

# Case Studies: Conclusion

## Proposed Answers

### Case 1: Be My Eyes - Ethics Canvas

#### Scenario:

Be My Eyes is a mobile application designed to connect visually impaired individuals with sighted volunteers through a live video call. The app allows visually impaired users to request assistance with tasks that require vision, such as reading labels, navigating new environments, or identifying objects. Volunteers provide real-time guidance, offering an easy and accessible way to bridge the gap between sighted and visually impaired communities. This app uses community and technology to help improve the independence and quality of life for visually impaired people.

Source: <https://www.bemyeyes.com/>

#### Task:










The goal of this exercise is to evaluate the impacts this system has and to organize them by filling the Ethics Canvas.

Here is a reminder of the 3 stages of the strategy:

- Stage 1: Identify relevant stakeholders**  
 Assessment question: Who might be affected by the technology?  
 → fill out blocks 1 and 2
- Stage 2: Identify ethical impacts**  
 Assessment question: What are the potential ethical impacts for these people and groups?  
 → fill out blocks 3, 4, 5, 6, 7 and 8
- Stage 3: Discuss remedial actions**  
 Assessment question: How can we address these ethical impacts?  
 → fill out block 9

The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

Ethics Canvas v1.8 - ethicscanvas.org © ADAPT Centre & Trinity College Dublin & Dublin City University, 2017.

Ethics Canvas		Project Title:	Date:
<b>Individuals affected</b> Identify the types or categories of individuals affected by the product or service, such as men/women, user/non-user, age-category, etc.	<b>Behaviour</b> Discuss problematic changes to individual behaviour that may be prompted by the application e.g. differences in habits, time-schedules, choice of activities, people behaving more individualistic or collectivist, people behaving more or less materialistic.	<b>What can we do?</b> Select the four most important Ethical impacts you discussed. Identify ways of solving these impacts by changing your project's product/service design, organisation, or by providing recommendations for its use or spelling out more clearly to users the values driving the design.	<b>Worldviews</b> Discuss how the general perception of somebody's role in society can be affected by the project.
 <b>1</b>	 <b>3</b>	 <b>4</b>	 <b>5</b>
	<b>Relations</b> Discuss problematic differences in individual behaviour such as differences in habits, time-schedules, choice of activities, etc.		<b>Group Conflicts</b> Discuss the impact on the relationships between the groups identified, e.g. employers and unions.
	 <b>4</b>	 <b>6</b>	 <b>2</b>
<b>Product or Service Failure</b> Discuss the potential negative impact of your product or service failing to operate as intended, e.g. technical or human error, financial failure/ receivership/acquisition, security breach, data loss, etc.		<b>Problematic Use of Resources</b> Discuss possible negative impacts of the consumption of resources of your project, e.g. climate impacts, privacy impacts, employment impacts etc.	
 <b>7</b>		 <b>8</b>	

The Ethics Canvas is adapted from Alex Osterwalder's Business Model Canvas. The Business Model Canvas is designed by Business Model Roundry AG. This work is licensed under the Creative Commons Attribution-Share Alike 3.0 unported license. To view a copy of this license, visit <https://creativecommons.org/licenses/by-sa/3.0/>. To view the original Business Model Canvas, visit <https://strategyzer.com/canvas>

You can find the [original PDF template of the Ethics Canvas here](#) and below is a corresponding table that you can fill out:

<b><u>1. Individuals affected</u></b>	<b><u>3. Behaviour</u></b>	<b><u>9. What can we do ?</u></b>	<b><u>5. Worldviews</u></b>	<b><u>2. Groups affected</u></b>
	<b><u>4. Relations</u></b>		<b><u>6. Group Conflicts</u></b>	
<b><u>7. Product or Service Failure</u></b>			<b><u>8. Problematic Use of Resources</u></b>	

**Proposed answer**

<b><u>1. Individuals affected</u></b> Visually impaired individuals who use the app for assistance in daily tasks  Sighted volunteers who offer their help through the app.  Elderly users who may be less familiar with technology but may have the need to use the app for assistance.  Family members or friends who might previously have assisted visually impaired individuals.	<b><u>3. Behaviour</u></b> Visually impaired users may become dependent on the app for tasks they would otherwise attempt to do independently  The app might change how users and volunteers interact socially. For instance, visually impaired users may increase their use of virtual assistance over in-person help, leading to a decrease in social interactions leading to potentially more isolation.  Users might adjust their schedules around the	<b><u>9. What can we do ?</u></b> Provide ethics training for all volunteers to ensure they are well-informed.  Maintain a dedicated team of volunteers available for immediate assistance in urgent situations.  Include a team of health professionals or disability support workers to offer guidance or direct help in complex cases.  Implement a client service system for reporting incidents or a moderation service for	<b><u>5. Worldviews</u></b> The app might reinforce the view of technology as a primary solution for disability support, potentially overshadowing the importance of human relationships and in-person assistance.  The app might reinforce the idea that visually impaired individuals need constant external assistance, affecting perceptions of their autonomy.	<b><u>2. Groups affected</u></b> Disability support workers: group of people who are specialized in helping at home/in person visually impaired people  Visually impaired communities  Inclusion advocate politic groups/legislators  The app's developing team
--	---	---	--	--

