

# A Co-Designed Checklist for Well-Calibrated Trust in Human-AI Collaboration

Dogtooth – Deliverable 3

## 1 Introduction

Artificial intelligence is being integrated into healthcare workflows to support diagnosis, decision-making, and treatment planning. However, workers must constantly decide whether the output should be trusted and to what extent. Overtrust in AI may lead to unsafe decisions, while insufficient trust may reduce the practical benefits AI systems can provide in healthcare workflows.

Our scoping review, “*Drivers of Well-Calibrated Trust in Human-AI Collaboration: A Scoping Review of Task Dynamics and Regulatory Frameworks*”, explored factors influencing trust and trust calibration in human-AI collaboration in healthcare contexts. The article identified some recurring themes such as AI reliability, transparency, uncertainty, the need of human oversight, user expertise, and responsibility.

Despite growing research on trust in AI, translating theoretical findings into practical decision-support tools remains challenging. Users may struggle to critically evaluate AI outputs without structured guidance.

Taking our review into account, the aim of this co-designed report is to transform our theoretical findings into a tool by structuring a checklist intended to support calibrated trust in AI-assisted healthcare tasks.

## 2 From Scoping Review to Checklist

The initial checklist was developed based on the four main thematic areas identified in the scoping review:

- AI system characteristics
- Human factors
- Task characteristics
- Organizational and contextual factors

These themes were transformed into checklist categories to help stakeholders evaluate whether trust in an AI system is appropriate in a given context.

The first dimension, *AI system characteristics*, highlighted concerns related to reliability, explainability, uncertainty communication, bias, and the possibility of AI hallucinations or misleading outputs. Studies highlighted explainability, communicated limitations, and uncertainty indicators as key drivers of calibrated trust.

The second dimension, *human factors*, focused on the role of user expertise, familiarity with AI systems, confidence in identifying AI mistakes, and attitudes towards AI's help in decision-making. Trust calibration depends not only on AI performance, but also on the user's ability to critically interpret outputs.

*Task characteristics* also represented an important area influencing trust. The research highlighted that trust in AI may change depending on task complexity, time sensitivity,

risk level, and the possibility of verifying AI outputs. High-risk situations increased the need for human validation and cautious AI reliance.

Finally, *organizational and contextual factors* accentuate the importance of workflow integration, regulations, accountability protocols, and the distribution of responsibility in AI-assisted environments. Several studies suggested that trust in AI is also shaped by the context in which the system is implemented, including the presence of clear guidelines and policies.

A co-design-oriented approach was adopted to incorporate stakeholders' perspectives into the process.

The interviews were therefore structured around these four dimensions and designed to understand how users perceive trustworthiness, overreliance, human validation, accountability, and AI-assisted decision-making.

The first version of the checklist focused on:

- Reliability
- Explainability
- Uncertainty in Communication
- Human Oversight
- Verification of Output
- Accountability
- Risk Of Overreliance

The resulting feedback then supported the refinement of the preliminary checklist and contributed to the development of a more practical and useful tool.

## 3 Method

### 3.1 Participants Recruitment

To support the co-design process, stakeholder feedback will be collected from participants coming from at least two different stakeholder groups.

Participants include 11 people in total with interest or experience in healthcare decision-making contexts, as well as individuals with technical familiarity with AI systems.

The participants were selected on a voluntary basis through personal connections within our group and they include:

- Med students
- Dentists
- Biomedical engineers
- Dental hygienists
- Workers dealing with AI daily, such as computer scientists, business analysts and financial economists.

The reason behind the inclusion of participants with different backgrounds is to provide different perspectives regarding our research question: “*What factors determine whether a*

*worker trusts AI to perform a task, and when is that trust well calibrated?"* and to make sure the checklist could be applicable to any field of work interacting with AI-assisted systems.

### 3.2 Procedure

The development process of the checklist was structured as a co-design-oriented workflow founded in the results of our scoping review. The initial phase consisted in analyzing the recurring factors influencing trust in healthcare contexts that surfaced in our initial review. These findings were then used to guide the development of both the raw checklist structure and the stakeholder interview.

In short interviews of approximately 20 minutes, participants were asked to answer questions regarding perceptions of trust in AI systems, factors influencing trust calibration, and feedback regarding the checklist. The interviews combine open and structured questions to encourage both reflective and practical feedback and to maintain engagement. Open-ended responses were qualitatively reviewed to identify recurring themes and concerns relevant to trust calibration.

The collected product was then analyzed to identify recurring issues, suggestions, and limitations of the first checklist structure.

### 3.3 Checklist Structure

The checklist was designed following the four main dimensions identified in the scoping review: AI system characteristics, human factors, task characteristics, and organizational and contextual factors.

The first section investigates the participants' familiarity and experience with AI systems.

The second section investigates trust calibration across the four identified dimensions.

The last section gives the participants the opportunity to help with the deliverable by giving feedback regarding the structure and the content of the checklist, regarding clarity, usefulness, completeness, and practicality.

**Table 1: Checklist dimensions focus areas**

Checklist Dimension	Focus Areas
AI Systems Characteristics	reliability, explainability, uncertainty communication, bias, hallucinations, need for human validation.
Human Factors	user expertise, familiarity with AI systems, confidence in identifying AI mistakes, personal comfort while interacting with AI systems
Task Characteristics	trust in high-risk, complex, and time-sensitive situations, behaviors related to verification and cross-checking of AI outputs
Organizational and Contextual Factors	workflow integration, institutional regulations, accountability, responsibility distribution in AI-assisted environments.

### 3.4 Planned Checklist Refinement

The first version of the checklist developed from the scoping review was then revised by paying attention to stakeholder feedback that was collected during the interviews. The refinement process was used as an opportunity to improve clarity, practicality, and applicability of the checklist in AI-assisted healthcare contexts and in broader ones as well.

Feedback that emerged from participants was analyzed thoroughly to identify recurring observations, unclear items from the checklist, missing considerations, and usability problems. Particular attention was given to suggestions related to ease of understanding of the checklist, with the intention of making the questions as generalizable as possible.

Another version of the checklist was produced, making sure that all the issues brought up by the stakeholders were fixed.

## 4 Stakeholders' Feedback and Iteration

### 4.1 Stakeholders' Perspectives Trust in Human-AI Collaboration

After analyzing the feedback received on the checklist, it was noted that some recurring themes regarding the topic of our research emerged.

Stakeholders consistently emphasized the importance of human oversight, particularly in high-risk healthcare contexts. Many participants stressed that AI outputs should not be accepted passively, suggesting that trust calibration depends heavily on the possibility of independently verifying AI recommendations.

Another recurring theme revolved around explainability and uncertain communication. Stakeholders reported that trust in AI systems increases when users are provided with explanations, evidence, confidence scores, warnings, or uncertainty-related information: this suggests that transparency alone may not be sufficient, unless users are also able to critically interpret the information provided. This suggests that trust in AI is highly context-dependent rather than determined solely by technical performance.

Organizational and contextual characteristics were also considered important drivers of trust calibration. Stakeholders involved frequently emphasized the importance of regulations, protocols, workflow integration, and safeguards capable of assuring safer and transparent AI use; some participants associated increased trust with the presence of clear regulations and policies governing AI systems.

