WOMEN IN DATA SCIENCE
DHAHRAN

# From Smart Cities to Social Cities: Smart Cities at the Service of Social Good

WOMEN IN DATA SCIENCE – DHAHRAN, SAUDI ARABIA

ANTONELA TOMMASEL

ISISTAN, CONICET-UNICEN, ARGENTINA

CONICET

ISISTAN

# Who am I?


- Dr Antonela Tommasel
  - PhD in Computer Sciences at UNICEN
- Researcher at ISISTAN, CONICET-UNICEN.
- Teacher Assistant at UNICEN.
- Research Interests:
  - Recommender systems
  - Natural language processing & Text mining
  - Social media
  - Social computing
  - eCitizenship
  - Hate speech
  - …

CONICET

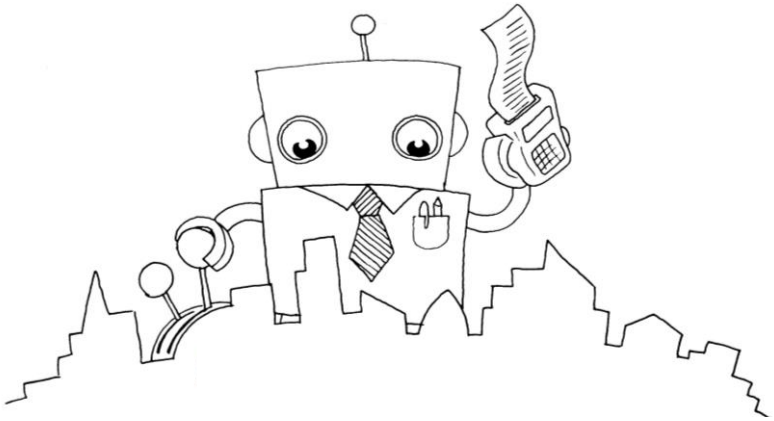
UNIVERSIDAD NACIONAL DEL CENTRO DE LA PROVINCIA DE BUENOS AIRES

I S I S T A N

# Table of contents

# Table of contents

# Smart cities or social cities?



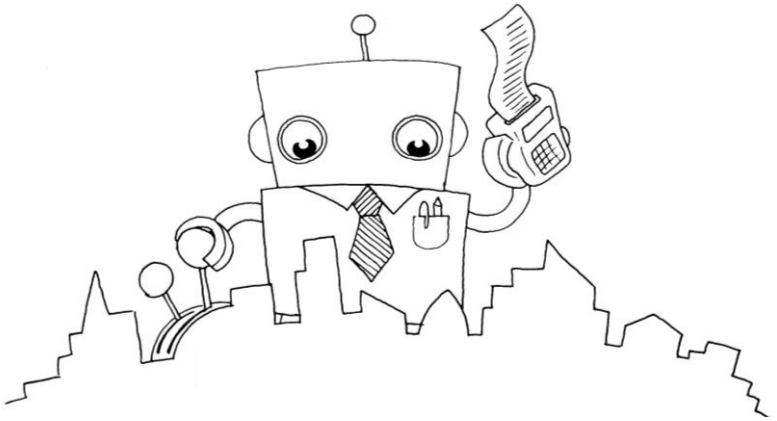
- Urbanization’s rapid progress has modernized many people’s lives.
- The **future** of the cities is being planned based on **new technologies to support their activities**.

- This scenario aims at **building cities’ infrastructure**, **not attending citizens’ needs** to gather and form communities.
  - It does not consider the social nature of human beings.
  - It does not analyze **how disruptive technologies can be** to the sense of community and citizenship.
  - There is more than information processing!

Manifesto for the Clever City: https://theclevercity.net/

# Smart cities or social cities?

## A city is not just a computer



- Urbanization's rapid progress has modernized many people's lives.
- The **future** of the cities is being planned based on **new technologies to support their activities**.

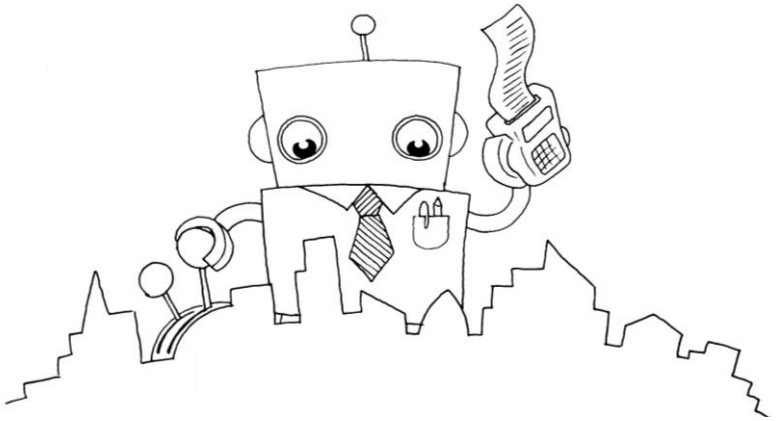
- This scenario aims at **building cities' infrastructure**, **not attending citizens' needs** to gather and form communities.
  - It does not consider the social nature of human beings.
  - It does not analyze **how disruptive technologies can be** to the sense of community and citizenship.
  - There is more than information processing!

**"Smart cities require more than just building band width, but instead need means to connect citizens and governance structures 'to ensure that all citizens have the opportunity to benefit from the knowledge-based networked economy"**

Manifesto for the Clever City: https://theclevercity.net/

Coe, A., Paquet, G., and Roy, J., 2001. E-Governance and Smart Communities: A Social Learning Challenge. Social Science Computer Review, 19 (1), 80–93

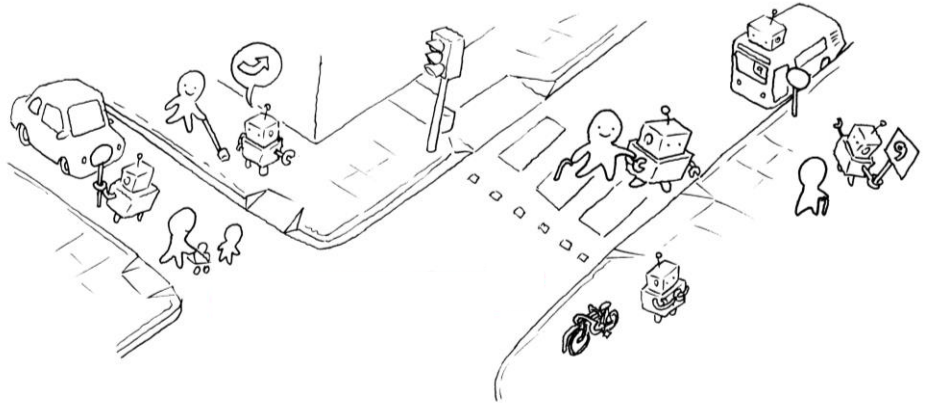
# Smart cities and social cities



- Urbanization's rapid progress has modernized many people's lives.
- The **future** of the cities is being planned based on **new technologies to support their activities**.