	<p>availability of volunteers, leading to changes in their routines or social life.</p>	<p>conflicts involving both volunteers and users.</p>		
	<p><b>4. Relations</b> Users might start relying more on virtual interactions through the app rather than seeking help from family, friends, or local community members, altering their social habits.</p> <p>Users might avoid tasks that require assistance outside of the app's scope or at times when volunteers are not available, limiting their daily activities, relationships outside of the app, and independence.</p> <p>The app could unintentionally create power dynamics where volunteers feel overly indispensable to visually impaired users, potentially leading to unbalanced relationships.</p>	<p>Implement strong data security measures and clear guidelines on data use to protect user information.</p> <p>Promote the app as a supplement to, rather than a replacement for, in-person assistance to maintain social bonds and prevent isolation, and the development of complementary skills to maintain independence.</p>	<p><b>6. Group Conflicts</b> While the app aims to empower users, it could inadvertently lead to a dependency on volunteers, which might conflict with broader societal efforts to make environments and products more accessible without assistance.</p> <p>The app could create tension between volunteers and professional caregivers/disability support workers, as it might be seen as undermining the need for paid care services.</p>	
<p><b>7. Product or Service Failure</b> If the app fails during critical moments, or volunteers are not available to help at some point, visually impaired users might be left without needed assistance, leading to dangerous situations.</p> <p>If the access authorizations are not monitored properly by the app, it could lead to important privacy breaches for the visually impaired individual (volunteers could witness parts of the disabled person's life that they aren't supposed to see)</p> <p>Errors in volunteer guidance could result in accidents or harm to the user, raising concerns about the reliability of the service.</p> <p><i>There may be (fairness) issues with how fairly the app distributes volunteer help, possibly leading to some users receiving more assistance than others.</i></p>		<p><b>8. Problematic Use of Resources</b> Relying on volunteers raises ethical questions about the sustainability of unpaid labor and the long-term viability of the service.</p> <p>While the app is free for users, the cost of maintaining and scaling the platform could become an issue, potentially leading to monetization strategies that can impact access or service quality/usability for people with fewer resources.</p>		

## Case 1: Emotion Cancelling AI - Digital Ethics Canvas

### Scenario:

Softbank's "Emotion Canceling AI" is a tool designed to support call center staff by altering the tone of angry or emotional customer voices in real-time. The system uses advanced voice recognition and emotion-detection technology to identify aggressive or frustrated tones in callers' voices and modifies the audio before it reaches the call center staff. This creates a neutral, less confrontational interaction, aiming to reduce stress and improve focus for employees handling challenging customer situations.

Source: [Softbank plans to cancel out angry customer voices using AI \(June 2024\)](#). *Ars Technica*.

**Task**

Your goal is to **apply the Digital Ethics Canvas** to evaluate the ethical implications of this system and identify ways in which ethical risks could be limited.

Follow these steps:

1. **Describe briefly:**
  - **The solution:** Briefly describe the software product. What services/features does it offer? How does it work?
  - **The context:** Who develops the software? Who uses it? Which other stakeholders are concerned?
2. **Evaluate the benefits:** list 3-4 expected benefits (1 sentence each).
3. **Evaluate the risks:**
  - A. Identify **potential risks** using the five ethical lenses: describe each risk you identify (1-2 sentences each).
  - B. **For each of the risks you identified, assess its level** by evaluating:
    - The **probability** of occurrence (low/ mid/ high) + 1 sentence justification.
    - The **severity** of impacts (low/ mid/ high) + 1 sentence justification.
    - Then determine the **overall level of risk** (low/ mid/ high) using the risk matrix from the cheatsheet:

		Severity		
		low	mid	high
Probability	low	low	low	mid
	mid	low	mid	high
	high	mid	high	high

4. **Reduce the risks:** For each identified risk, propose **one mitigation measure** (briefly describe it: 1-2 sentences).  
 You can think about:
  - o Technical mitigations, that the companies developing the product could (or should have) put in place
  - o Organizational mitigations, that rely on human ways of reducing the risks (e.g. that users can put in place)

Here is the original [template for the digital ethics canvas](#) and you will find below a corresponding table that you can fill out.

**Proposed answer**

**Solution:** We are analyzing Softbank’s ‘Emotion Canceling AI’, a real-time voice-processing tool that modifies the audio of a caller’s voice to sound more neutral for the call center staff. The goal of the system is to reduce stress and help call center staff manage difficult interactions.

**Context:** The tool is developed by Softbank. The tool is used by call centers. The main stakeholders involved are call center operators, managers and IT teams, and of course customers who call the center.

**Benefits:**

- **Reduced stress:** By neutralizing the tone of angry or emotional customer voices, the system can reduce the emotional strain on the call center staff and mitigate the psychological toll of handling emotionally charged conversations.

