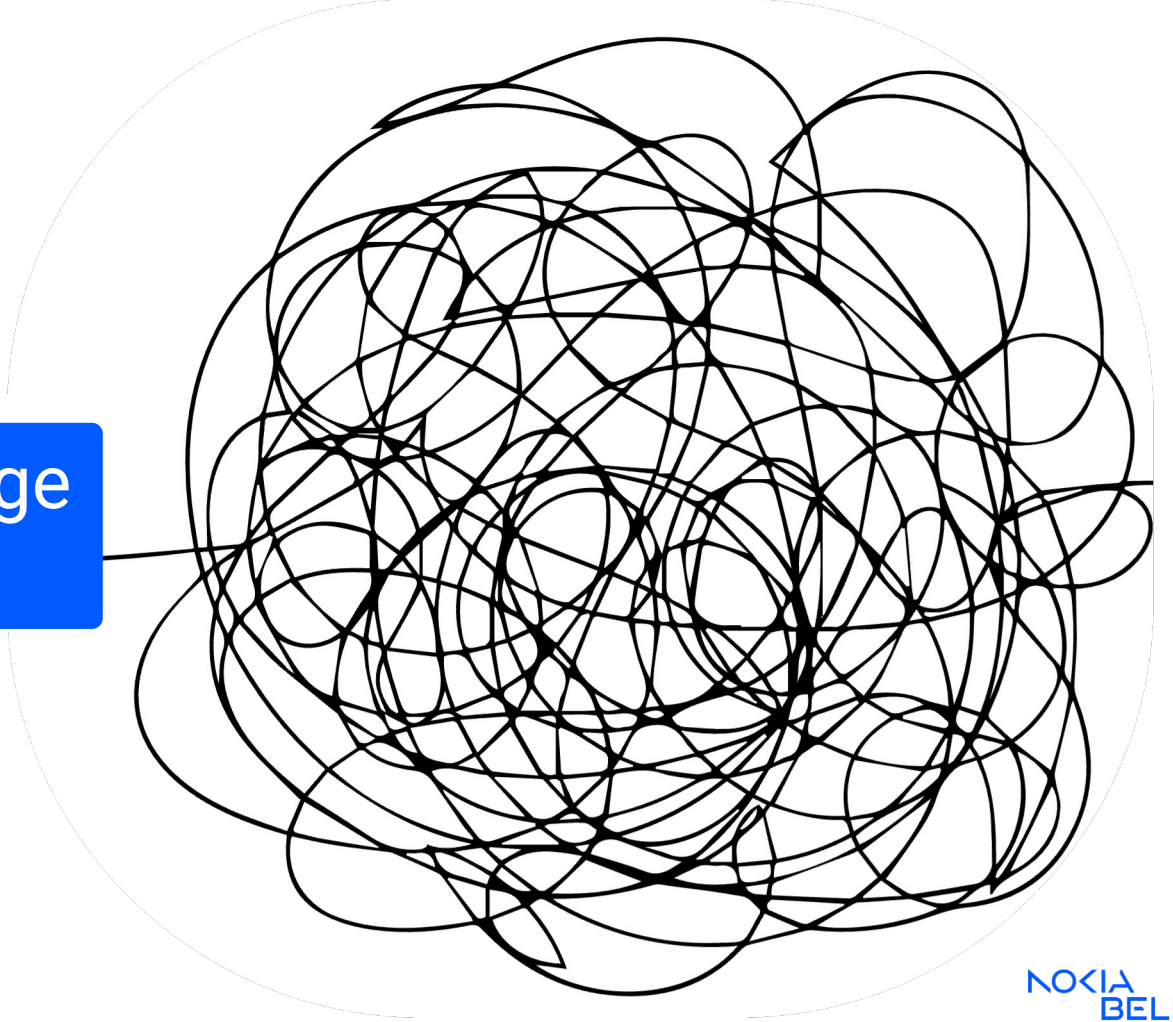
The search box let me quickly get an idea of **my daily activities or interests** that this technology may be integrated into and what the risks were.



