

ResponsibleX

Mohammad Tahaei @Nokia Bell Labs

Lead user researcher

Qualitative: Interviews, cognitive walkthroughs, usability tests

Overview

Business Problem: Need to **comply with new AI regulations, build trust** in AI products, and promote **brand as trustworthy**.

Product Idea: Introducing an **interactive toolkit** into **AI developer workflows** to **prompt** consideration of decision and model ramifications (**Responsible AI / Ethics by Design**)



Context

- Old method: paper/PDF **checklists** with **top-down** approach.
- ResponsibleX: **interactive**, promotes **ethical AI**, considers bias, privacy, security, harm **in the design**.
- Aimed to **replace** checklists, but **lacked user research**.



Transitioning to Interactive Gamified Tool

Screenshot: Ethical AI Model Checklist

- Members of stakeholder groups, including demographic groups
 - Consider whether the system will impact human rights
 - Consider whether these uses or applications should be prohibited
- Expected deployment contexts (e.g., geographic regions, time periods)
- Expected stakeholders (e.g., people who will make decisions about system adoption, people who will use the system, people who will be directly or indirectly affected by the system, society), including demographic groups (e.g., by race, gender, age, disability status, skin tone, and their intersections)
- Expected benefits for each stakeholder group, including demographic groups
- Relevant regulations, standards, guidelines, policies, etc.
- 1.1.b Scrutinize resulting system vision for potential fairness-related harms to stakeholder groups, considering:
 - Types of harm (e.g., allocation, quality of service, stereotyping, denigration, over- or underrepresentation)
 - Tradeoffs between expected benefits and potential harms for different stakeholder groups
 - Consider who the system will give power to and who it will take power from
 - Consider which expected benefits you are willing to sacrifice to mitigate potential harms
- 1.1.c Revise system vision to mitigate any potential harms; if this is not possible, document why, along with future mitigation or contingency plans, etc., and consider aborting development
- 1.2 Solicit input and concerns on system vision
- 1.2.a Solicit input on system vision and potential fairness-related harms from diverse perspectives, including:
 - Members of stakeholder groups, including demographic groups
 - Consider whether any stakeholder groups would prefer that the system not exist or not be deployed in all contexts, what alternatives they would prefer, and why
 - Domain or subject-matter experts
 - Team members and other employees
- 1.2.b Revise system vision to mitigate any potential harms; if this is not possible, document why, along with future mitigation or contingency plans, etc., and consider aborting development
- 1.3 Escalate potential harms involving sensitive, premature, dual, or adversarial uses or applications to leadership

Screenshot: ResponsibleX Prototype

C Interacting with the guidelines

Your AI System: Face recognition system

14 development cards left

Harm reporting

Provide mechanism(s) for incentivizing reporting of system harms.

Has this been done?