### 4.2 Stakeholders' Initial Feedback

Several stakeholders that participated in the co-design checklist process gave some valuable feedback that helped us in the refinement process.

The checklist was generally considered understandable, relevant, and coherent with concerns surrounding AI-assisted decision-making. Several stakeholders also suggested that the checklist could potentially be generalized beyond healthcare-related contexts and adapted to other professional fields.

The most significant observations that emerged concerned the overall framing of the preliminary checklist. The first version of the checklist was perceived as strongly risk-centered. This feedback highlighted the importance of balancing risk-awareness with recognition of the practical value AI systems may provide. Stakeholders noted that the checklist focused heavily on risks such as loss of control, hallucinations, overreliance, and incorrect outputs, approaching the topic from a “fear perspective” rather than considering the practical benefits AI could bring to professional workflows.

Stakeholders also requested stronger attention to institutional safeguards and existing regulations such as the European AI Act (The European Union AI Act is the first comprehensive legal framework created by the European Union to regulate artificial intelligence systems based on their level of risk to people’s safety, rights, and society).

In addition, some other responses emphasized the importance of including considerations regarding healthcare data and privacy protection, which were perceived as insufficiently addressed in the first version.

A thematic analysis was conducted to identify recurring themes emerging from stakeholder feedback.

**Table 2: Count of times Stakeholders mentioned a certain theme in their interviews**

Recurring Theme	Stakeholders Mentioning It
Human oversight and verification	9/11
Explainability and uncertainty communication	8/11
Importance of regulation and accountability	7/11
Privacy and data concerns	5/11
Concern about overreliance	8/11

### 4.3 Checklist Refinement Plan

The stakeholders’ feedback was taken into serious consideration during the refinement process to try and satisfy every single concern that was brought to light.

Several recurring observations emerged regarding clarity, balance, accessibility, and contextual relevance of the interviews.

Based on the observations, multiple modifications were introduced into the revised version of the checklist. The main improvements are shown in Table 3.

**Table 3: Planned adjustment based on Stakeholders’ Feedback**

Stakeholder’s Feedback	Changes
The preliminary checklist focused on risks and loss of control associated with AI systems	Some questions regarding the practical benefits and appropriate use of AI systems were introduced
Existing regulations and government systems were not addressed enough	Some considerations regarding institutional safeguards and AI regulations were added
Certain AI-related terms were not explained for everybody to understand	Additional explanations, definitions, and simplified wording were modified

Certain checklist items were considered redundant or too similar

Redundant questions were merged and questions that had already been answered were deleted

## 5 Final Checklist

The final version of the checklist was developed through the refinement process described in the previous sections.

The revised checklist was designed not only as a reflective questionnaire, but also as a practical decision-support tool intended to support well-calibrated trust in AI systems. It encourages users to:

- evaluate whether AI outputs are verifiable,
- assess the level of risk associated with the task,
- identify the presence of institutional safeguards,
- determine whether human expertise is sufficient to critically review AI outputs.

When multiple high-risk conditions are identified simultaneously, such as low explainability or the absence of human validation, the checklist suggests reducing reliance on the AI system and increasing human oversight.

**Table 4: Decision Conditions Checklist and suggested Action**

Condition Identified	Suggested Action
AI output cannot be independently verified	Increase human oversight
High-risk task with low explainability	Reduce reliance on AI
Presence of uncertainty/confidence indicators	Trust may be increased cautiously
User lacks sufficient expertise	Require expert validation
Lack of institutional safeguards	Apply additional verification

The checklist is intended not only as a reflective tool, but also as a practical support mechanism for AI-assisted decision-making. Multiple high-risk conditions may indicate the need for stronger human oversight or reduced reliance on AI outputs. High-risk conditions may include:

- unverifiable AI outputs,
- low explainability,
- high-risk tasks,
- low user expertise,
- unclear accountability,
- absence of safeguards or uncertainty indicators.

**Table 5: Suggested Trust Calibration Actions**

High-Risk Conditions	Suggested Action
0–1	Standard human supervision
2–3	Increased verification
4+	Strong oversight, reduced AI reliance
High-risk + unverifiable	Mandatory human validation

Compared to the first version, the revised checklist adopts a more balanced approach by considering both AI risks and practical benefits. The refinement process also improved accessibility through simplified terminology and reduced redundancy. The main differences between the first and the second versions of the checklist are summarized in Table 6.

**Table 6: Comparison between 1<sup>st</sup> and 2<sup>nd</sup> version of the checklist**

1 <sup>st</sup> Version	2 <sup>nd</sup> Version
Focus on risks, loss of control and over trust issues	More balanced questions, able to explore both risks and benefits
Limited attention on already existing regulations	Inserted questions that added context to already existing governance
Technical terminology not explained enough	Added explanations to any term that may not be known to any kind of user
Redundant questions	Merged questions that seemed repetitive to give space to some new ones
Too theoretical-oriented	Added some practical questions that were still able to be as generalizable as possible
Not exploring privacy and data concerns enough	Included questions that focused on users' thoughts about privacy and data sharing

## 6 Worked Use Case

### 6.1 Scenario Description

Building on the findings of our scoping review, several application scenarios regarding AI-assisted healthcare decision-making were considered. Within these, AI-assisted diagnostic in the dentistry field was selected as a representative worked use case due to its strong dependence on human-AI collaboration, interpretation of complex information, and need for well-calibrated trust.

The healthcare professional may use AI output as support during the diagnostic process and treatment planning.

This scenario was selected because it reflects key trust-calibration challenges identified in the scoping review, including explainability, uncertainty communication, and human oversight.

To give an example, although AI systems may give faster support in image interpretation and access to additional diagnostic insights, healthcare professionals need to continuously verify if the AI output is reliable, if additional verification is needed, and to what extent AI's output should influence the final decision.

Both overreliance and underreliance may negatively affect AI-assisted decision-making.

### 6.2 Preliminary Checklist Considerations

Before collecting stakeholder feedback, the preliminary version of the checklist was primarily shaped by the theoretical findings of the scoping review. In the selected dentistry-related scenario, the initial checklist focused on factors that emerged consistently throughout the literature and were summarized in Table 1.

Particular attention was given to the possibility of independently verifying AI outputs before integrating them into the final decision-making process. Based on the literature, explainability and the communication of uncertainty were identified as essential elements for supporting informed and well-calibrated interactions with AI systems.

Concerns regarding overreliance, automation bias, and inappropriate trust also influenced the inclusion of checklist items designed to encourage critical evaluation rather than the passive acceptance of AI-generated recommendations.

However, at this stage of development, the checklist still relied mainly on theoretical findings and assumptions derived from the literature. One of the central objectives of the co-design process was therefore to evaluate whether these considerations would remain equally relevant when compared with the perspectives of dentistry professionals working in real-world contexts.

### 6.3 Applying the Final Checklist to the Dentistry Scenario

A dentist uses an AI-assisted tool to analyze a dental X-ray during a routine patient evaluation. The AI system identifies a possible early-stage lesion and recommends further investigation with a confidence score of 78%, but it does not clearly explain how the result was reached. Since the patient also presents a complex medical history, the dentist must decide whether the AI recommendation should be trusted directly or verified further before influencing the final clinical decision.