- **Improved focus during calls:** With less confrontational tones to handle, employees can concentrate better on resolving customer issues effectively.
- **Increased employee retention:** By making challenging calls less stressful, the system may improve job satisfaction and reduce turnover rates among call center staff.
- **Improved customer service quality:** By improving the ability of the call center staff to maintain calm and professional interactions, the system can help deliver a better overall experience for customers.

WELFARE	
RISK	MITIGATION
<ul style="list-style-type: none"> <li>● Altering customer voices may cause misunderstandings due to the change in tone [MEDIUM RISK]</li> <li>● Errors from the system may alter voices in ways that are inaudible or can generate hearing pain [MEDIUM RISK]</li> <li>● Prolonged exposure to artificially neutralized conversations might make the tasks more dull and repetitive and reduce job satisfaction [LOW RISK]</li> </ul>	<ul style="list-style-type: none"> <li>● Allow employees to access both the original and altered voices for verification if needed.</li> <li>● Provide a reporting mechanism for employees to signal errors in order to investigate and correct</li> <li>● Allow employees to balance the use of the system with personal judgment when handling calls.</li> </ul>
FAIRNESS	
RISK	MITIGATION
<ul style="list-style-type: none"> <li>● The system may not equally detect or adjust to emotions across different accents or cultural speech patterns [HIGH RISK]</li> <li>● Some types of voices may be altered in biased way (e.g. high pitched voices, even when not angry, could be systematically lowered) [MEDIUM RISK]</li> </ul>	<ul style="list-style-type: none"> <li>● Train and test the system on diverse accents, languages, and cultural speech styles</li> <li>● Monitor the fairness of the system with fairness metrics applied over usage data</li> </ul>
AUTONOMY	
RISK	MITIGATION
<ul style="list-style-type: none"> <li>● Employees may feel constrained if they are required to use the tool without the option to disable it [MEDIUM RISK]</li> <li>● The system's real-time alterations may reduce employee's ability to rely on their own judgment in emotionally charged situations, leading to a loss of autonomy in how they manage interactions. [MEDIUM RISK ]</li> <li>● Customers might feel mistreated if they discover their voices are being altered without their consent [HIGH RISK ]</li> </ul>	<ul style="list-style-type: none"> <li>● Provide employees with the option to enable, disable, or adjust the system's settings based on their needs/preferences during calls</li> <li>● Encourage a workplace culture where employees can use their own judgment alongside automated tools, allowing them to balance their skills with support from the system</li> <li>● Add transparency features, such as notifying customers their voices may be moderated during calls</li> </ul>
PRIVACY	
RISK	MITIGATION

<ul style="list-style-type: none"> <li>• Voice data is personally identifiable data that can also be considered sensitive (e.g. part of biometric information) [HIGH RISK]</li> <li>• Identifying customer emotions can lead to the disclosure of information the customers might consider private [HIGH RISK]</li> <li>• The perception of being monitored or emotionally analyzed might raise privacy concerns among customers, even if data is not stored [HIGH RISK]</li> </ul>	<ul style="list-style-type: none"> <li>• The system should work in real time without storing or transmitting voice data (unless there is explicit consent from the customer)</li> <li>• Data about the emotions detected by the system should not be stored or transmitted or used for other purposes</li> <li>• Provide clear communication to customers about how the tool operates and ensure transparency regarding its functionality.</li> </ul>
SUSTAINABILITY	
RISK	MITIGATION
<ul style="list-style-type: none"> <li>• Continuous emotion processing could increase energy consumption, water consumption and carbon footprint [HIGH RISK]</li> <li>• The infrastructure required to process the voice and detect emotions involves equipment with high embodied carbon footprints [HIGH RISK]</li> </ul>	<ul style="list-style-type: none"> <li>• Optimize algorithms and choose infrastructure for processing with minimal energy , water consumption and carbon footprint.</li> <li>• Maximize the utilization of the infrastructure (e.g. use it for other ML tasks when the emotion detection is not used) and maximize the lifespan of the hardware by avoiding too frequent renewal</li> </ul>

For the analysis of the risk level here is the detail for the WELFARE risk "Errors from the system may alter voices in ways that are inaudible or can generate hearing pain":

- The probability of this issue occurring is LOW as the system includes safeguards that prevent extreme sound distortions, and it has been thoroughly tested and evaluated with a minimized error rate.
- The impact of this risk is HIGH since if errors do occur, they could directly affect the quality of the service if the voice becomes inaudible or employees' well-being and health if causing hearing pain, which is highly problematic.
- As a result, using the risk matrix above, we obtain a global risk level of MEDIUM which reflects that while the consequences of such errors are impactful, they are not very likely to occur.

We have applied the same process to all the risks in the canvas to attributed them an overall risk level (low / mid / high) combining probability (low / mid / high) and severity of impacts (low / mid / high) and using the provided risk matrix:

		Severity		
		low	mid	high
Probability	low	low	low	mid
	mid	low	mid	high
	high	mid	high	high

Except where otherwise noted, the content of this document is licensed under a Creative Commons Attribution 4.0 International License (CC BY)

<http://creativecommons.org/licenses/by/4.0/>