- This scenario aims at **building cities' infrastructure**, **not attending citizens' needs** to gather and form communities.
  - It does not consider the social nature of human beings.
  - It does not analyze **how disruptive technologies can be** to the sense of community and citizenship.
  - There is more than information processing!

**"Citizenship shouldn't just be a form of data collection but instead should emphasize deep and meaningful civic engagement or community control in questions about local urban planning and design"**

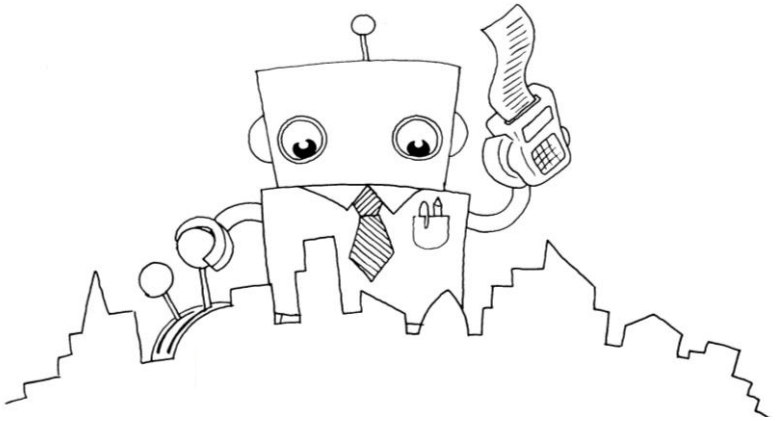


Manifesto for the Clever City: https://theclevercity.net/

Sadoway, D. and Shekhar, S., 2017. (Re)Prioritizing Citizens in Smart Cities Governance: Examples of Smart Citizenship from Urban India. The Journal of Community Informatics

# Smart cities and social cities

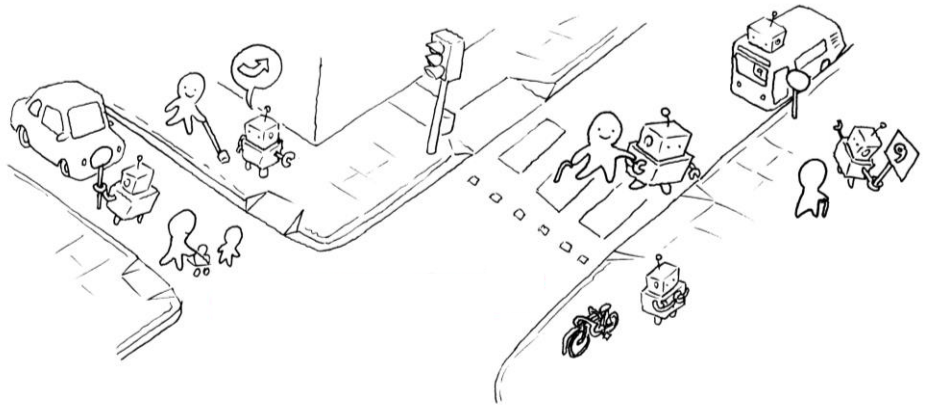
## AI and social good



- Urbanization's rapid progress has modernized many people's lives.
- The **future** of the cities is being planned based on **new technologies to support their activities**.

- This scenario aims at **building cities' infrastructure**, **not attending citizens' needs** to gather and form communities.
  - It does not consider the social nature of human beings.
  - It does not analyze **how disruptive technologies can be** to the sense of community and citizenship.
  - There is more than information processing!

**AI is not a silver bullet, but it could help tackle some of the world's most challenging social problems**



Manifesto for the Clever City: https://theclevercity.net/

Sadoway, D. and Shekhar, S., 2017. (Re)Prioritizing Citizens in Smart Cities Governance: Examples of Smart Citizenship from Urban India. The Journal of Community Informatics

# Table of contents

# AI areas of action

- The **UN sustainable development goals** set different areas in which AI could help!

| | | | |
|---|---|---|---|
| Public and social sector | Security and justice | Crisis response | Economic empowerment |
| Education | Infraestructure | Information verification and validation | Health and hunger |
| | Equality and inclusion | Environment | |

https://www.undp.org/content/undp/en/home/sustainable-development-goals.html

McKinsey Global Institute Analysis

# AI areas of action

- The **UN sustainable development goals** set different areas in which AI could help!

| | | | |
|---|---|---|---|
| Public and social sector | Security and justice | Crisis response | Economic empowerment |
| Education | Infraestructure | Information verification and validation | Health and hunger |
| | Equality and inclusion | Environment | |

https://www.undp.org/content/undp/en/home/sustainable-development-goals.html

McKinsey Global Institute Analysis

# AI areas of action

Public and social sector

- The UN goals set different areas in which AI could help!

- Initiatives related to efficiency and the effective management of public- and social-sector entities, including:
  - Strong institutions.
  - Transparency.
  - Financial management.

- For example, AI can be used to identify **tax fraud** using alternative data such as browsing data, retail data, or payments history.

Revitalize the global partnership for sustainable development.

Good health and well-being

# AI areas of action

## Digital Citizenship

Public and social sector

- Formal political process dependent on the effective communication about public issues amongst citizens.
- E-participation involves the extension and transformation of participation in societal processes mediated by technologies.
- Support active citizenship with the latest technology developments, increasing access to and availability of participation to promote fair and efficient societies and governments.
- What can we do with this?
  - Study the relationship of the government with the citizens.
  - Study the characteristics of citizens.
  - Create citizens' profiles.
  - Estimate the engagement cycle of citizens.
  - Try to foster the engagement of citizens.

# AI areas of action

Economic empowerement

- The UN goals set different areas in which AI could help!

- Emphasis on currently vulnerable populations.

- Opening access to economic resources and opportunities:
  - Jobs.
  - Market information.

- For example,
  - AI can be used to detect plant damage early through low-altitude sensors (including smartphones and drones), to improve yields for small farms.
  - AI can be used to determine the body condition score from cows.