To demonstrate the practical applicability of the checklist, the scenario was evaluated using the revised Version 2 checklist developed through the co-design process. The worked use case shows how the checklist can help professionals identify risk conditions, critically evaluate AI outputs, and support well-calibrated trust during AI-assisted decision-making.

In this scenario, factors such as explainability, independent verification, contextual risk, organizational safeguards, and human oversight become particularly important when evaluating whether the AI recommendation is trustworthy.

The questions referring to the final checklist can be found in the Appendix D.

**Table 7: Example Application of the Final Checklist in the Dentistry Scenario**

Checklist Item	Example Response
Q7	Very important
Q8	Yes, previous incorrect detections reduce automatic trust
Q9	Very important
Q10	Explanations, evidence, and uncertainty information
Q11	Yes, patient medical data increases caution
Q12	Yes, diagnosis is considered high-risk
Q13	Yes
Q16	No
Q17	Partially, confidence score is present but explanations are limited

Q18	Yes, the dentist can manually review the X-ray
Q19	Yes
Q20	Partially
Q22	The human user
Q23	Very concerning

The application of the checklist highlights several conditions requiring caution before relying on the AI recommendation. The scenario involves a high-risk healthcare context, limited explainability, sensitive patient data, and the possibility of significant harm if the output is incorrect. Although the AI system provides a confidence score, the lack of detailed explanations limits full verification of the recommendation.

Based on the trust-calibration framework presented in Table 5, these conditions suggest the need for increased verification and stronger human oversight before integrating the AI output into the final clinical decision. In this case, the dentist should therefore use the AI recommendation as supportive information rather than as an autonomous decision-making tool.

This worked example shows how the checklist can support critical reflection, reduce overreliance, and encourage safer human-AI collaboration in healthcare workflows.

#### 6.4 Stakeholder Feedback Considerations

The stakeholder feedback confirmed several of the main considerations identified in the scoping review and incorporated into the first version of the checklist.

Participants emphasized the importance of explainability, human oversight, accountability, and the verification of AI outputs depending on the level of contextual risk.

These observations supported the relevance of the four dimensions identified in the scoping review and reinforced their applicability within real-world healthcare-related contexts.

Within the dentistry scenario, participants highlighted that trust in AI systems should not rely exclusively on technical performance, but also on the possibility of independently verifying AI outputs before integrating them into the decision-making process. They also noted that trust may vary depending on contextual factors such as task complexity, time pressure, and the potential consequences of incorrect outputs.

The stakeholder feedback further supported the contextual and human-centered perspective on trust calibration proposed in the scoping review.

The interview completed by the dentist who participated in the co-design process is included in Appendix C.

#### 6.5 Reflection on Checklist Application

The worked use case demonstrates how the final checklist may support professionals during AI-assisted tasks by structuring trust-related considerations into a practical reflective process. Rather than replacing professional expertise, the checklist is intended to encourage users to critically evaluate AI outputs, contextual risks, explainability, and verification possibilities before acting on AI recommendations.

The example also illustrates that well-calibrated trust depends not only on technical performance, but also on contextual and organizational factors such as human oversight, institutional safeguards, accountability, and data sensitivity. When multiple risk conditions emerge simultaneously, the checklist encourages reduced reliance on AI outputs and increased human validation.

Overall, the worked use case demonstrates how the co-designed checklist may support safer, more transparent, and context-aware AI-assisted decision-making in real-world healthcare environments.

## 7 Conclusion

Through the co-design process, stakeholder feedback contributed to the validation of some of the theoretical findings identified in the scoping review previously produced, and to the refinement of the checklist structure, wording, and framing.

Stakeholder involvement revealed practical concerns and contextual factors that may not emerge through literature analysis alone, demonstrating that the involvement of actual users is always important in these types of work.

Although the checklist was developed within a limited number of stakeholders, the project demonstrates how a co-design approach may help use theoretical findings to create an accessible and generalizable tool for any type of user.

Overall, the findings suggest that well-calibrated trust in AI depends not only on technical reliability, but also on human expertise, contextual risk, and organizational safeguards.

## Appendix

### Appendix A: Trust Calibration Decision Checklist – Version 1

#### Co-designed Checklist for Well-Calibrated Trust in Human-AI Collaboration

Thank you for participating in this interview. We are students of Politecnico di Torino working on a project about well-calibrated trust in AI systems, especially in healthcare-related contexts.

The goal is to collect feedback that will help us improve a practical checklist to support appropriate and well-calibrated trust in AI-assisted tasks.

Your answers will remain anonymous and will be used for academic purposes only.

#### 1 — Background and Experience with AI

1. What is your field of study/work?
2. Have you used AI systems or AI-assisted tools before? If yes, in what context?
3. How familiar would you say you are with AI systems? (Check what represents best)
  - Very familiar
  - Somewhat familiar
  - Slightly familiar
  - Not familiar
4. Have you ever experienced a situation where you trusted an AI system? If yes, what made you trust it?
5. Have you ever experienced a situation where you distrusted or questioned an AI output? If yes, what brought you to that conclusion?

#### 2 – Trust in AI Systems

##### 2.1 - AI System Characteristics

6. How important is AI accuracy when deciding whether to trust an AI system? (Check what represents best)
  - Very important
  - Somewhat important
  - Slightly important
  - Not important
7. Do you think AI systems remain reliable over time, or should their performance be continuously monitored?
8. Have you ever encountered situations where an AI system produced incorrect, misleading, or “hallucinated” outputs? If yes, how did this affect your trust?
9. How important is it for users to understand how the AI reached its result? (Check what represents best)
  - Very important
  - Somewhat important
  - Slightly important
  - Not important
10. What kind of information should an AI system provide to help users evaluate and verify its output?

(Check what you think is more important, even more than one)

- explanations
  - confidence score
  - evidence
  - warnings
  - uncertainty information
  - Other (please explain):
11. How concerned are you about possible bias in AI systems? (Check what represents best)
    - Very concerned
    - Somewhat concerned
    - Slightly concerned
    - Not concerned
  12. In your opinion, when is human validation of AI outputs most necessary?
  13. Does the type or sensitivity of the information given to the AI affect how much you trust it?
  14. Do you think trust in AI should change depending on whether the situation is high-risk or low-risk?

##### 2.2 - Human Factors

15. How comfortable do you personally feel when using AI systems? (Check what represents best)
  - Very comfortable
  - Somewhat comfortable
  - Slightly comfortable
  - Not comfortable
16. Do you think users need specific expertise or training to appropriately trust AI systems?
17. How confident do you feel in your ability to identify mistakes made by AI systems? (Check what represents best)
  - Very confident
  - Somewhat confident
  - Slightly confident
  - Not confident
18. Do you think familiarity with AI systems increases trust over time?
19. Have you ever trusted an AI output simply because it sounded confident or convincing?