What about other wicked problems?

NOKIA BELL LABS

Climate change awareness



NOKIA BELL LABS

From AI risks to wicked problems: a generalizable methodology

Step 1

Crowdsourcing design requirements for the tool

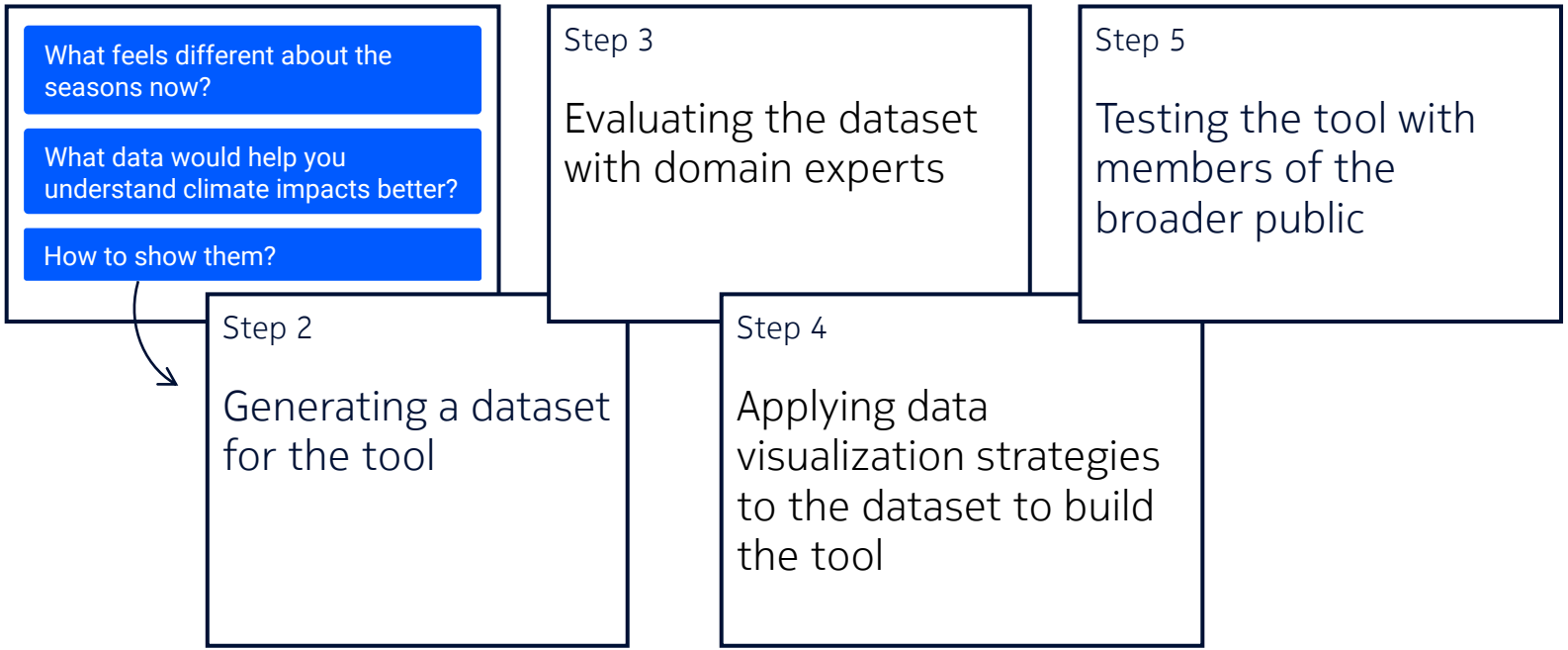
What feels different about the seasons now?

What data would help you understand climate impacts better?

How to show them?