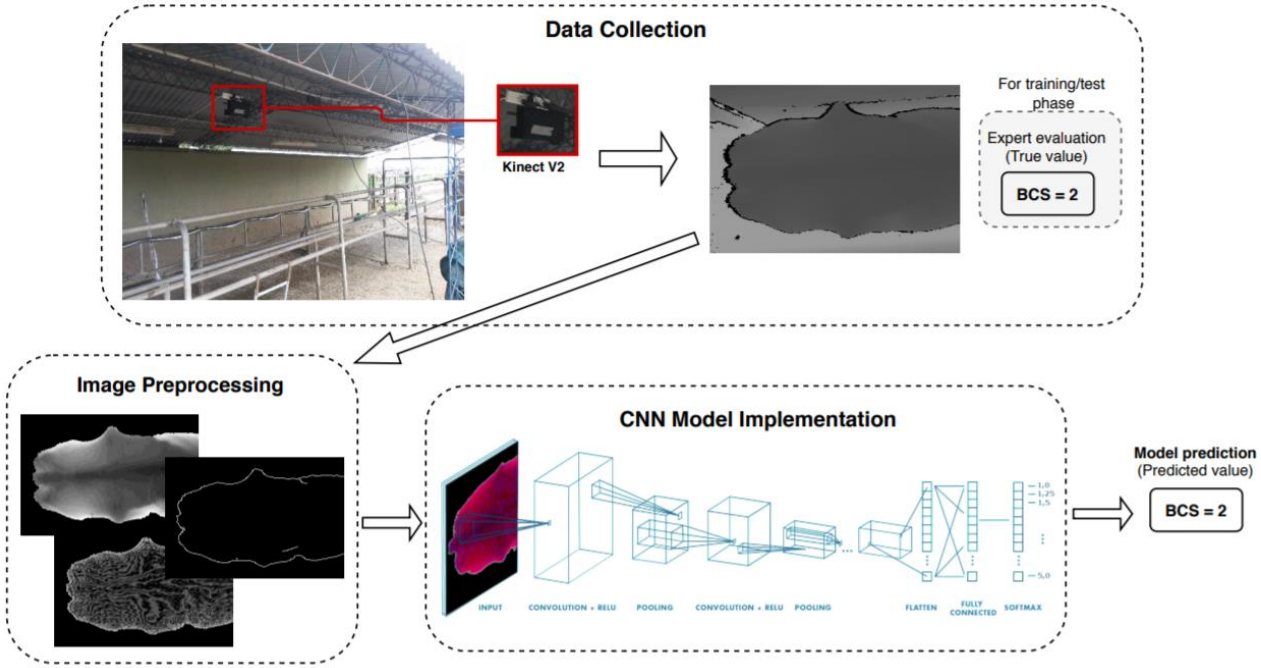
| |
|---|
| Zero hunger |
| Decent work and economic growth |
| No poverty |
| Peace, justice and strong institutions |

# AI areas of action

## Estimating body conditions of cows

Economic empowerement

- Body Condition Score is especially important for dairy cows as it is not only a measurement of **obesity** degree, but also a suitable assessment of **feeding management** according to each stage of lactation, which heavily influences milk production, reproduction, and cow health.
- Usually a time-consuming manual task performed by experts.
- Applied a novel CNN-based model to estimate BCS on cows from depth images.




Rodríguez Alvarez, J., Arroqui, M., Mangudo, P., Toloza, J., Jatip, D., Rodriguez, J.M., Teyseyre, A., Sanz, C., Zunino, A., Machado, C. and Mateos, C., 2019. Estimating body condition score in dairy cows from depth images using convolutional neural networks, transfer learning and model ensembling techniques. *Agronomy*, *9*(2), p.90.

# AI areas of action

Health and hunger

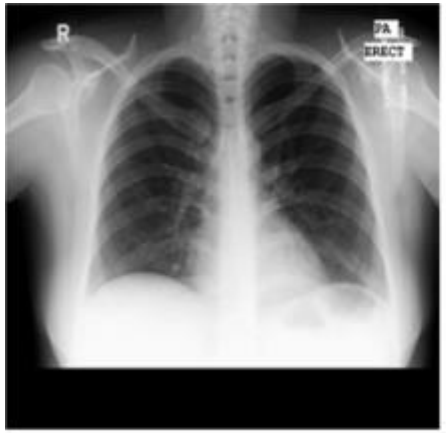
- The UN goals set different areas in which AI could help!
- Early-stage diagnosis.
- Optimized food distribution.
- For example,
  - At the University of Heidelberg and Stanford University have created a disease-detection AI system using images of skin lesions to determine if they are cancerous.
  - AI-enabled wearable devices can already detect people with potential early signs of diabetes using heart-rate sensor data.
  - Sedentary detection.

