Case Studies : Safety 2

Proposed Answers

Case 1: Edge Cases - How artificial intelligence is transforming professional meetings

Scenario
As a software engineer developing cutting-edge solutions for the future workplace, your goal is to seamlessly connect workers with their colleagues for effective collaboration, regardless of their location, device, or time zone. Your role involves redesigning the digital workspace to foster seamless interaction and collaboration among team members.

To achieve this, you have enhanced virtual meetings to replicate the immersive experience of in-person interactions. You have implemented advanced camera setups that leverage AI to detect who is speaking and where they are facing. This ensures that virtual participants always see the speaker's face, even when they turn towards a whiteboard or someone in the room. Additionally, AI streamlines meeting tasks such as transcription and scheduling follow-ups, ensuring that discussions are accurately captured and their context is preserved.

Edge case analysis
For each of the three categories Global Reach, Mass Adoption, and Longevity:

- Identify at least one edge case that represents a challenge or an opportunity for your virtual meeting solution;
- Describe how you could change the design of your software to take it into account.

Proposed answers (examples are provided for each category (global reach, mass adoption and longevity, many answers are possible):

I. **Global Reach**

A. Deploying in different countries means having a good enough support for languages other than English for voice to text features such as transcription. This can be difficult for some languages. Supporting a range of accents in English would also be necessary.

Design change: Anticipate and look for language models and tools for a range of languages that could be supported. An opportunity could also be to provide translation tools.

B. Different cultures have varying norms for communication and hierarchy. Eye contact is also interpreted differently in each culture. For instance, in Japan, there is a cultural emphasis on formality and hierarchy in business meetings and what is considered appropriate eye contact depends on the hierarchy among the persons involved.

Design change: Implement algorithms adapted to the local cultural nuances, such as emphasizing titles and formal language in countries like Japan, to ensure culturally appropriate communication in meeting transcripts and translations.

II. **Mass Adoption**

A. During a global event like a pandemic, remote work tools could see a surge in users. The system should be able to scale horizontally by adding more servers to handle increased load without downtime for users that already depend on the system.

Design change: Develop a robust, scalable architecture that can dynamically add server capacity and utilize cloud-based resources to handle sudden increases in user load, ensuring uninterrupted service and optimal performance.

B. As more organizations and individuals rely on the platform for daily work, any outages or disruption could have very severe consequences on the global economic and professional systems.

Design change: enhance platform's reliability with robust failover systems and backup mechanisms. You could as well ensure users have backup plans in place to mitigate the impact of potential disruptions.

C. If all meetings occur on this platform, users may develop digital fatigue and mental health concerns from prolonged screen time.

Design change: design the platform with features to mitigate fatigue, such as scheduling breaks and offering alternative communication modes (ex : audio only) to reduce screen time.

III. Longevity

A. The software could automate tasks such as note-taking, scheduling, and meeting management. This may reduce the need for certain roles currently occupied by humans, such as secretaries and administrative assistants, leading to job displacement. This effect could become more pronounced as the software continues to progress and incorporate more advanced features.

Design change: The features could be designed specifically to provide tools that are helpful to people in these roles. However that would not prevent a long term effect of job displacement. Being conscious of this effect, you would need to decide whether you want to keep these features or not - this is a dilemma, knowing that the features would help people who don't have access to administrative assistants...

B. Over time, vast amounts of recorded meeting data, including transcriptions and video, are accumulated. This will put pressure on the technical infrastructure and increase the risk that this data could be misused or accessed without proper authorization, leading to privacy violations.

Design change: Implement a data management strategy and infrastructure sized for 10 years of data, including data archival and data deletion mechanisms, making sure it complies with regulations in the deployed countries. Strengthen data encryption and implement rigorous access controls.

Case 2: The Ethics Canvas - VR holidays app

Scenario
RealVirtualHolidays is an innovative virtual reality app designed to revolutionize vacation. RealVirtualHolidays 's goal is to provide you with an exceptional vacation experience, minus the drawbacks such as missing flights, unresponsive staff, and unfavorable weather. In a world where carbon footprint matters, RealVirtualHolidays is the eco-friendly travel option.