NOKIA BELL LABS

From AI risks to wicked problems: a generalizable methodology



Atlas of AI Risks

Enhancing Public Understanding of AI Risks

www.social-dynamics.net/atlas

Edyta Bogucka, Sanja Šćepanović, Daniele Quercia
@edytapbogucka, @miki7s, @danielequercia

Tips & Tricks for Great Charts in Your Climate Perception Report

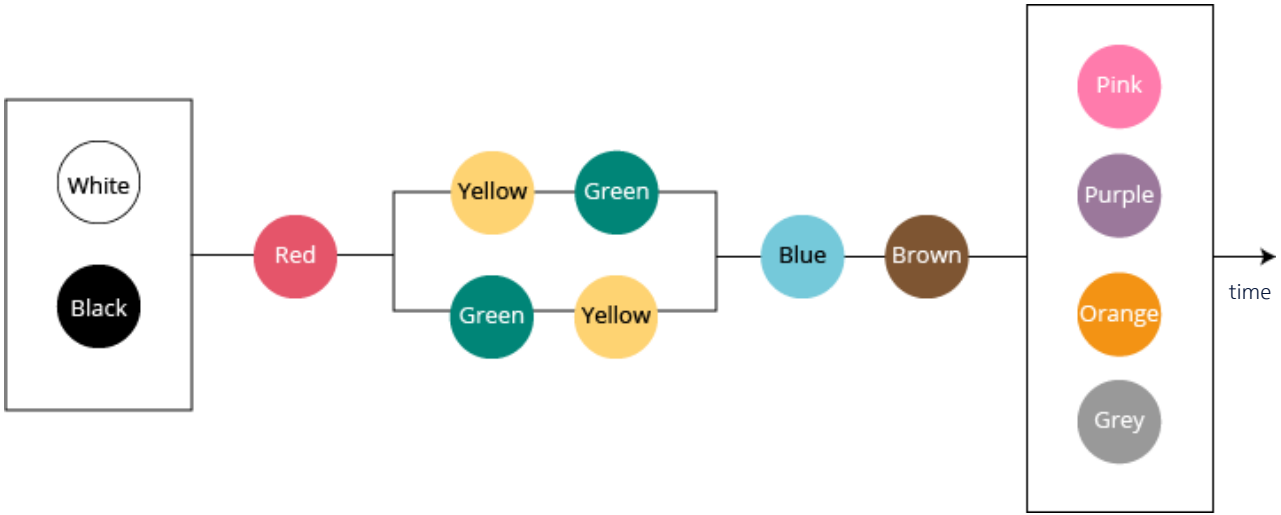
1. Chart organization

- **Prioritize quality over quantity:** 2–3 well-designed, relevant charts per research question are better than ten random ones straight from Matplotlib.
- **Keep it skimmable**: readers should be able to skim just your figures and their captions and grasp the flow and main results of your research.
- **Focus each chart on one idea:** ensure every chart answers one specific research question or illustrates one key insight.
- **Split when needed:** if a chart must cover multiple related insights, use panels instead of crowding everything into one image.
- **Run the Aha-test:** ask yourself, “What’s the one ‘Aha!’ this chart should deliver?” Design everything - layout, labels, annotations - around that insight.
- **Place charts strategically:** position each chart as close as possible to its first mention in the text, ideally at the top or bottom of the page for easy reference.

2. Chart design principles (1)

Don't use too many colors!

The order of appearance of colour names in languages around the world is fixed and far away from rainbow ;-)