##### 2.3 - Task Characteristics

20. Do you trust AI differently in routine situations compared to highly sensitive or critical situations?
21. How does task complexity influence your willingness to trust AI?
22. When using AI, do you usually cross-check the results? Why or why not?
23. Have there been situations where you trusted AI enough not to verify the output?
24. What do you think are the risks of relying too much on AI systems?
25. What do you think could happen if an AI output is wrong in a high-risk situation?
26. Do you think AI systems are more useful or riskier in time-sensitive situations?

### 2.4 - Organizational and Contextual Factors

27. Do you think organizations should establish clear guidelines for how AI systems should be used?
28. How important is it for AI systems to integrate smoothly into existing workflows? (Check what represents best)
  - Very important
  - Somewhat important
  - Slightly important
  - Not important
29. Who do you think should remain responsible for the final decision when AI is involved? (Check what represents best)
  - The human user
  - The organization
  - The AI developer
  - Shared responsibility
  - Other (please explain):
30. Do you think there should be clear accountability protocols in case an AI-assisted decision causes harm or error?
31. How important are regulations, standards, or institutional policies for building trust in AI systems? (Check what represents best)
  - Very important
  - Somewhat important
  - Slightly important
  - Not important
32. Would you trust AI systems more if organizations clearly communicated how they are monitored and regulated? Why or why not?

### 3 — Checklist Feedback

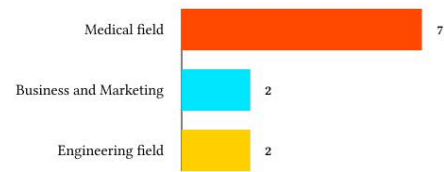
33. Do you think the checklist addresses the most important aspects of trusting AI systems? Why or why not?
34. Which checklist items do you find the most useful?
35. Are there any checklist items that seem unclear or vague?
36. Are there any important aspects missing from the checklist that you feel are important to this field?
37. Do you think the checklist is practical and understandable for all types of users?
38. Would this checklist help someone think more critically about whether to trust an AI system? Why or why not?
39. Do you think this checklist could be expanded to any field of work?
40. Do you have any suggestions for improving the checklist?

## Appendix B: Version 1 Checklist Stakeholders' Answers

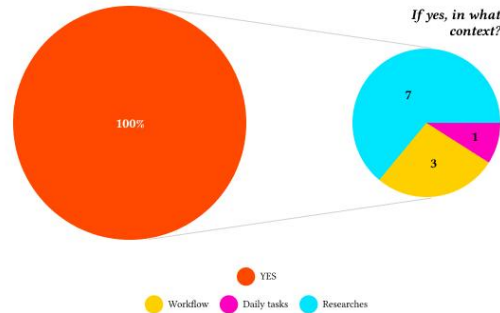
Graphics and Tables of the answers were made manually by group members using Figma program.

### 1 — Background and Experience with AI

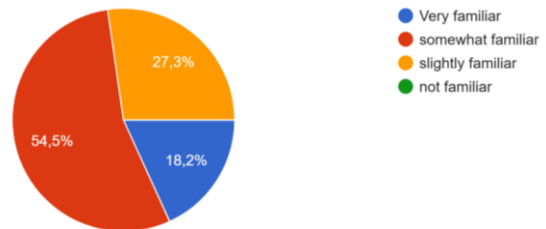
1. What is your field of study/work?



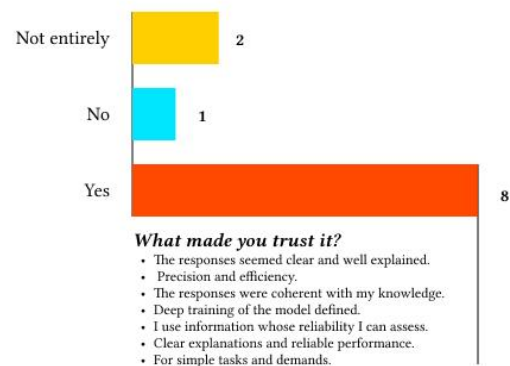
2. Have you used AI systems or AI-assisted tools before? If yes, in what context?



3. How familiar would you say you are with AI systems?

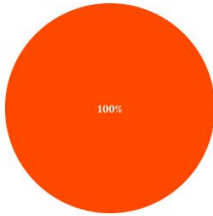


4. Have you ever experienced a situation where you trusted an AI system? If yes, what made you trust it?



5. Have you ever experienced a situation where you distrusted or questioned an AI output? If yes, what brought you to that conclusion?

*Have you ever experienced a situation where you distrusted or questioned an AI output?*

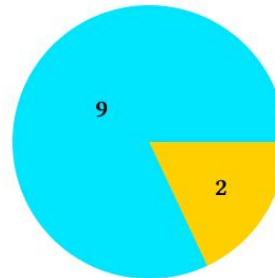


● YES

*If yes, what brought you to that conclusion?*

- AI often tends to agree with you.
- The answer AI gave me did not go along with other informations I already had.
- Sometimes if the prompt is not specifically detailed the conclusion often is wrong.
- The answers didn't add up and it seemed like it didn't understand my words.
  - Common sense.
- Even if it was a simple task the results weren't properly calculated and did not match my manual computations.

8. Have you ever encountered situations where an AI system produced incorrect, misleading, or "hallucinated" outputs? If yes, how did this affect your trust?



● Yes ● No

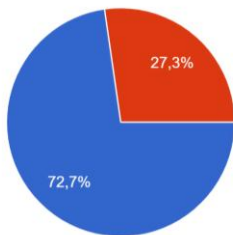
*If yes, how did this affect your trust?*

- It didn't.
- My confidence faltered and I had to pay more attention to my commands.
  - I had to double check everything.
- I had to filter every information according to my knowledge.

## 2 - Trust in AI Systems

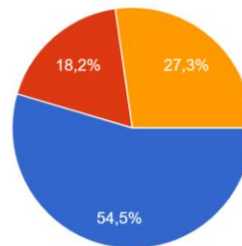
### 2.1 - AI System Characteristics

6. How important is AI accuracy when deciding whether to trust an AI system? (Check what represents best)



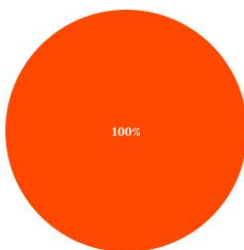
- very important
- somewhat important
- slightly important
- not important

9. How important is it for users to understand how the AI reached its result? (Check what represents best)



- very important
- somewhat important
- slightly important
- not important

7. Do you think AI systems remain reliable over time, or should their performance be continuously monitored?



● YES

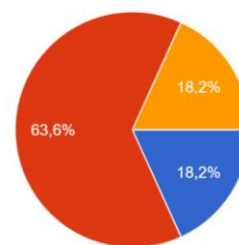
*If yes, what brought you to that conclusion?*

- Their reliability can vary depending on the context, the updates they receive, the quality of the information they provide, and technological innovation.
- Otherwise they could encounter some hallucinations.
- Their performance can change over time because data, users, and real-world situations change, so regular checks help keep them accurate and reliable.