Users start by selecting one of the vacation packages: domestic, international, or exotic. Upon registration, they'll meet with a confidential consultant. Their role is to assist users in creating the ideal VR vacation, tailored to their desires. Whether you seek great weather, friendly locals, tranquility, or even an exclusive VIP dance party within the Great Pyramid with your favorite celebrity, choose from a variety of destinations.

The Ethics Canvas and group-scale impacts
The goal of this exercise is to evaluate the **impacts this system can have on groups** by focusing on the following blocks in the Ethics Canvas: 2 "groups affected", 5 "worldview", and 6 "group conflicts" (right side of the canvas, the blocks from the the left and bottom sides have already been completed).
Follow the steps below and fill the corresponding blocks of the Ethics Canvas.

Step 1: Identify relevant stakeholder - Block 2 "Groups affected"

General question : Which groups might be affected by the technology?
Specific questions :

- Which groups are involved in the design, production, distribution and use of your product or service?
- Which groups might be affected by it? Are these work-related organizations, interest groups, etc.?

Step 2: Identify ethical impacts - Blocks 5 "Worldviews" and 6 "Group conflicts"

General question**:** What are the potential ethical impacts for these groups?
Specific questions:

- Worldviews: How might people's worldviews be affected by your product or service? Their ideas about consumption, religion, work, etc.?
- Group conflicts: How might group conflict arise or be affected by your product or service? Could it discriminate between people, put them out of work, etc.

Fill out the Ethics Canvas:

Note: the left and bottom sides of the canvas (blocks 1, 3, 4, 7, 8 and 9) have already been completed and are just here for the example, they are not a part of the case study.

Proposed answer:

<table>
<tr>
<td rowspan="2"><u>1.Individuals affected</u>
- Everyone who wants to go on vacation and who can afford the technology.
- Local population in touristic areas.
- Tourism and transport workers.
- Local business workers.</td>
<td><u>3.Behaviour</u>
People would reduce their travel.</td>
<td rowspan="2"><u>9.What can we do?</u>
- In order to reduce the negative impacts on the local economy, we could make the virtual shops in the app connected to real local ones.
- Involve locals and in particular cultural minorities in the construction of the virtual travels.
- Communicate on the positive effects of reducing transportation and overtourism on the natural environment and local communities</td>
<td><u>5.Worldviews</u>
Information about countries and their culture are selected by some stakeholders, which could be reductive, foster stereotypes or only show the bright side of the "visited" countries .

Some parts of the culture and experience of a country are lost through this type of tourism, like local food, weather, etc…</td>
<td rowspan="2"><u>2.Groups affected</u>
- Tourism-related professionals (agencies, hotels, restaurants, etc..)

- Transport industry (planes, buses, car rental agencies, etc.)

- Cultural minorities

- Environmental advocates

- VR headset vendors

- Cultural institutions (museums etc.)</td>
</tr>
<tr>
<td><u>4.Relations</u>
Social interactions between tourists and locals would be reduced.</td>
<td><u>6.Group Conflicts</u>
The product might trigger anger from tourism- and transport-related staff, as it is a direct threat to them.

It can also foster tensions in the local population between people who are losing their jobs and those who are happy with the reduction of tourism.

We could see social tensions between people who can afford real travel and those who can afford only the VR version (class conflicts).

Cultural minorities might be "erased" in the virtual visit, thus generating anger and potentially conflicts with other groups.</td>
</tr>
</table>

7.Product or Service Failure Potential private information leaks, such as where some people went on vacation.	**8.Problematic Use of Resources** - Construction of many VR headsets, which is requires lots of material resources and would generate e-waste - Energy consumption of the usage of the VR technology - On the other hand, a large reduction in travel (flights in particular) would actually reduce environmental impacts related to transport

Case 3: Systems thinking - Predictive policing

Context
Predictive policing is an application of algorithmic governance in which police activities, such as intervention and prevention, are supported by algorithms designed to collect and analyze data in order to predict individuals or geospatial areas with an increased probability of criminal activity (Meijer and Wessels, 2019). Based on statistical analysis of past data, an algorithm computes probabilities that crime happens in different areas. The police then use the output of the system to optimize the allocation of police patrols and decide where to send them.