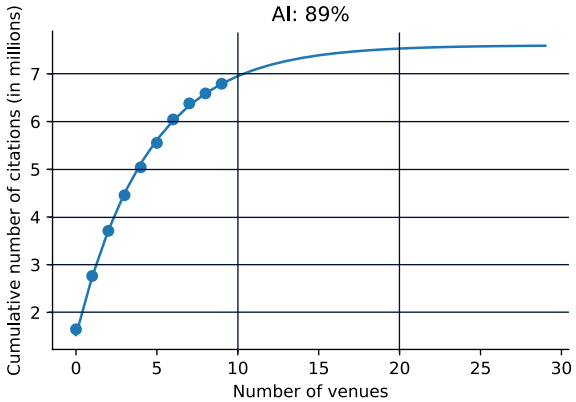


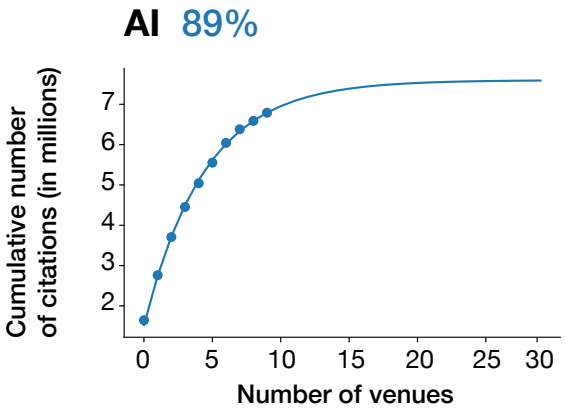
Berlin and Kay (1969); https://www.eea.europa.eu/data-and-maps/daviz/learn-more/chart-dos-and-donts

2. Chart design principles (2)

Use less “ink”

- Remove redundant gridlines, 3D effects, and backgrounds to make your insights pop -> Save your charts as a .pdf and edit them in vector graphic software like Adobe Illustrator, Canvas, or Inkscape
- Use legible fonts: choose a clear sans-serif typeface (e.g., Arial, Helvetica) and set font sizes large enough to remain readable when the chart is resized or printed