| |
|---|
| Zero hunger |
| Good health and well-being |

# AI areas of action

## COVID & Deep Learning

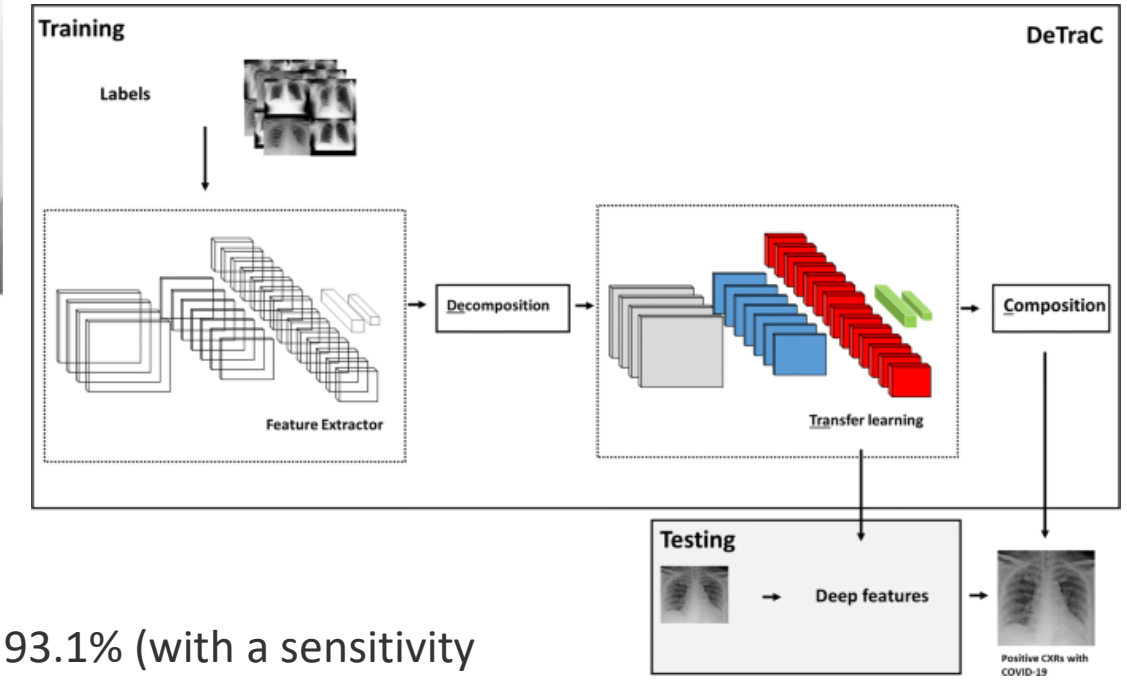
Health and hunger

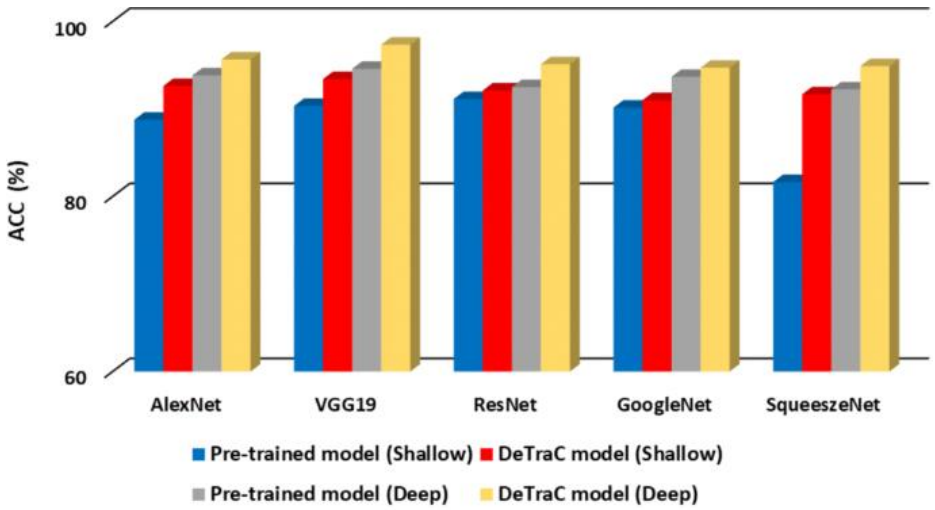


Normal

COVID-19

SARS




- High accuracy of 93.1% (with a sensitivity of 100%) was achieved by *DeTraC* .

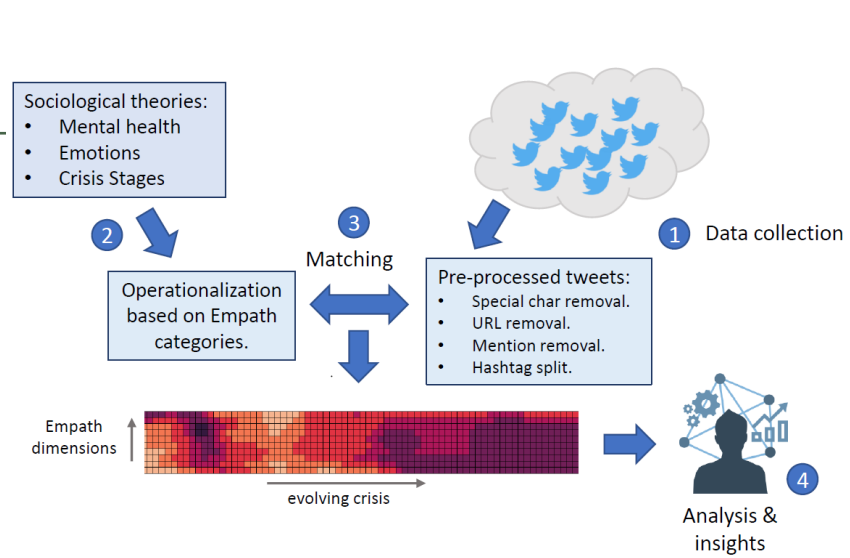
Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network

# AI areas of action

## COVID & Mental Health

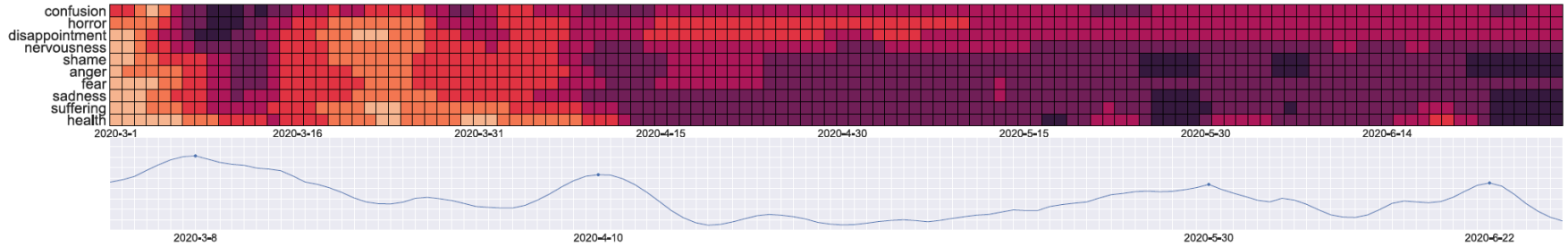
**Health and hunger**

- To what extent social media reflects changes in anxiety, stress and depression markers, as the COVID-19 crisis evolved?
- Are there changes in the manifestation of the selected disorders?

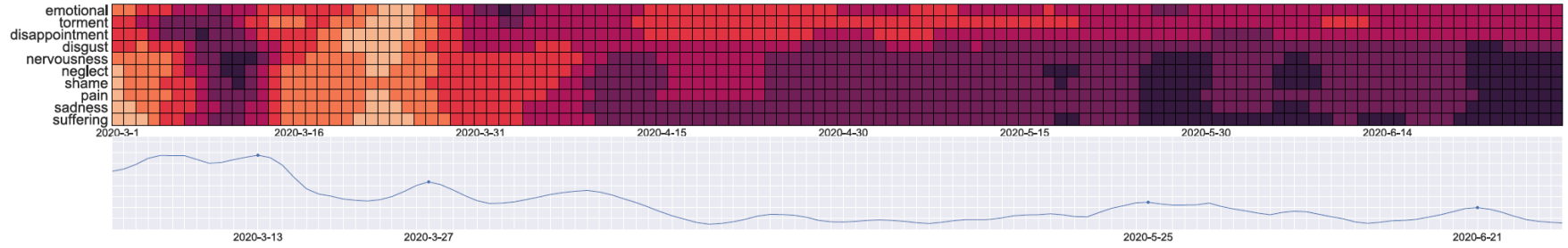



- Recent evidence has suggested that people who are kept under lock-down experience significant levels of anxiety, anger, confusion, depression and stress.
- The variations observed in the markers show that social media reflects the events in the real world.
- Found matching between the changes in the prevalence of markers and relevant COVID events in Argentina.

Anxiety




Depression



# AI areas of action

Information validation and verification

- The UN goals set different areas in which AI could help!
- Fake news is **made-up stuff**, masterfully manipulated to **look like credible** journalistic reports that are **easily spread** online to large audiences willing to believe the fictions and spread the word
- Facilitating the provision, validation, and recommendation of helpful, valuable, and reliable information to all.
- Focuses on filtering or counteracting misleading and distorted content, including false and polarizing information disseminated through Internet and social media.
- A threat to the access to reliable and **trustworthy** information and the establishment of **reliable** social relations.