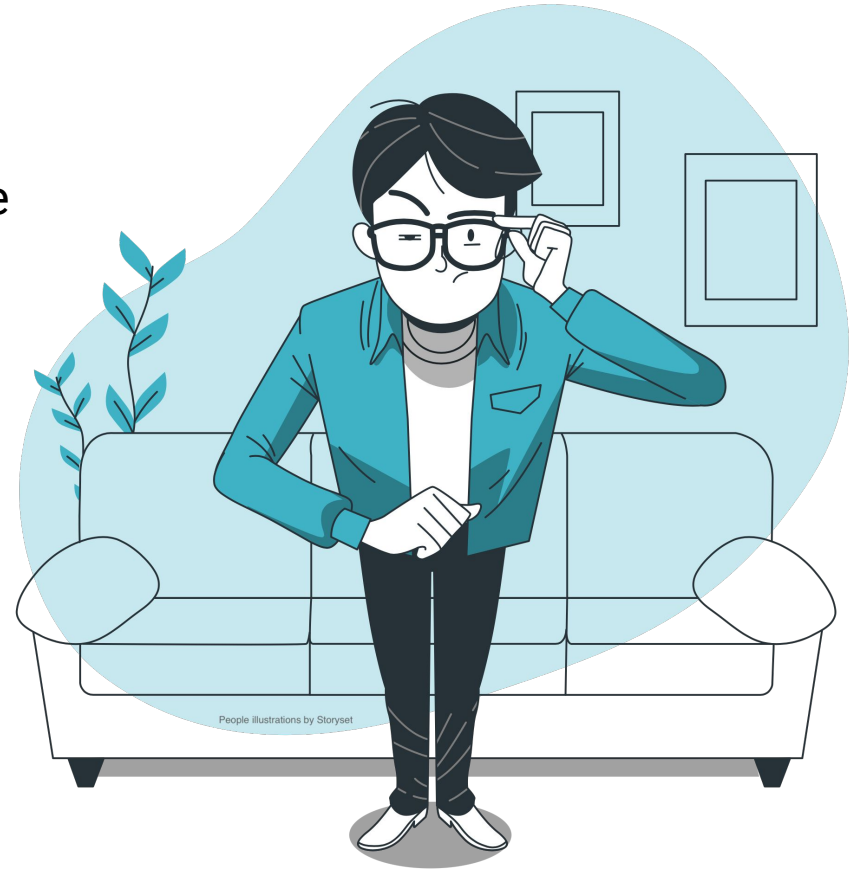
✗
✓

Exit game
Download report

Edit last card

Research Questions

- How can the tool be improved to ensure **seamless adoption** by AI developers?
- Which **features** are currently **absent**?
Which features are **redundant**?
- Is it advisable for ResponsibleX to **replace** traditional checklists?



Constraints

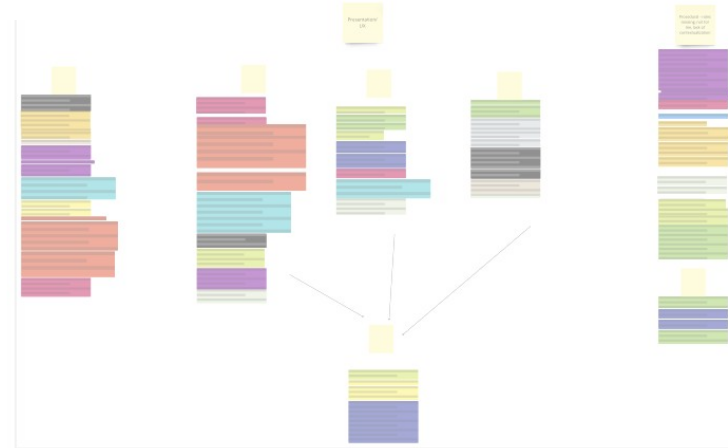
- Timeline: **2-3 weeks**
- Recruitment challenges: **expensive & busy AI developers**
- **Exploratory** nature and limited previous user research



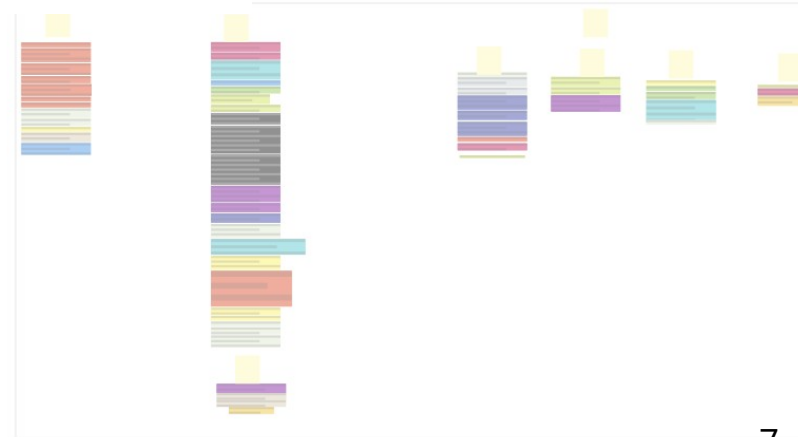
Method

- **Interviews and usability tests** with 14 AI developers, testing checklist & ResponsibleX.
- 5 additional **cognitive walkthroughs** to dig deeper into gaps and issues.
- Synthesized using **affinity diagrams** and extracted user needs.
- **Prioritized features** based on time, effort, and significance in **collaboration** with the designer, developer, manager, and project owner.

Checklist



ResponsibleX



High-Level Findings

- 86% believe the tool **boosts awareness and self-learning** for responsible AI.
- SUS rates ResponsibleX's **usability higher** than the checklist (73 vs. 42).
- 71% want clarity on **project specifics, stages, and roles**.
- 57% want the **checklist as an addition**.
- Issues with **unclear terms**.
- Importance of **remembering past interactions** highlighted in the walkthroughs.