10. What kind of information should an AI system provide to help users evaluate and verify its output? (Check what you think is more important, even more than one)

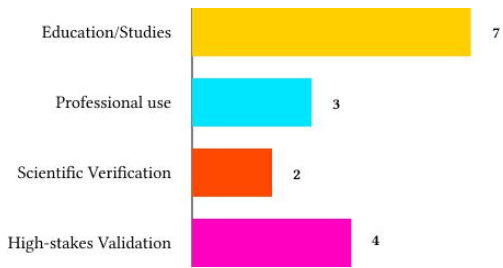


11. How concerned are you about possible bias in AI systems? (Check what represents best)

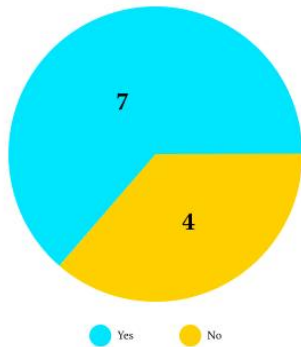


- very concerned
- somewhat concerned
- slightly concerned
- not concerned

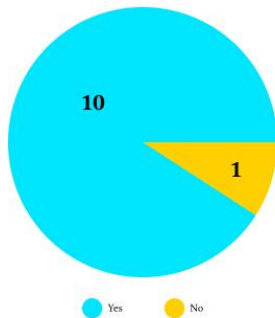
12. In your opinion, when is human validation of AI outputs most necessary?



13. Does the type or sensitivity of the information given to the AI affect how much you trust it?

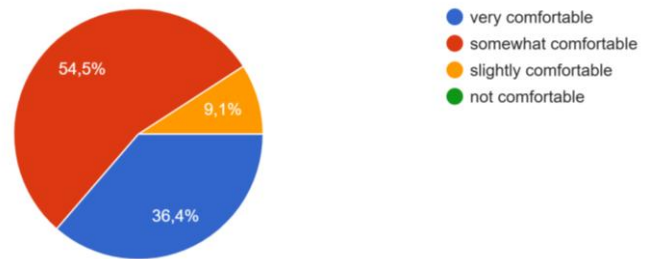


14. Do you think trust in AI should change depending on whether the situation is high-risk or low-risk?

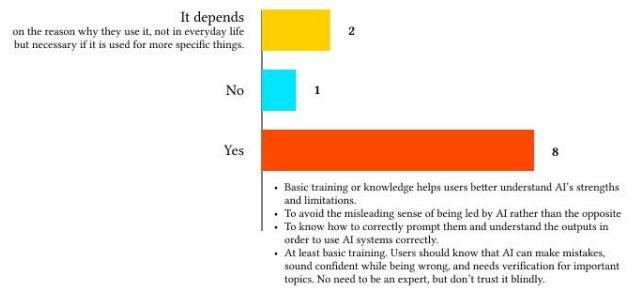


## 2.2 - Human Factors

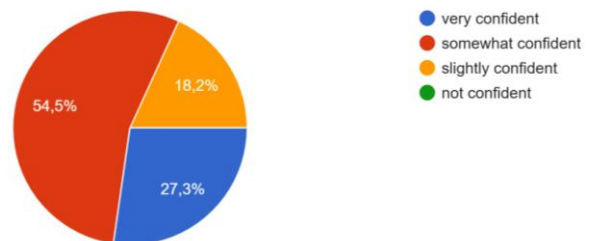
15. How comfortable do you personally feel when using AI systems? (Check what represents best)



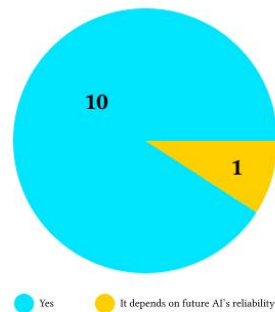
16. Do you think users need specific expertise or training to appropriately trust AI systems?



17. How confident do you feel in your ability to identify mistakes made by AI systems? (Check what represents best)



18. Do you think familiarity with AI systems increases trust over time?

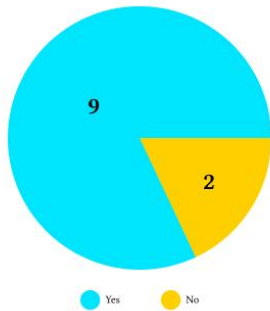


19. Have you ever trusted an AI output simply because it sounded confident or convincing?



### 2.3 - Task Characteristics

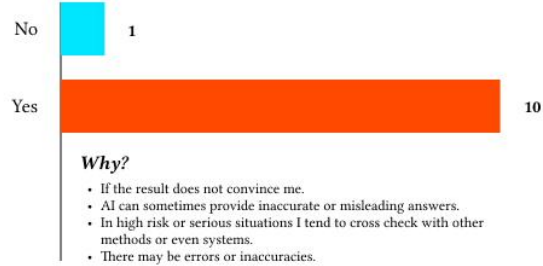
20. Do you trust AI differently in routine situations compared to highly sensitive or critical situations?



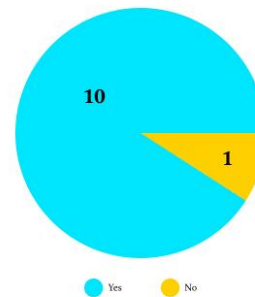
21. How does task complexity influence your willingness to trust AI?

- The more complex the task is, the more careful I become when using AI.
- I think a lot
- The more complex the task is, the more careful I am about trusting AI without verification.
- I would double check the answer with my knowledge in complex tasks
- Not so much. It's not a matter of complexity but a matter of impact
- I tend to only use AI for complex tasks as if it's easy, I can do it by myself. However I am always checking the outputs
- The more complex the task, the more prudent it is to evaluate the reliability of the answers.
- The more complex the task is, the more careful I will be in asking information and asking the AI itself to double check or test if its answer is rooted in reasonably.
- The more complex the task is, the less I would trust AI blindly. For simple tasks, I'd rely on it more, but for complex tasks I'd want explanations, sources, and human review.
- If the request is very difficult to verify the AI information by consulting other sources
- I trust AI used step by step

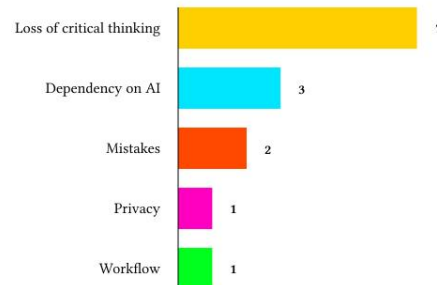
22. When using AI, do you usually cross-check the results? Why or why not?



23. Have there been situations where you trusted AI enough not to verify the output?



24. What do you think are the risks of relying too much on AI systems?

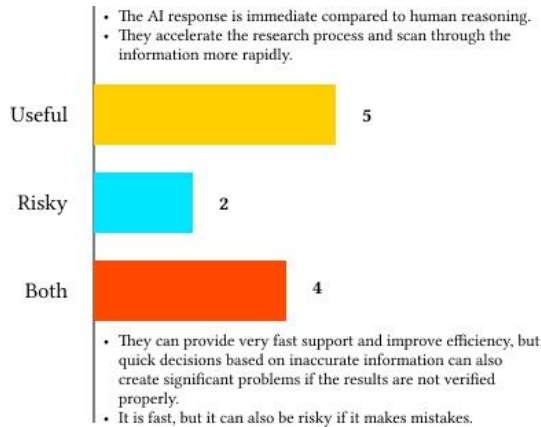


25. What do you think could happen if an AI output is wrong in a high-risk situation?

- If an AI is wrong in a high-risk situation, it can cause harm, financial loss, unfair decisions, or even put lives at risk.
- If an AI output is wrong in a high-risk situation, it could lead to serious consequences.
- It would end badly.
- Depends on the situation.
- It depends on the user.
- Putting someone or something in danger.
- It could have serious consequences.
- An incorrect response in a high-risk situation can cause catastrophic damage if not checked in time.
- I honestly don't know.