Peace, justice and strong institutions

# AI areas of action

**Information validation and verification**

- The UN goals set different areas in which AI could help!
- Fake news is **made-up stuff**, masterfully manipulated to **look like credible** journalistic reports that are **easily spread** online to large audiences willing to believe the fictions and spread the word
- Facilitating the provision, validation, and recommendation of helpful, valuable, and reliable information to all.
- Focuses on filtering or counteracting misleading and distorted content, including false and polarizing information disseminated through Internet and social media.
- A threat to the access to reliable and **trustworthy** information and the establishment of **reliable** social relations.

Peace, justice and strong institutions

Additional “real life” consequences!

- Mob killings
- Stock changes
- Health
- Political manipulation

# AI areas of action

## Fake news, social bots & more

Information validation and verification

- Social media are vulnerable to deceptive social bots, which can impersonate humans to amplify misinformation and manipulate opinions.

- Some social bots are designed with benign intentions and serve useful purposes, but they also have many harmful applications:
  - Amplify misinformation.
  - Create the appearance that people or opinions are more popular than they are.
  - Influence public opinion.
  - Commit financial fraud.
  - Infiltrate vulnerable communities.
  - Disrupt communication.

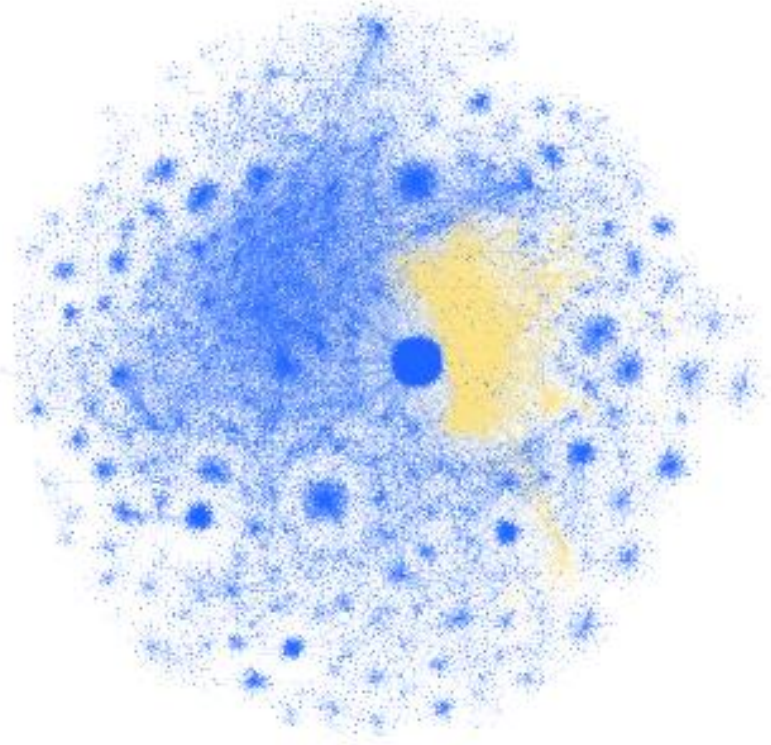
- Some social bots behave in ways that cannot be distinguished by those of humans on an individual basis because automation is mixed with or copied from human behaviour.

Lou, X., Flammini, A. and Menczer, F., 2019. Information Pollution by Social Bots. *arXiv preprint arXiv:1907.06130*.

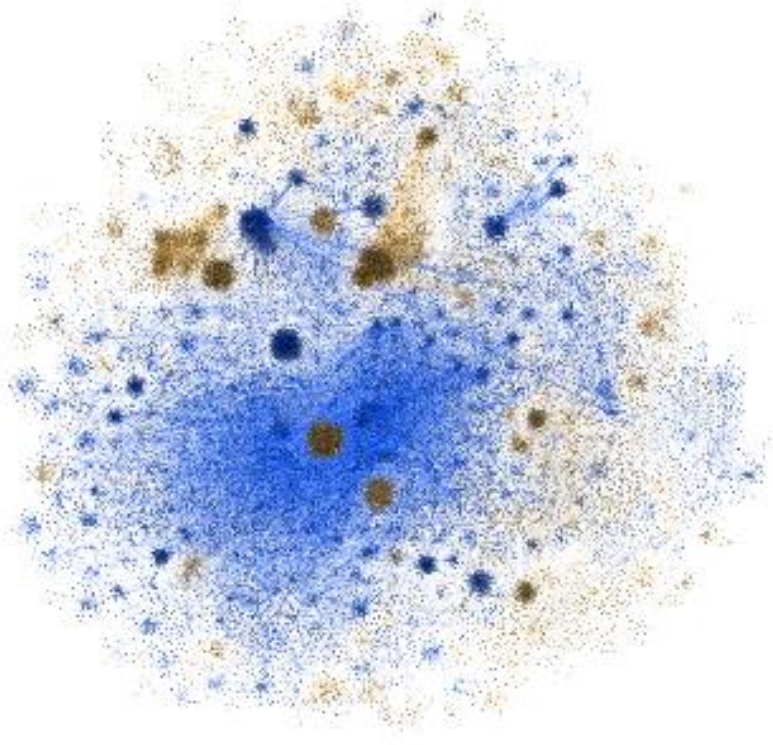
# AI areas of action

## Fake news, social bots & more

Information validation and verification



chemtrails conspiracies (yellow) mix with conversations about the sky



antivax campaigns (brown) penetrate discussions about the flu

# AI areas of action

## COVID-19 & Fake news

**Information validation and verification**

We are working closely together on COVID-19 response efforts. We're helping millions of people stay connected while also jointly combating fraud and misinformation about the virus, elevating authoritative content on our platforms, and sharing critical updates in coordination with government healthcare agencies around the world. We invite other companies to join us as we work to keep our communities healthy and safe.

FACEBOOK Google LinkedIn Microsoft reddit YouTube

Joint Industry Statement from Facebook, Google, LinkedIn, Microsoft, Reddit, Twitter and YouTube

**Alberto Fernández difundió una "fake news" sobre la pandemia del coronavirus: qué recomendó**

"La Organización Mundial de la Salud, entre las cosas que recomienda, **es que uno tome muchas bebidas calientes, porque precisamente el calor mata el virus**", dijo Fernández.

**Alberto Fernández** @alferdez

Según un informe publicado por el@@washingtonpost y realizado por el Instituto Tecnológico de Massachusetts (MIT) Evo Morales ganó los comicios electorales del año pasado por más de 10 puntos de diferencia, sin que mediara fraude alguno clarin.com/mundo/informe-...

3:06 pm · 29 Feb 2020 · Twitter for iPhone

**10.6K** Retweets **23.3K** Likes

**Aníbal Fernández** @FernandezAnibal

Llegó el Corralito
El BCRA anuncia que no se venden mas dólares


1.136 13:51 - 13 ago. 2018

1.308 personas están hablando de esto

# AI areas of action

## Security and justice

- The UN goals set different areas in which AI could help!