"made me reflect on my previous choices and how I would describe my decisions when I had to develop the system"

"would be helpful if the guidelines were tailored to the specific tasks"



Product Impact

All high priority research findings were applied!

Responsible AI Guidelines

Addition of roles, project type & stage

System name
Face recognition system

Short summary of your system
Biometric authentication method to unlock a smartphone

What system phase you were involved in?
Development Deployment Use

When the AI System was designed and developed

Your role
Your role: Decision maker or Advisor Engineer or Researcher Designer
Your e-mail: researcher@mail.com
AI/ML Engineer, AI/ML Researcher, Data Scientist, Software Engineer, UX Engineer or UX Researcher

Your AI System
Upload your .json card file: cards-face-recognition-system.json

System name: Face recognition system
System summary: Biometric authentication method to unlock a smartphone
Your role: Engineer or Researcher
Your e-mail: researcher@mail.com

AI system phase you were involved in: Development

A **B**

Clarification of terms

Based on what you know

System evaluation

Report evaluation metrics for various groups based on factors such as age, gender, and ethnicity.

Which ones have been reported and how?
We created datasets for performance evaluation in terms of precision and recall for face detection and tracking. We used open-source tools like FiftyOne for the evaluation.

What else could have been done?

Save Skip

Importing past interactions

Actions to improve development by 21%
See the actions in PDF and save them for later use in JSON

Save actions → cards-face-recognition-system.pdf

cards-face-recognition-system.json **C**

Responsible AI Guideline Summary

Face recognition system

About
AI system: Face recognition system
AI system phase: Development
Most recent label update: 2023-5-30
Updated by: Researcher: researcher@mail.com

System description
Biometric authentication method to unlock a smartphone

Responsible AI blind spots
Unnoticed biases and potential blind spots
21%
3 actions to take 11 taken

Actions to take
critical pressing done inapplicable

- Uses
- Oversight
- Team
- Harms
- Data
- System

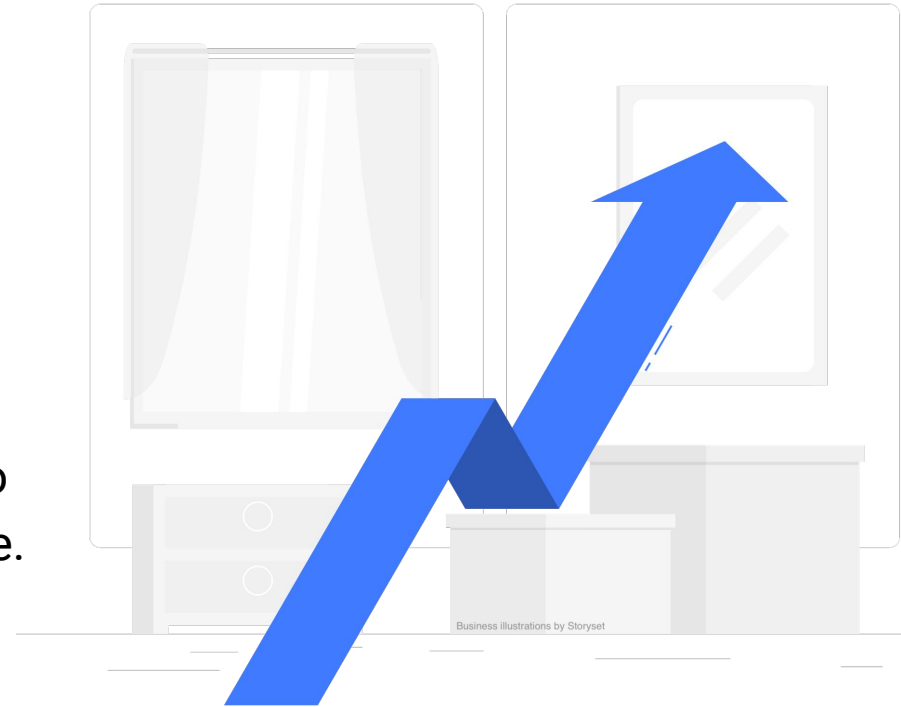
Identify potential harms associated with the intended uses.
Document all system components, including the AI models, to enable reproducibility and scrutiny.
Provide mechanisms for interpretable outputs and auditing.