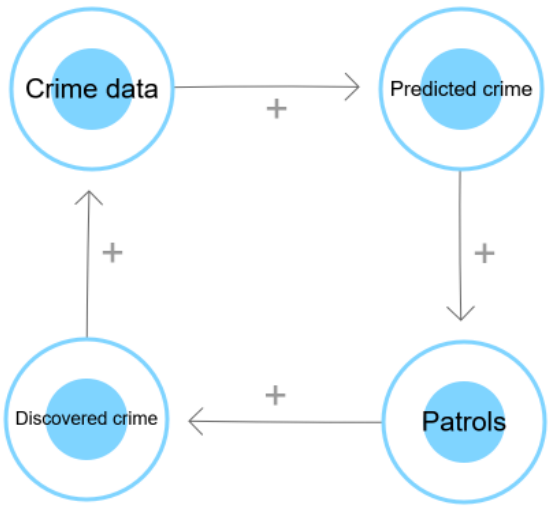
Scenario (fictional and simplified)
A Swiss Canton hears about predictive policing systems and wants to analyze pros and cons of such systems. They hire a team of system thinkers to get a complete in-depth analysis. This team produces a report including causal loop diagrams that you will analyze below.

Part 1: overall operation

The analysis team first provides a diagram presenting in simplified terms how the predictive policing system works overall.

They provide the causal loop diagram below, where the variables are the following:
- Crime data: quantity of data available on crimes in a considered zone
- Predicted crime: quantity of crimes predicted by the system for the considered zone
- Patrols: quantity of police patrols deployed in the considered zone
- Discovered crime: quantity of crimes discovered by police patrols in the considered zone

Question: Describe the diagram by detailing how the variables influence each other. Indicate the nature of the feedback loop (balancing or reinforcing).



Proposed answer

Since the diagram forms a loop, we can start by any variable we want:

- An increase in the quantity of police patrols deployed in a zone will lead to an increase in the quantity of crimes discovered by police patrols in that zone
- An increase in the quantity of crimes discovered by police patrols in the zone will lead to an increase in the quantity of data available on crimes in that zone
- An increase in the quantity of data available on crimes in the zone will lead to an increase in the quantity of crimes predicted by the system in that zone
- An increase in the quantity of crimes predicted by the system for that zone will lead to an increase in the quantity of police patrols deployed in that zone.
- This will again lead to an increase in the quantity of crimes discovered in that zone, etc.

This diagram illustrates a reinforcing feedback loop: an increase leads to an increase. Reinforcing feedback loops are also called "runaway" feedback loops since the behavior can get out of control.

Part 2: geographical effects

The analysis team has looked at the geographical effects of deploying a predictive policing system that would optimize the deployment of patrols in two geographical zones A and B.

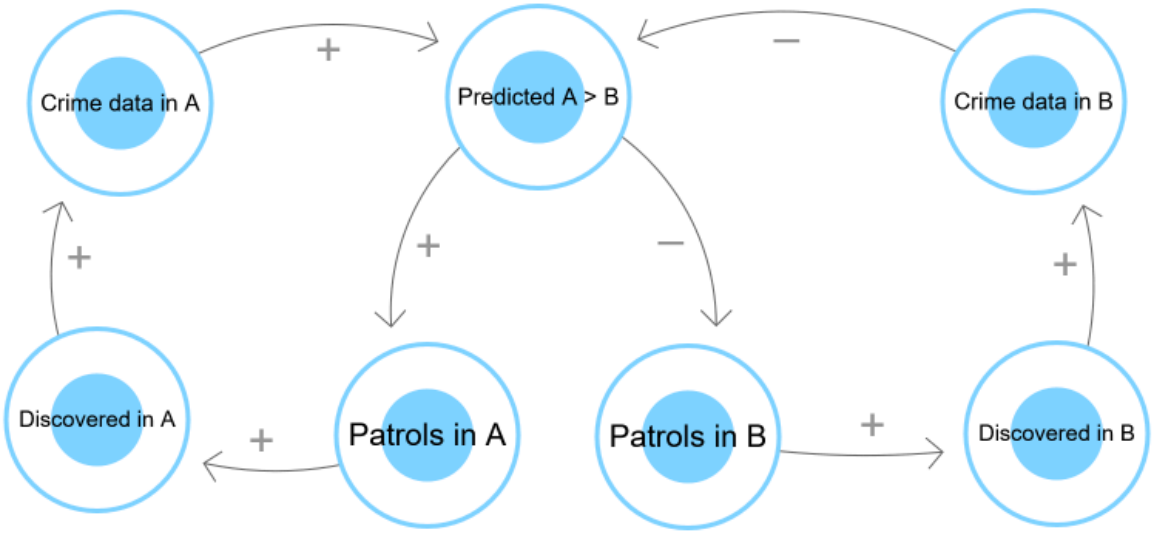
They provide the causal loop diagram below, where the variables are the following:

- Crime data in A/B: quantity of data available on crimes in zone A/B
- Discovered in A/B: quantity of crimes discovered by patrols in zone A/B
- Patrols in A/B: quantity of patrols deployed in zone A/B
- Predicted A > B: quantity of crimes predicted in zone A compared to zone B
  - An increase in Predicted A > B means that the crime rate is predicted to be higher in A than in B
  - A decrease in Predicted A > B means that the crime rate is predicted to be lower in A than in B

Questions:

a) Describe the two feedback loops by detailing how the variables influence each other. Indicate the nature of each feedback loop (balancing or reinforcing).

b) While the true crime rates are equal within zones A and B, for historical reasons zone A has been more policed in the past which means that more crime data is available for zone A. What would be the long term effects of deploying the predictive policing system? Use the diagram to identify the consequences. (NB: you can use loopy to make a simulation, https://ncase.me/loopy/v1.1/)



Proposed answer

a) Analysis of the diagram

Description of Loop A (left):
- An increase in Predicted A > B (crime rate predicted to be higher in A than in B) leads to an increase in Patrols A (more patrols are sent to zone A)
- The increase in Patrols in A results in an increase in Discovered in A (more crimes are discovered in zone A)
- The increase in Discovered in A leads to an increase in Crime data in A (more data is available on crimes in zone A)
- The increase in Crime data in A leads to an increase in Predicted A > B

Therefore Loop A is a reinforcing loop: an increase leads to an increase.

Description of Loop B (right):
- An increase in Predicted A > B (crime rate predicted to be higher in A than in B) leads to a decrease in Patrols B (less patrols are sent to zone B)
- The decrease in Patrols in B results in an decrease in Discovered in B (less crimes are discovered in zone B)
- The decrease in Discovered in B leads to an decrease in Crime data in B (less data is available on crimes in zone B)
- The decrease in Crime data in B leads to an increase in Predicted A > B

Therefore Loop B is a reinforcing loop: an increase leads to an increase.

In this diagram, both loops are reinforcing loops: an initial change will result in a reinforcement of the change.

b) Consequences of deploying the predictive system.

To be able to identify the consequences of deploying the predictive policing system using the diagram, we need an initial condition, i.e. an initial change in one of the variables that will trigger all the other changes. We are told that historically more data has been collected in zone A, which can be represented as an increase in the variable Crime data in A. As a result the system will predict more crime in zone A, and the two reinforcing loops will progressively reinforce this prediction and lead to the over-policing of zone A and the under-policing in zone B. As a result, residents in zone A might feel targeted, leading to potential strained police-community relations. Conversely, residents in zone B might experience a lack of police presence, potentially leading to unreported crimes and a sense of impunity for criminals.
It would be the opposite if more crime data had been collected in zone B.
Because of this data imbalance and despite true crime rates being equal, the introduction of the system results in discrimination against residents based on their location.