- Focuses on security, policing, and criminal-justice issues.
  - Preventing crime and other physical dangers.
  - Tracking criminals.
- Mitigating bias in police forces.

- For example,
  - Using AI and IoT devices to create solutions that to help firefighters determine safe paths through burning buildings.

  - Design and simulate evacuation plans.

| |
|---|
| Sustainable cities and communities |
| Reduced inequalities |
| Peace, justice, and strong institutions |

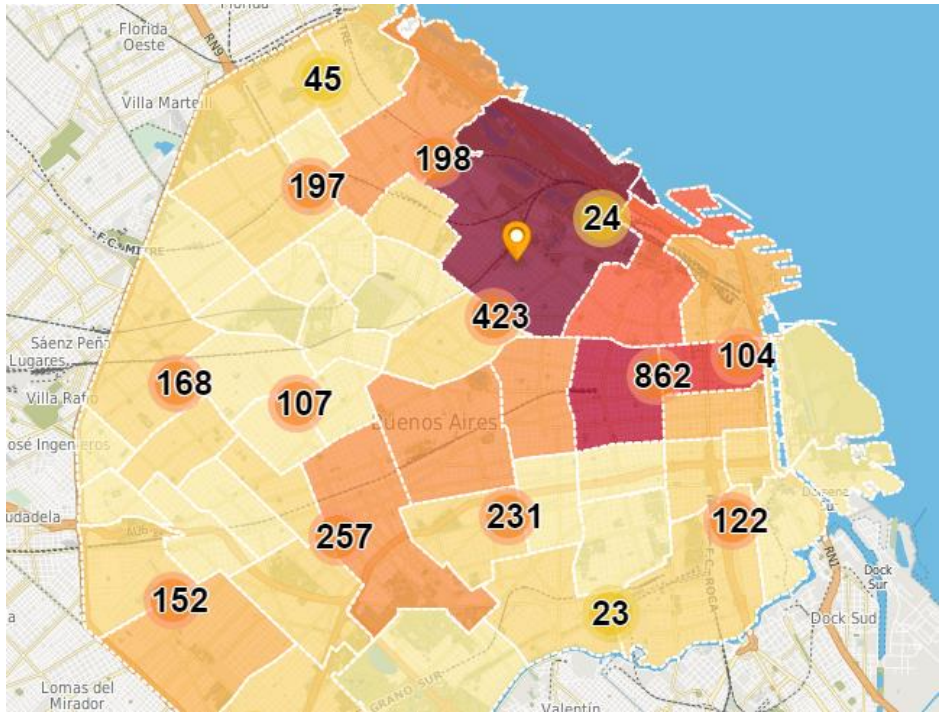
# AI areas of action

## Crime

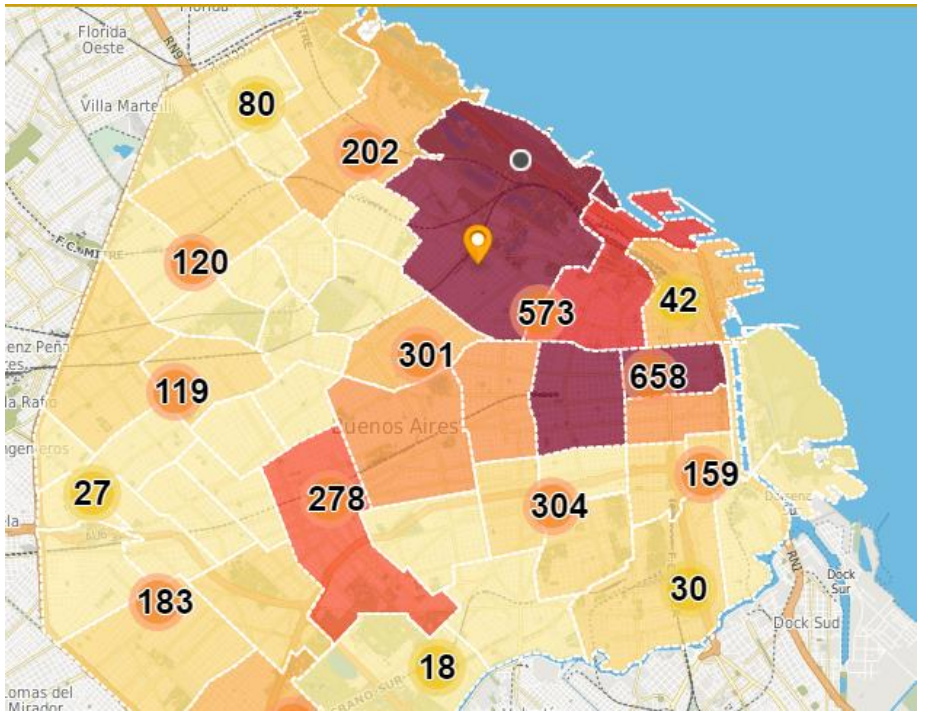
Security and justice

Buenos Aires Ciudad | Mapa del Delito

September 2018




December 2018



# Limitations for AI use

## Critical barriers for most domains

- Data availability
- Data quality
- High-level AI-expertise availability
- Regulatory limitations
- Data volume
- Data labelling
- Access to computing capacity

## Contextual challenges

- Data integration
- Access to technology
- Privacy concerts
- Receptiveness

# Limitations for AI use

## Critical barriers for most domains

- **Data availability**
- **Data quality**
- High-level AI-expertise availability
- Regulatory limitations
- Data volume
- Data labelling
- Access to computing capacity

## Contextual challenges

- **Data integration**
- Access to technology
- Privacy concerts
- Receptiveness

- Most of the data essential or useful for social-good applications are in **private hands** or in public institutions that might not be willing to share their data.

- These data owners include:
  - Telecommunications and satellite companies.
  - Social-media platforms;
  - Financial institutions (e.g. credit histories).
  - hospitals, doctors, and other health providers.
  - Governments (e.g. tax information for private individuals).

- Why is data difficult to obtain?
  - Regulations on data use.
  - Privacy concerns.
  - Bureaucratic inertia.
  - Data has business value → **commercial use!**

# Limitations for AI use

## Critical barriers for most domains

- Data availability
- Data quality
- **High-level AI-expertise availability**
- Regulatory limitations
- Data volume
- Data labelling
- Access to computing capacity

## Contextual challenges

- Data integration
- Access to technology
- Privacy concerts
- **Receptiveness**

- Not only expertise in building the models but also on interpreting the results.

- Even if a model achieves a desired level of accuracy on test data, new or unanticipated failure cases often appear in real-life scenarios.
  - Need someone to “translate” it.
  - Without “translation” the model could be trusted too much!

**Usually, models do not work on the 100% of the cases!**

# Table of contents

# Responsible AI practices

The development of AI is creating new opportunities to **improve the lives of people** around the world, from business to healthcare to education.