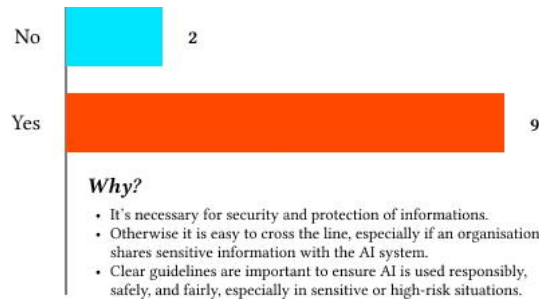
- Bad decisions that could affect a lot of people could be taken.
- It can lead to complex situations where people's lives are in danger.

26. Do you think AI systems are more useful or riskier in time-sensitive situations?

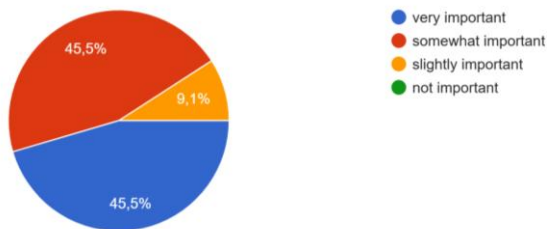


### 2.4 - Organizational and Contextual Factors

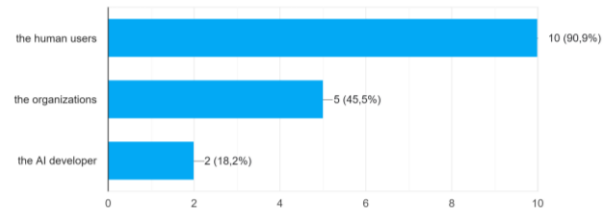
27. Do you think organizations should establish clear guidelines for how AI systems should be used?



28. How important is it for AI systems to integrate smoothly into existing workflows? (Check what represents best)



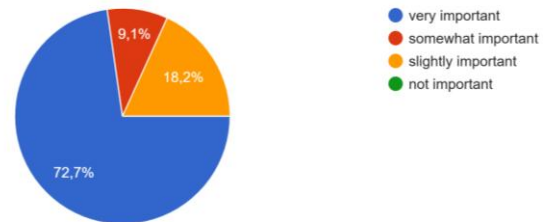
29. Who do you think should remain responsible for the final decision when AI is involved? (Check what represents best)



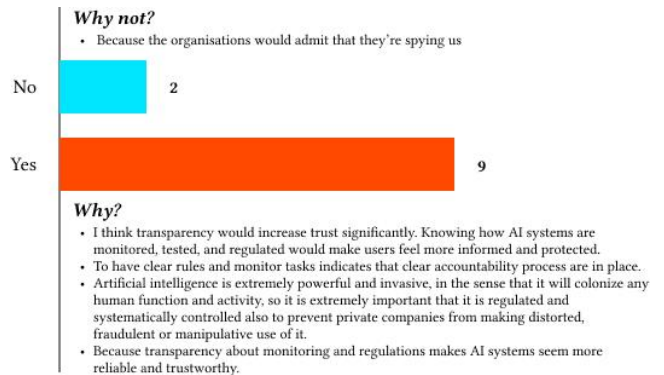
30. Do you think there should be clear accountability protocols in case an AI-assisted decision causes harm or error?



31. How important are regulations, standards, or institutional policies for building trust in AI systems? (Check what represents best)



32. Would you trust AI systems more if organizations clearly communicated how they are monitored and regulated? Why or why not?



### 3 - Checklist Feedback (anonymous notes written by the stakeholders in a questionnaire)

33. Do you think the checklist addresses the most important aspects of trusting AI systems? Why or why not?

- Yes, because it covers reliability, risks, human validation, and transparency.

- Yes, because it encourages users to think critically instead of trusting AI automatically.
  - Yes, it covers everything
  - Yes, but not the most important
  - Maybe not all, but mostly yes
  - Yes, it was very effective
  - Yes
  - It may be helpful
  - Yes
  - I think it's pretty complete
  - Yes
34. Which checklist items do you find the most useful?
- The section Human Factors
  - The rusting theme
  - Database security
  - The regulations and accountability since mistakes cannot be eliminated completely
  - The aspect about accountability
  - Whether we are aware that AI does hallucinate and make mistakes
  - I find them redundant.
  - Regulation and critical usage of AI
  - The items about accuracy, verification, and risk level.
  - I don't know
  - 2- trust in AI, 2.1 more precisely
35. Are there any checklist items that seem unclear or vague?
- No
  - No
  - I don't know
  - None
  - None
  - Some
  - Some terms like "reliability" could be explained more clearly.
  - Some
  - No
  - No
  - No
36. Are there any important aspects missing from the checklist that you feel are important to this field?
- No
  - No
  - I don't know
  - None
  - Maybe asking if people would be interested in learning how to handle an AI system. And learn about transparency and explainability
  - No
  - It's not my problem
  - Privacy and data protection could be included more strongly.
  - Not everyone knows how much AI pollutes
  - No
- No
37. Do you think the checklist is practical and understandable for all types of users?
- Yes
  - Yes
  - Quite clear
  - Some things could be found a bit confusing if they are read by someone who doesn't study or work with AI
  - Yes
  - It depends on the level of the people who have to answer these questions. For some, it might be difficult.
  - Yes
  - Yes, it seems simple and practical for most users
  - Yes
  - Yes
  - Yes
38. Would this checklist help someone think more critically about whether to trust an AI system? Why or why not?
- Yes, I believe it would. The checklist encourages users to reflect on reliability, risks, and verification instead of accepting AI outputs automatically. This could promote a more balanced and critical use of AI systems.
  - Yes, it made you reflect about important questions
  - I don't know
  - It can help, but if someone is already familiar with AI systems and workflows, they probably have concluded these themselves
  - I think so
  - Yes it would
  - This question about whether we should trust an AI system has been asked many times with different nuances, it seems excessive to me.
  - Yes, it underlines important but not obvious aspects of using AI
  - Yes, because it encourages users to question and verify AI outputs.
  - Yes
  - Yes I hope so
39. Do you think this checklist could be expanded to any field of work?
- Yes
  - Yes
  - Not all but many fields
  - I think it's quite general, so... yes
  - This might be helpful.
  - Yes, especially scientific research and journalism
  - Yes
  - Yes, especially in healthcare, education, business, and technology.