References:
Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C., & Venkatasubramanian, S. (2018). *Runaway Feedback Loops in Predictive Policing*. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 160–171. https://proceedings.mlr.press/v81/ensign18a.html

European Union Agency for Fundamental Rights (EU body or agency). (2022). *Bias in algorithms : Artificial intelligence and discrimination*. Publications Office of the European Union. https://data.europa.eu/doi/10.2811/25847

Meijer, A., & Wessels, M. (2019). *Predictive Policing : Review of Benefits and Drawbacks*. International Journal of Public Administration, 42(12), 1031-1039. https://doi.org/10.1080/01900692.2019.1575664

Part 3 [Optional]: community trust and reported crime

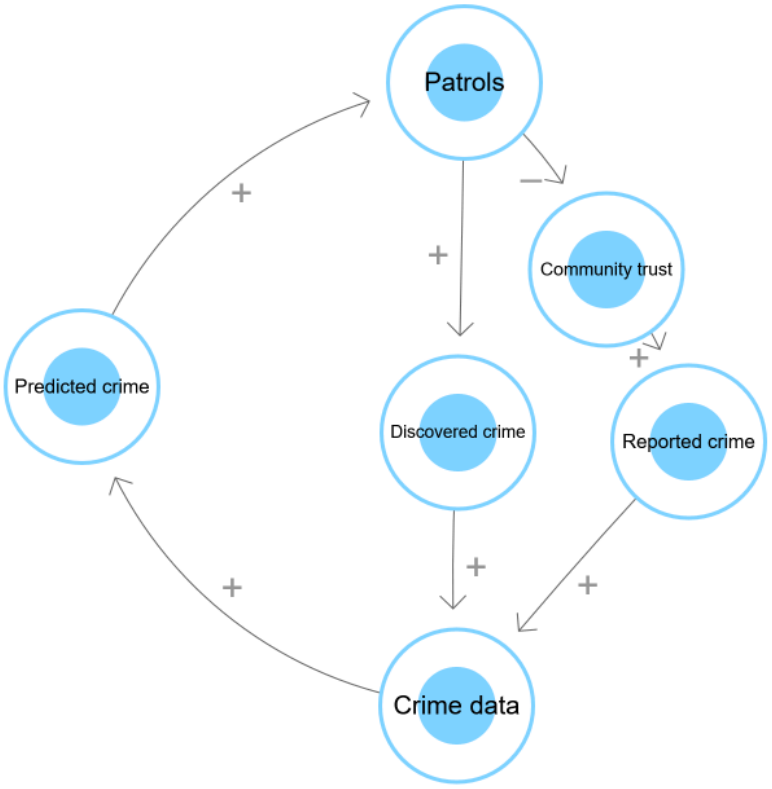
The analysis team has looked at the effect of police presence on the confidence that the community has in the police and the quantity of crime that the population reports itself (e.g. theft reported by victims).

They provide the causal loop diagram below, where the variables are the following:

- Patrols: quantity of police patrols deployed
- Discovered crime: quantity of crimes discovered by police patrols
- Community trust: confidence that the population has in the police
- Reported crime: quantity of crimes reported by the population to the police
- Crime data: quantity of data available on crimes
- Predicted crime: quantity of crimes predicted by the predictive policing system

Question: The diagram presents both an inner feedback loop (involving discovered crime) and an outer feedback loop (involving community trust and reported crime). Describe these feedback loops by detailing how the variables influence each other. Indicate the nature of each feedback loop (balancing or reinforcing).

Important note: research on the effects of police presence on community trust indicates a complex relationship with various factors that depend on local contexts, the causal loop diagram in this exercise is an extremely simplified depiction provided for pedagogical purposes only.



Proposed answer

Description of the inner loop:

- Patrols ↗️
- Discovered crime ↗️
- Crime data ↗️
- Predicted crime ↗️
- Patrols ↗️

Therefore the inner loop is a reinforcing loop: an increase leads to an increase.

Description of outer loop:

- Patrols ↗️
- Community trust ↘️
- Reported crime ↘️
- Crime data ↘️
- Predicted crime ↘️
- Patrols ↘️

Therefore the outer loop is a balancing loop: an increase leads to a decrease.