# Responsible AI practices

The development of AI is creating new opportunities to **improve the lives of people** around the world, from business to healthcare to education.

**AI tools and techniques can be misused, so principles for their use must be established.**

# Responsible AI practices

The development of AI is creating new opportunities to **improve the lives of people** around the world, from business to healthcare to education.

**AI tools and techniques can be misused, so principles for their use must be established.**

It is also raising new questions about the best way to build the solutions.

“with great power comes great responsibility”

# Responsible AI practices

The development of AI is creating new opportunities to **improve the lives of people** around the world, from business to healthcare to education.

**AI tools and techniques can be misused, so principles for their use must be established.**

It is also raising new questions about the best way to build the solutions.

"with great power comes great responsibility"

How to be responsible?

# Responsible AI practices

The development of AI is creating new opportunities to **improve the lives of people** around the world, from business to healthcare to education.

**AI tools and techniques can be misused, so principles for their use must be established.**

It is also raising new questions about the best way to build the solutions.

"with great power comes great responsibility"

How to be responsible?

- Fairness
- Interpretability
- Privacy
- Security

https://ai.google/responsibilities/responsible-ai-practices/

# Responsible AI practices

## Fairness

- Systems have the potential to be **fairer** and **more inclusive** at a broader scale than decision-making processes based on **ad hoc rules or human judgments.**

- AI solutions can unintentionally harm the very people they are supposed to help.

- The **risk** is that any **unfairness** in such systems can also **have a wide-scale impact**.

- Models can even be used to identify some of the conscious and unconscious human biases and barriers to inclusion that have developed and perpetuated throughout history, bringing about positive change.

# Responsible AI practices

## Fairness

- Systems have the potential to be **fairer** and **more inclusive** at a broader scale than decision-making processes based on **ad hoc rules or human judgments.**

- AI solutions can unintentionally harm the very people they are supposed to help.

- The **risk** is that any **unfairness** in such systems can also **have a wide-scale impact**.

- Models can even be used to identify some of the conscious and unconscious human biases and barriers to inclusion that have developed and perpetuated throughout history, bringing about positive change.

AI may perpetuate and aggravate existing prejudices and social inequalities,
affecting already-vulnerable populations and amplifying existing cultural prejudices.

# Responsible AI practices

## Fairness

- Systems have the potential to be **fairer** and **more inclusive** at a broader scale than decision-making processes based on **ad hoc rules or human judgments.**

- AI solutions can unintentionally harm the very people they are supposed to help.

- The **risk** is that any **unfairness** in such systems can also **have a wide-scale impact**.

- Models can even be used to identify some of the conscious and unconscious human biases and barriers to inclusion that have developed and perpetuated throughout history, bringing about positive change.

AI may perpetuate and aggravate existing prejudices and social inequalities,
affecting already-vulnerable populations and amplifying existing cultural prejudices.

It is critical to work towards systems that are fair and inclusive for all.

# Responsible AI practices

## Fairness

- **But… easier said than done**

1. Models learn from existing data collected from the real world, and so **an accurate model may learn or even amplify problematic pre-existing biases** (e.g. culture, gender, religion).

2. It is a challenge to ensure that a system will be fair across every possible situation.

3. Identifying appropriate fairness criteria for a system requires accounting for **user experience, cultural, social, historical, political, legal, and ethical considerations**.

# Responsible AI practices

## Fairness

- **But… easier said than done**

1. Models learn from existing data collected from the real world, and so **an accurate model may learn or even amplify problematic pre-existing biases** (e.g. culture, gender, religion).

favour certain gender in interviews, assume pronouns

2. It is a challenge to ensure that a system will be fair across every possible situation.

something trained with adults may be fair in such context, but fail when used with teenagers

3. Identifying appropriate fairness criteria for a system requires accounting for **user experience, cultural, social, historical, political, legal, and ethical considerations**.

# Responsible AI practices

## Fairness – Recommended practices

- It is important to identify whether or not machine learning can help provide an adequate solution to the specific problem at hand.

- If it can, just as there is **no single "correct" model** for all tasks, there is **no** single technique that ensures fairness **in every situation.**

### Design models using concrete goals for fairness and inclusion

Consider how the technology and its development over time will impact different use cases

### Use representative datasets to train and test your model

Assess fairness in your datasets, which includes identifying representation and corresponding limitations, and prejudicial or discriminatory correlations between features, labels, and groups.

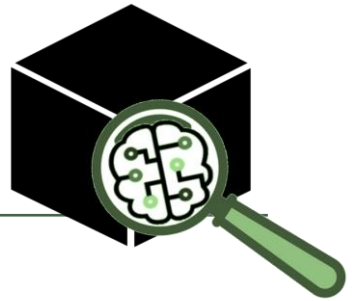
Understand the various perspectives, experiences, and goals of the people annotating the data.

### Check the system for unfair biases

While designing metrics to train and evaluate your system, also include metrics to examine performance across different subgroups.

# Responsible AI practices

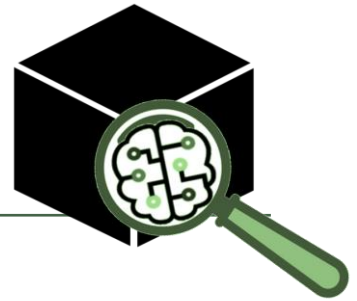
## Interpretability



- It is crucial to being able to **question, understand, and trust** AI systems.
- Reflects **domain knowledge** and **societal values**.
- Provides scientists and engineers with **better means of designing, developing**, and debugging models.
- Helps to ensure that AI systems are **working as intended**.
- Understanding and testing AI systems also offers new challenges compared to traditional software.

# Responsible AI practices

## Interpretability



- It is crucial to being able to **question, understand, and trust** AI systems.
- Reflects **domain knowledge** and **societal values**.
- Provides scientists and engineers with **better means of designing, developing**, and debugging models.
- Helps to ensure that AI systems are **working as intended**.
- Understanding and testing AI systems also offers new challenges compared to traditional software.

Complex AI models include millions of parameters and mathematical operations.

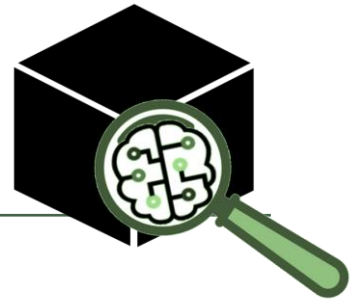
It is much harder to pinpoint what that leads to a faulty decision.

Traditional software is essentially a series of if-then rules.

While that can be gnarly, a human can generally track the path taken through the code, and understand a given result.

# Responsible AI practices

## Interpretability



- It is crucial to being able to **question, understand, and trust** AI systems.
- Reflects **domain knowledge** and **societal values**.
- Provides scientists and engineers with **better means of designing, developing**, and debugging models.
- Helps to ensure that AI systems are **working as intended**.
- Understanding and testing AI systems also offers new challenges compared to traditional software.