- I don't know
  - Yes
  - Yes
40. Do you have any suggestions for improving the checklist?
- No
  - No
  - No
  - I think that this checklist is perhaps a little biased from a "fear perspective" of losing control, rather than to deep dive actual regulatory European rules that are already in place
  - None
  - Add more real-life examples and clearer explanations for technical terms.
  - No
  - No good work, maybe too much on trust AI, but it's good to raise awareness
  - No
  - No
  - No

## Appendix C: Dentist Answers to the Checklist version 1

### 1 - Background and Experience with AI

1. What is your field of study/work?  
Dentist.
2. Have you used AI systems or AI-assisted tools before? If yes, in what context?  
Yes, after testing it to evaluate its reliability and level of detail, I now use artificial intelligence in a wide variety of contexts: scientific insights, news, tourism, history—practically any topic requiring information or insights, including those related to my profession.
3. How familiar would you say you are with AI systems?
  - Somewhat familiar
4. Have you ever experienced a situation where you trusted an AI system? If yes, what made you trust it?  
When it comes to an important topic, I've never trusted it 100%. I use information whose reliability I can assess, because I've had the opportunity to verify that artificial intelligence sometimes makes mistakes and occasionally makes trivial errors. I've also had the impression that it often tries to please the user.
5. Have you ever experienced a situation where you distrusted or questioned an AI output? If yes, what brought you to that conclusion?  
Sometimes the AI gave completely wrong answers. I pointed this out, and it admitted its mistake, apologized, and corrected it.

## 2 - Trust in AI Systems

### 2.1 - AI System Characteristics

6. How important is AI accuracy when deciding whether to trust an AI system? (Check what represents best)
  - Very important
7. Do you think AI systems remain reliable over time, or should their performance be continuously monitored?  
I think that currently it is better to have continuous monitoring, at least for the levels of detail dedicated to ordinary users.
8. Have you ever encountered situations where an AI system produced incorrect, misleading, or "hallucinated" outputs? If yes, how did this affect your trust?  
Yes, as already reported, it has happened that it provided incorrect and/or more or less complacent results; therefore I always use the information provided, filtering it according to my knowledge.
9. How important is it for users to understand how the AI reached its result? (Check what represents best)
  - Somewhat important
10. What kind of information should an AI system provide to help users evaluate and verify its output? (Check what you think is more important, even more than one)
  - explanations
  - confidence score
  - evidence
  - warnings
  - uncertainty information
11. How concerned are you about possible bias in AI systems? (Check what represents best)
  - Slightly concerned
12. In your opinion, when is human validation of AI outputs most necessary?  
I think mostly in response to scientific arguments.
13. Does the type or sensitivity of the information given to the AI affect how much you trust it?  
Yes.
14. Do you think trust in AI should change depending on whether the situation is high-risk or low-risk?  
The riskier the situation, the more reliable the answers should be.

### 2.2 - Human Factors

15. How comfortable do you personally feel when using AI systems? (Check what represents best)
  - Very comfortable
16. Do you think users need specific expertise or training to appropriately trust AI systems?  
Perhaps training could be useful when working on highly difficult topics.

17. How confident do you feel in your ability to identify mistakes made by AI systems? (Check what represents best)
  - Somewhat confident
18. Do you think familiarity with AI systems increases trust over time?  
It depends on how reliable the AI proves to be during continuous use.
19. Have you ever trusted an AI output simply because it sounded confident or convincing?  
Especially if the answers are about topics that are not particularly important, we tend to accept the answers received in the hope that they are reliable.

### 2.3 - Task Characteristics

20. Do you trust AI differently in routine situations compared to highly sensitive or critical situations?  
Yes.
21. How does task complexity influence your willingness to trust AI?  
The more complex the task, the more prudent it is to evaluate the reliability of the answers.
22. When using AI, do you usually cross-check the results? Why or why not?  
If the answers concern important topics, it is advisable to check them because there may be errors or inaccuracies.
23. Have there been situations where you trusted AI enough not to verify the output?  
I never check the information I receive if it concerns trivial topics.
24. What do you think are the risks of relying too much on AI systems?  
The more important the topic in question, the more a wrong answer could cause serious damage.
25. What do you think could happen if an AI output is wrong in a high-risk situation?  
An incorrect response in a high-risk situation can cause catastrophic damage if not checked in time.
26. Do you think AI systems are more useful or riskier in time-sensitive situations?  
It can be very useful because the AI response is immediate compared to human reasoning.

### 2.4 - Organizational and Contextual Factors

27. Do you think organizations should establish clear guidelines for how AI systems should be used?  
It would be helpful.
28. How important is it for AI systems to integrate smoothly into existing workflows? (Check what represents best)
  - Very important
29. Who do you think should remain responsible for the final decision when AI is involved? (Check what represents best)
  - The human user

30. Do you think there should be clear accountability protocols in case an AI-assisted decision causes harm or error?  
Yes.
31. How important are regulations, standards, or institutional policies for building trust in AI systems? (Check what represents best)
  - Very important
32. Would you trust AI systems more if organizations clearly communicated how they are monitored and regulated? Why or why not?  
Artificial intelligence is extremely powerful and invasive, in the sense that it will colonize any human function and activity, so it is extremely important that it is regulated and systematically controlled also to prevent private companies from making distorted, fraudulent or manipulative use of it.

## Appendix D: Trust Calibration Decision Checklist - Version 2

### Co-designed Checklist for Well-Calibrated Trust in Human-AI Collaboration

Thank you for participating in this short interview. We are students of Politecnico di Torino working on a project about well-calibrated trust in AI systems, especially in healthcare-related contexts.

The goal is to collect feedback that will help us improve a practical checklist intended to support appropriate and well-calibrated trust in AI-assisted tasks.

Your responses will remain anonymous and will only be used for academic purposes only.

#### 1 — Background and Experience with AI

1. What is your field of study/work?
2. Have you used AI systems or AI-assisted tools before? If yes, in what context?
3. How familiar would you say you are with AI systems? (Check what represents best)
  - Very familiar
  - Somewhat familiar
  - Slightly familiar
  - Not familiar
4. Have you ever experienced a situation where you trusted an AI system? If yes, what made you trust it?
5. Have you ever experienced a situation where you distrusted or questioned an AI output? If yes, what brought you to that conclusion?
6. If you are not familiar with AI systems, would you be interested in learning more about how to use it? And if you are already familiar with AI, would you be interested in learning even more?

## 2 - Trust in AI Systems

### 2.1 - AI System Characteristics

7. How important is AI accuracy when deciding whether to trust an AI system?
  - Very important
  - Somewhat important
  - Slightly important
  - Not important
8. Were there any situations where an AI system produced incorrect, misleading, or "hallucinated" \* outputs?  
*\*When AI (like ChatGPT or Gemini) produces confident but inaccurate or nonsensical information that is not grounded in reality*
9. How important is it for users to understand how the AI reached its result?
  - Very important
  - Somewhat important
  - Slightly important
  - Not important
10. What kind of information should an AI system provide to help users evaluate and verify its output?
  - explanations
  - confidence score
  - evidence
  - warnings
  - uncertainty information
  - Other (please explain):
11. Does the type or sensitivity of the information given to the AI affect how much trust it should be given?
12. Should trust calibration change depending on whether the situation is high-risk or low-risk?  
*An example of a high-risk situation could be handling finances, whilst an example of a low-risk situation could be scheduling appointments.*

### 2.2 - Human Factors

13. Should users have specific expertise or get training to appropriately trust AI systems?
16. Would you rely on this AI system without supervision after repeated successful use?
17. Does the AI provide enough evidence or explanation to justify trusting its output?