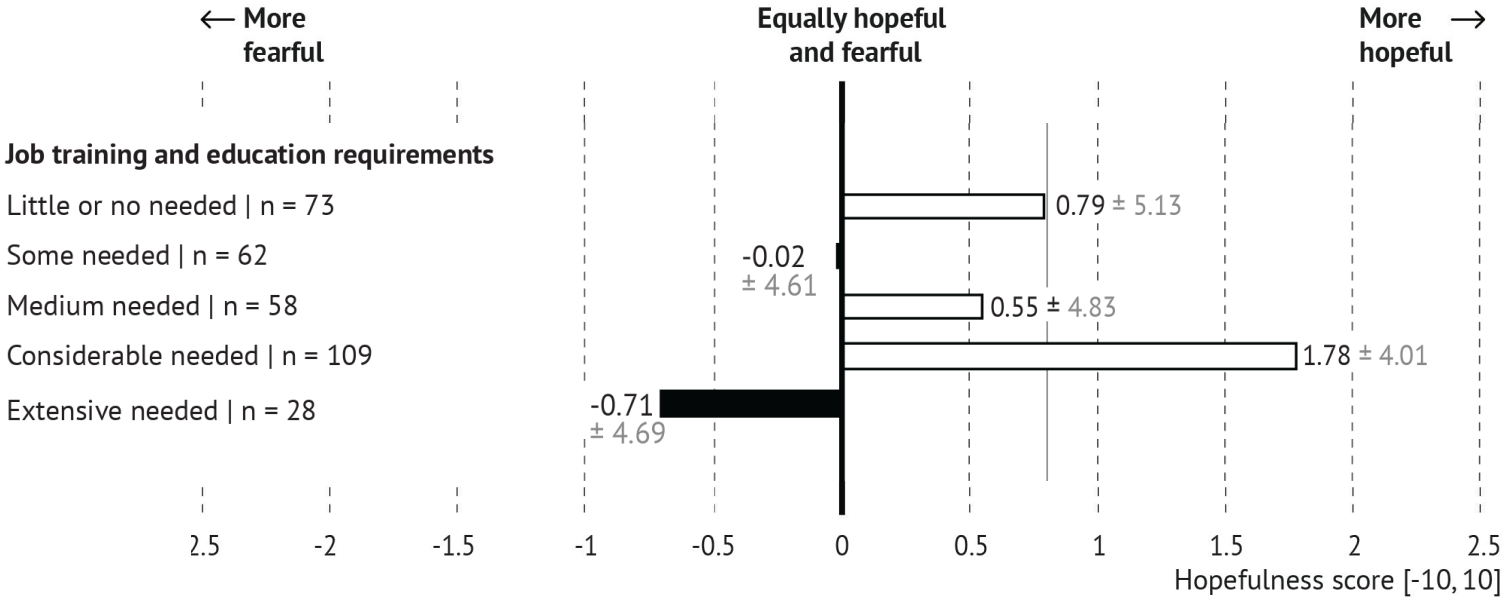




https://social-dynamics.net/docs/rai-impact.pdf

2. Chart design principles (3)

Sort and group by size or relevance - not A to Z!



https://social-dynamics.net/docs/fears-and-hopes.pdf

NOKIA BELL LABS

2. Chart design principles (34)

Label like a human

- Use real words, not variable names
- Make axes and legends understandable without reading the full report and twisting your head
- Include sample size (N) when relevant