Complex AI models include millions of parameters and mathematical operations.

It is much harder to pinpoint what that leads to a faulty decision.

values can be traced to the training data or model

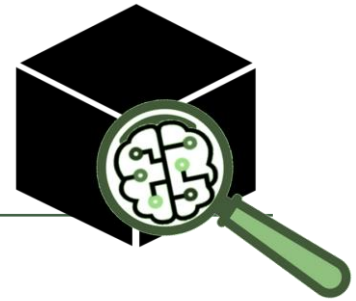
Traditional software is essentially a series of if-then rules.

While that can be gnarly, a human can generally track the path taken through the code, and understand a given result.

"magic numbers", magic thresholds or now-forgotten rules, personal intuition

# Responsible AI practices

## Interpretability – Recommended practices



- An AI system is best understood by the underlying training data and training process, as well as the resulting AI model.

### Plan out your options to pursue interpretability

Pursuing interpretability can happen before, during and after designing and training your model.
What degree of interpretability is really needed?
Are you providing too much transparency, potentially opening up vectors for abuse?

### Design the model to be interpretable

Use the **smallest** set of inputs necessary for your performance goals to make it clearer what factors are affecting the model.
Use the **simplest** model that meets your performance goals.
Learn causal relationships not correlations when possible.

### Treat interpretability as a core part of the user experience

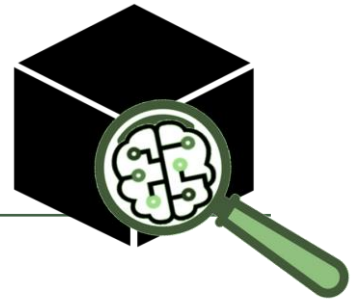
If not given clear and compelling information, users may make up their own theories about how an AI system works, which can negatively affect how they try to use the system.

### Choose metrics to reflect the end-goal and the end-task

The metrics you consider must address the particular benefits and risks of your specific context. (e.g. precision vs recall)
Analyse the model's sensitivity to different inputs, for different subsets of examples.

# Responsible AI practices

## Interpretability – Recommended practices



- An AI system is best understood by the underlying training data and training process, as well as the resulting AI model.

**Communicate explanations to model users**

Provide explanations that are understandable and appropriate for the user.

**Explanations should be informed by a careful consideration of philosophical, psychological, computer science (including HCI), legal and ethical considerations about what counts as a good explanation in different contexts.**

Identify if and where **explanations may not be appropriate** (could they cause more confusion?)

Prioritize explanations that suggest clear actions a user can take to correct inaccurate predictions going forward.
Don't imply that explanations mean causation unless they do.

Recognize human psychology and limitations (e.g., confirmation bias, cognitive fatigue).

Be mindful of the **limitations** of explanations.

# Responsible AI practices

## Privacy

- Sometimes the training data, input data, or both can be quite **sensitive**.
  - We don’t want it to go public!

- Although there may be enormous **benefits** to building a model that operates on sensitive data, it is essential to consider the **potential privacy implications** in using sensitive data.

- Not only respecting the legal and **regulatory requirements**, but also considering **social norms** and typical individual expectations.

# Responsible AI practices

## Privacy

- Sometimes the training data, input data, or both can be quite **sensitive**.
  - We don't want it to go public!

- Although there may be enormous **benefits** to building a model that operates on sensitive data, it is essential to consider the **potential privacy implications** in using sensitive data.

- Not only respecting the legal and **regulatory requirements**, but also considering **social norms** and typical individual expectations.

  - What safeguards need to be put in place **to ensure the privacy** of individuals considering that ML models may remember or reveal aspects of the data they have been exposed to?

  - What steps are needed to ensure users have adequate **transparency and control** of their data?

# Responsible AI practices

## Privacy – Recommended practices

- There is no single correct approach to ML privacy protection across all scenarios.
- Balance privacy and utility.
- Privacy needs to be clearly defined.

### Collect and handle data responsibly

Identify whether your ML model can be trained **without** the use of sensitive data.
Anonymize and aggregate incoming data using best practice data-scrubbing pipelines.

### Leverage on-device processing where appropriate

Consider collecting only statistics that have been computed locally, on-device, rather than raw interaction data, which can include sensitive information.
When feasible, apply aggregation, randomization, and scrubbing operations on-device.

### Appropriately safeguard the privacy of ML models

Consider the privacy impact of how the models were constructed and may be accessed.
Estimate whether your model is unintentionally memorizing or exposing sensitive data.
Train ML models using techniques that establish mathematical guarantees for privacy.

# Responsible AI practices

## Security

- Safety and security entails ensuring AI systems **behave as intended**, regardless of how attackers try to **interfere**.

- It is essential to consider and address the security of an AI system before it is widely relied upon in safety-critical applications.

- **Challenges**.
  - It is **hard to predict all scenarios** ahead of time, especially when ML is applied to problems that are difficult for humans to solve.
  - It is hard to build systems that provide **both the necessary restrictions for security** as well as the **necessary flexibility to generate creative solutions** or adapt to unusual inputs.
  - **New ways of attacking constantly appear**. Hard to keep up.

# Responsible AI practices

## Security – Recommended practices

- Developers should think about whether their system is likely to come under attack, consider the likely consequences of a successful attack and in most cases should simply not build systems where such attacks are likely to have significant negative impact.

**Identify potential threats to the system**

Consider whether anyone would have an incentive to make the system misbehave.
Identify what unintended consequences would result from the system making a mistake, and assess the likelihood and severity of these consequences.
Build a rigorous threat model to understand all possible attack vectors.

# Summary

- AI is not a silver bullet, but it could help tackle some of the world's most challenging social problems.
- The development of AI is creating **new opportunities to improve the lives of people** around the world, from business to healthcare to education.
- AI tools and techniques **can be misused** so **principles for their use** must be established.

Crucial to being able to question, understand, and trust AI systems.
Reflects domain knowledge and societal values.

It is essential to consider and address the security of an AI system before it is widely relied upon in safety-critical applications.

Fairness | Interpretability | Privacy | Security

AI may perpetuate and aggravate existing prejudices and social inequalities,
affecting already-vulnerable populations and amplifying existing cultural prejudices.

Although there may be enormous **benefits** to building a model that operates on sensitive data, it is essential to consider the **potential privacy implications** in using sensitive data.

# Thanks!

## Questions?

WOMEN IN DATA SCIENCE
DHAHRAN

# From Smart Cities to Social Cities: Smart Cities at the Service of Social Good

WOMEN IN DATA SCIENCE – DHAHRAN, SAUDI ARABIA

ANTONELA TOMMASEL

ISISTAN, CONICET-UNICEN, ARGENTINA

CONICET

ISISTAN