### 2.3 - Task Characteristics

18. Can the AI output be independently verified before action is taken?
19. Would incorrect AI output create significant harm in this situation?

### 2.4 - Organizational and Contextual Factors

20. Are clear organizational rules or accountability procedures available for this AI use? (for example: the European AI Act\*)  
*\*The European Union AI Act is the first comprehensive legal framework created by the European Union to regulate artificial intelligence systems based on their level of risk to people's safety, rights, and society.*
  - Very familiar
  - Somewhat familiar
  - Slightly familiar
  - Not familiar
21. How important is it for AI systems to integrate smoothly into existing workflows?
  - Very important
  - Somewhat important
  - Slightly important
  - Not important
22. Who should remain responsible for the final decision when AI is involved?
  - The human user
  - The organization
  - The AI developer
  - Shared responsibility
  - Other (please explain):
23. How concerned should organizations and society be about the use, sharing, and privacy of sensitive personal data in AI systems?
  - Very concerned
  - Somewhat concerned
  - Slightly concerned
  - Not concerned

### 3 - Checklist Feedback

24. Does the checklist address the most important aspects of trusting AI systems? Why or why not?
25. Which checklist items do you find the most useful?
26. Are there any checklist items that seem unclear or vague?
27. Are there any important aspects missing from the checklist that you feel are important to this field?
28. Is the checklist practical and understandable for all types of users?
29. Would this checklist help someone think more critically about whether to trust an AI system? Why or why not?
30. Could this checklist be expanded to any field of work?
31. Do you have any suggestions for improving the checklist?

## Appendix E: Operational Trust Calibration Checklist

- AI output can be independently verified
- AI provides evidence/explanations
- Human expert reviewed the output
- Task is high-risk
- Accountability protocol exists
- AI uncertainty/confidence score is visible
- User has sufficient expertise
- Sensitive/private data is involved
- Organizational safeguards are present
- AI recommendation conflicts with human judgment

High-Risk Conditions	Suggested Action
0-1	Standard human supervision
2-3	Increased verification
4+	Strong oversight, reduced AI reliance
High-risk + unverifiable	Mandatory human validation

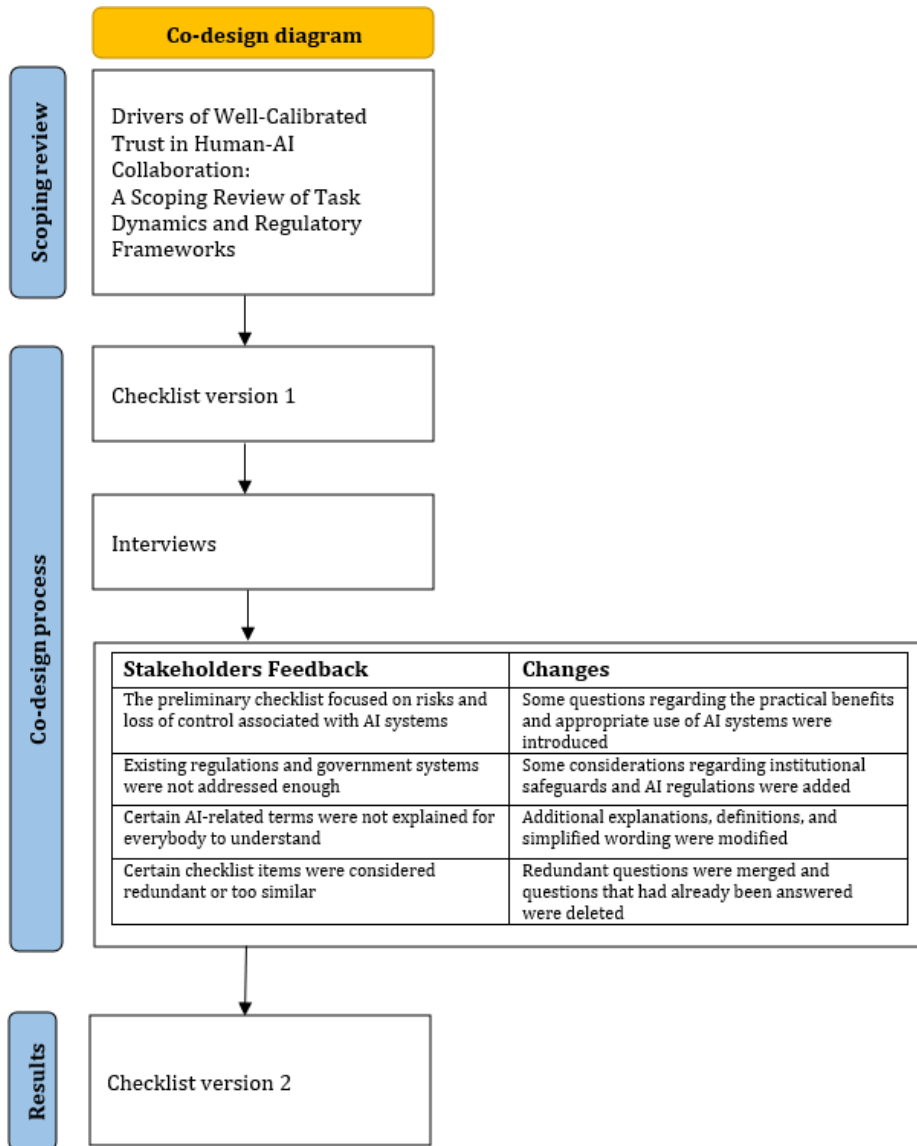
## Appendix F: Disclosure of AI Tools

AI tools such as ChatGPT, Gemini, and Claude were utilized during the project to better comprehend the practical aspects of creating a co-designed checklist and to gain insights on how to systematically organize the qualitative feedback collected from the interviews.

However, it is important to clarify that all the charts, diagrams, and visual representations visible in Appendix B were made manually by the team members on Figma. AI was strictly used as a conceptual and linguistic support to structure the data aggregation process, ensuring complete human oversight, transparency, and accuracy in the visual translation of our research findings.

## Appendix G: Prisma SCR-Diagram of the Co-design Process

Diagram 1: Workflow diagram of the development of the Co-design Checklist



## Appendix H: Dogtooth Team Members

Samin Nour Azaran; s324057@studenti.polito.it  
 Francesco Sergio Blandino; s320524@studenti.polito.it  
 Giorgia Brusa; s341200@studenti.polito.it  
 Elisa Ciani; s340642@studenti.polito.it  
 Elena Corio; s342703@studenti.polito.it  
 Bruno Yael Zambrano Rios; s337077@studenti.polito.it  
 Alessandro Russo; s328917@studenti.polito.it  
 Sarvar Toshmurodov; s334729@studenti.polito.it  
 Sonia Visco; s341763@studenti.polito.it  
 Erika Zhou; s335819@studenti.polito.it