D

Addition of hierarchy & categories to the output

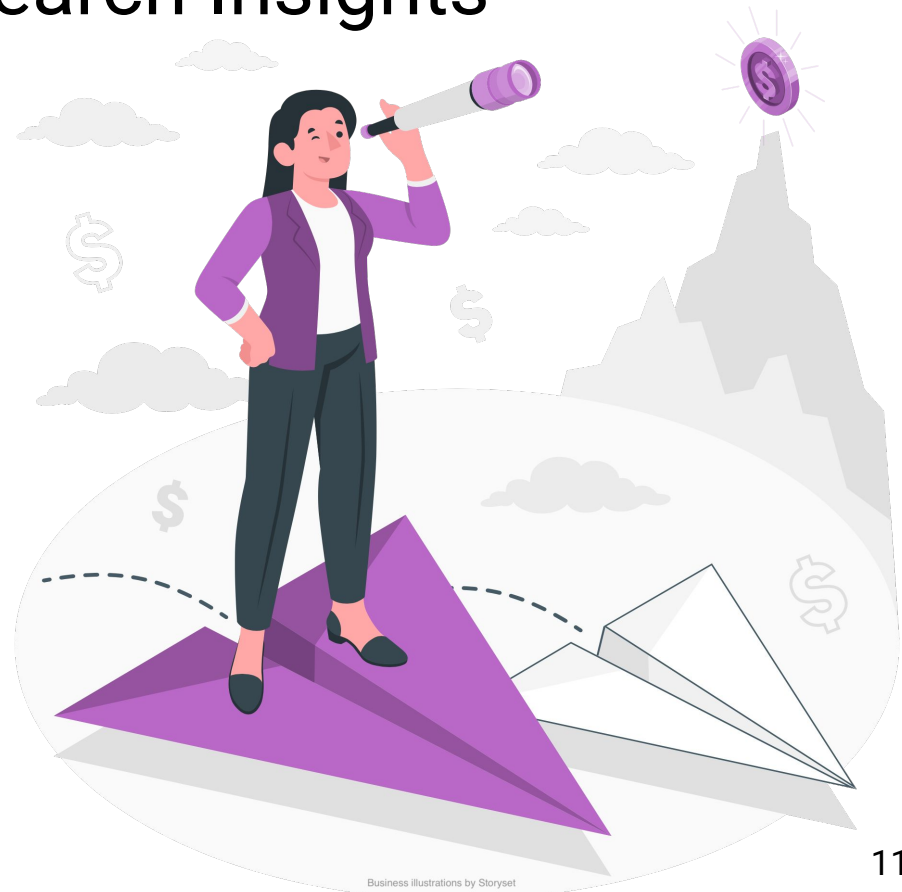
Business Impact

- **Adoption** by **2 business units**.
- External impact as a **research publication** and a **corporate blog post** for **brand promotion**.
- Strategically **repositioned** as a **complement product** and **method** to the checklist rather than a substitute.
- Filing a **patent**



Next Steps Based on Research Insights

- **Broaden** ResponsibleX's reach to additional domains (scalability & adaptability).
- Integrate a **user interaction-based recommender system** to enhance initial user engagement for newcomers.
- Create an **executive dashboard**, furnishing key insights into AI developers' on-the-ground practices to elevate business value & impact.



Reflections - What I Learned?

- **Engage** with **external users** beyond the company at the initial stages to explore **open-sourcing** opportunities.
- **Optimize recruitment efficiency** by creating a dedicated internal platform for usability studies and developing strategies to incentivize participation.
- Reframing as internal partnerships and **collaboration instead of recruitment.**



Growth Illustrations by Storyset

What Do Former Teammates Have to Say?

[Quotes from LinkedIn recommendations]

“efficient in **time management**, allowing him to keep the research projects **on track** and **deliver** the results **on time**, without losing the quality.” [Alisa Frik, Senior UXR]

“**highly professional and amiable** colleague . . . was involved in a number of projects, **worked with a colleagues at varying levels of seniority** and experience, and **acted as a mentor** for junior colleagues.” [Louise Evans, Research Manager]

“easily one of my **most productive** students. He has an **excellent eye for interesting research problems** and the **attention to detail** needed to realize them.”

[Kami Vaniea, Associate Professor]

Contact

<https://mohammad.tahaei.com/>

<https://www.linkedin.com/in/tahaei/>

mohammad@tahaei.com