2. Chart design principles (4)

Label like a human

patient, image, planning

module, robot, machine

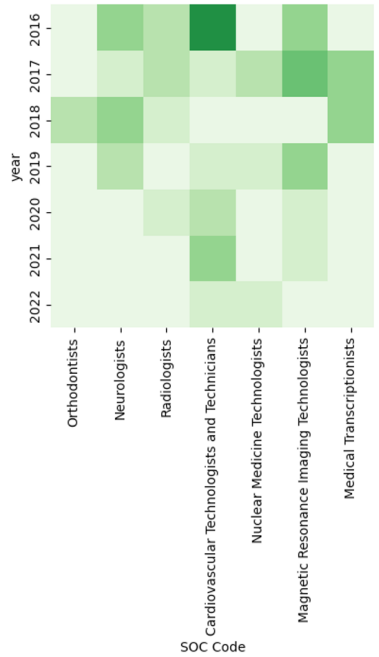
patient, device, medical

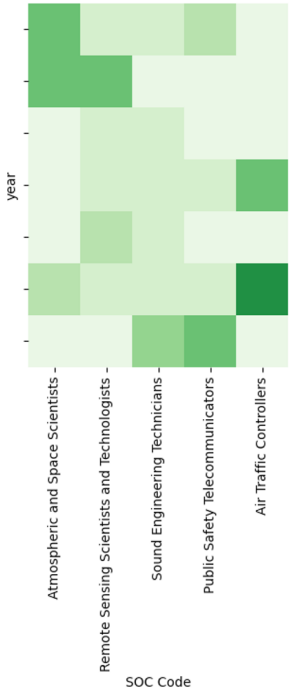
computer, test, state

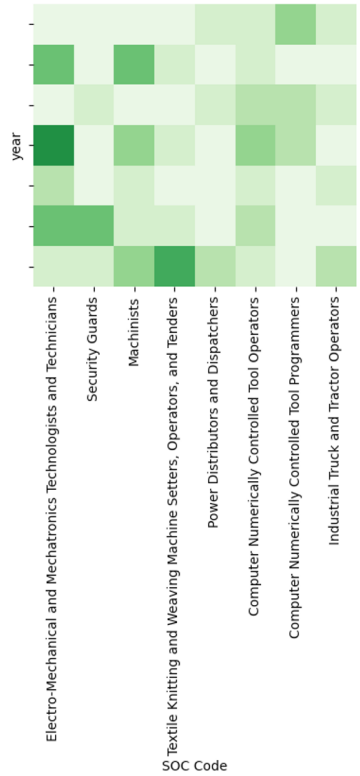
neural network, data, analysis

control, planning, path

control, image, neural network



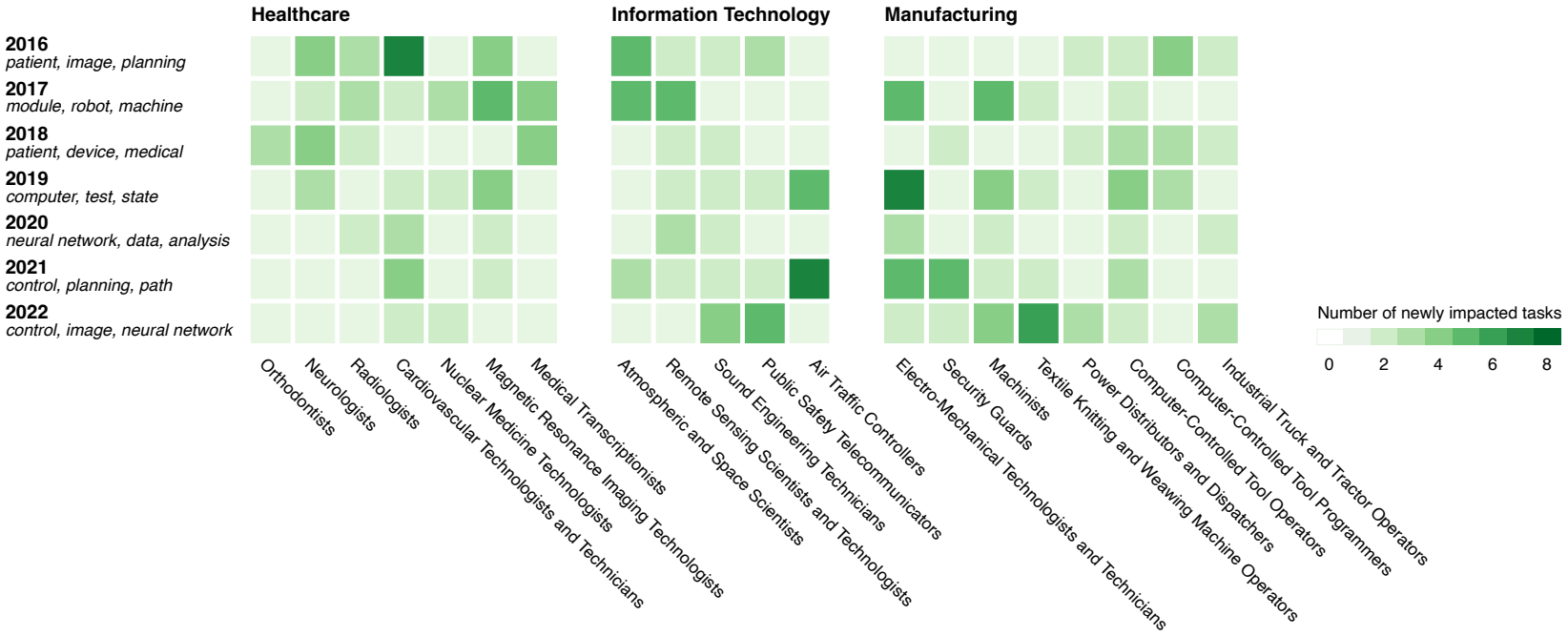




NOKIA BELL LABS

2. Chart design principles (4)

Label like a human



https://social-dynamics.net/docs/aii.pdf

NOKIA BELL LABS

2. Chart design principles (5)

Write captions that do more than just name the chart

State what is being measured and what the reader should notice - highlight the pattern or anomaly.

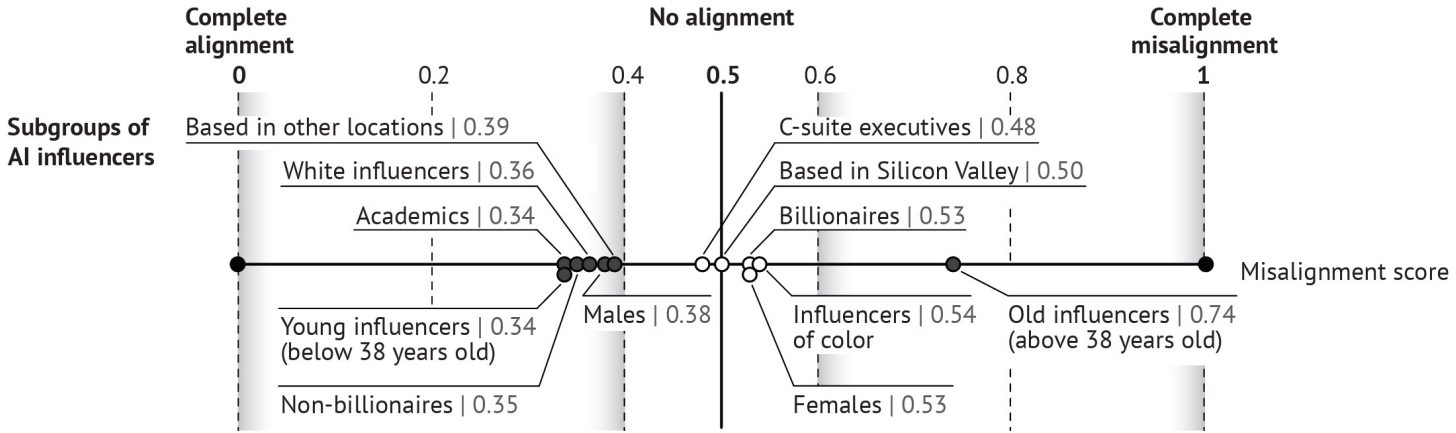


Figure 6: Misalignment scores between subgroups of AI influencers and participants representative of the U.S. population. Young influencers' views are most closely aligned with those of our participants, followed by academics and non-billionaires.

https://social-dynamics.net/docs/fears-and-hopes.pdf

NOKIA BELL LABS

2. Chart design principles (5)

Write captions that do more than just name the chart

Include extra details for more complex charts:

- Any transformations or scales used
- Statistical notes (e.g., error bars show 95% CI)

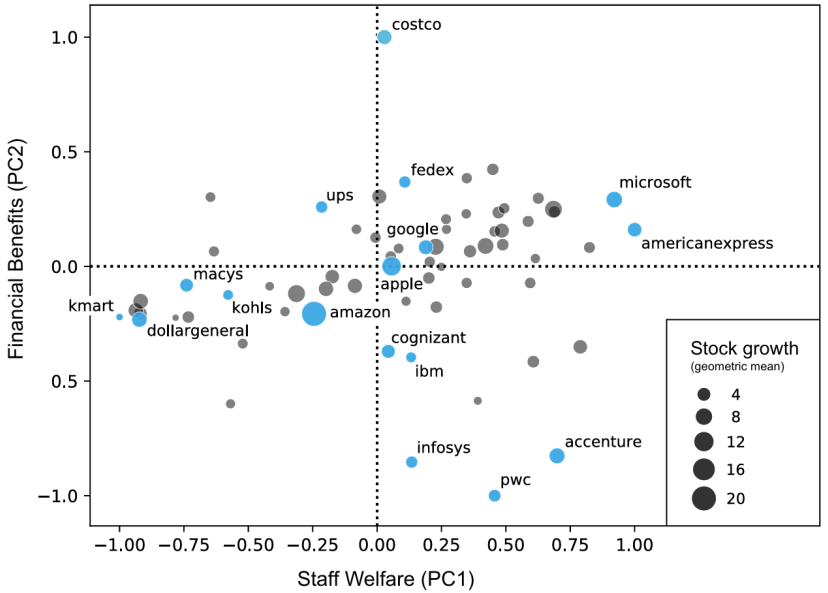


Fig. 7 Scatterplot of the scores of each company's staff welfare vs. financial benefits. The size of a company's dot represents its stock growth. We highlighted in blue some of the companies to assess them qualitatively. Consumer staples and discretionary companies like Kmart, Macy's, and Kohl's scored low for both types of sustainability. Traditional IT companies like Infosys, IBM, and Accenture scored high for staff welfare sustainability but not for financial benefits sustainability.

https://social-dynamics.net/docs/insider-stories-nature.pdf

NOKIA BELL LABS


NOKIA
BELL
LABS