

PRIVY: Operationalizing Privacy Policy with AI-Assisted Privacy Impact Assessment Workflow

Hank Lee, Yu-Ju Yang, Matthew Bilik, Isadora Krsek, Thomas Serban von Davier, Kyzyl Monteiro, Jason Lin, Shivani Agarwal
Jodi Forlizzi, Sauvik Das



PhD student
Carnegie Mellon University
hankplee.com

Privacy and Public Policy Conference
Feb. 10, 2026



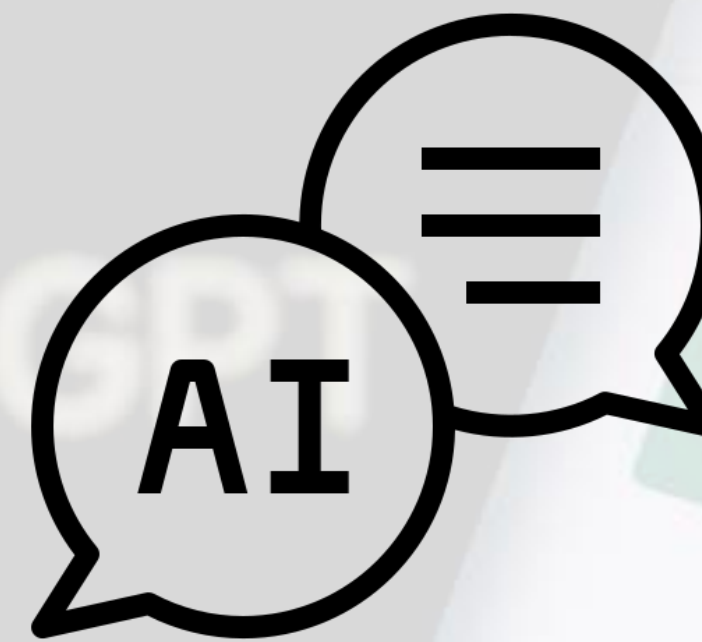
**Human-
Computer
Interaction
Institute**

Carnegie Mellon University
Security and Privacy Institute

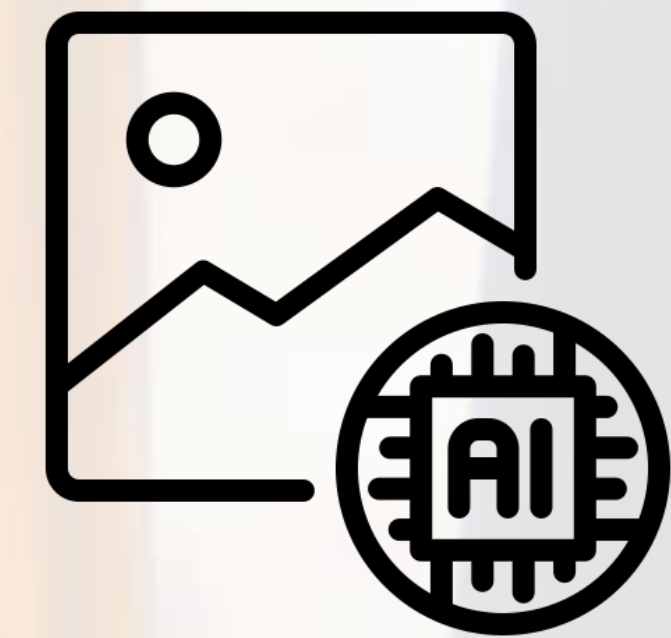




Facial recognition



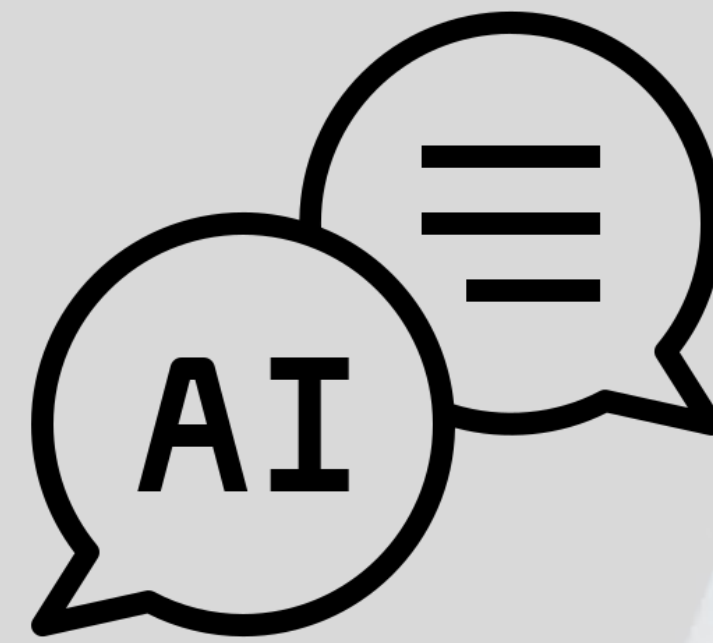
Large language model



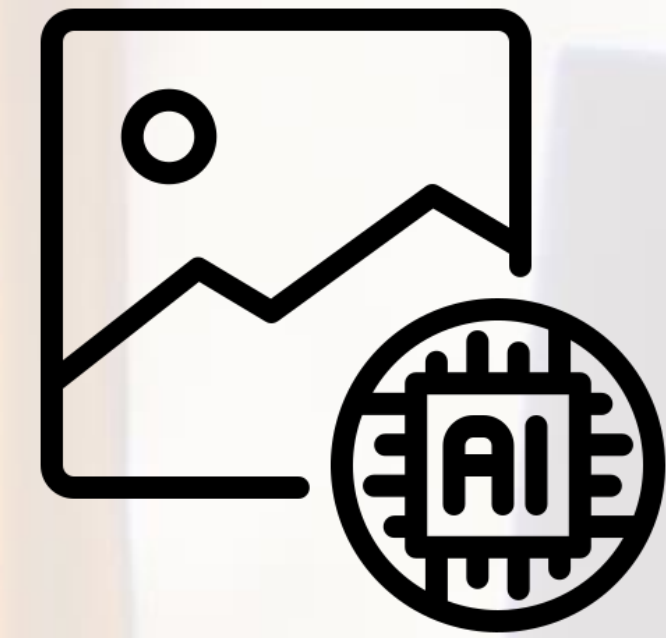
Diffusion-based algorithm



Facial recognition



Large language model



Diffusion-based algorithm

The New York Times



The Secretive Company That Might End Privacy as We Know It

A little-known start-up helps law enforcement match photos of unknown people to their online images — and “might lead to a dystopian future or something,” a backer says.

ars TECHNICA


BIZ & IT TECH SCIENCE POLICY CARS GAMING & CULTURE STO

ADVENTURES IN 21ST-CENTURY PRIVACY —

Artist finds private medical record photos in popular AI training data set

LAION scraped medical photos for AI research use. Who's responsible for taking them down?

BENJ EDWARDS - 9/21/2022, 11:43 AM




Pokimane, QTCinderella, & Sweet Anita slam deepfakes

One of the biggest and most influential streamers on Twitch, [QTCinderella](#), expressed her outrage at those who were sharing the explicit images and the website that hosted them.

“Everybody f*cking stop. Stop spreading it. Stop advertising it.”

She also gave her perspective on how this violation feels, saying “[b]eing seen ‘naked’ against your will should NOT BE A PART OF THIS JOB.”

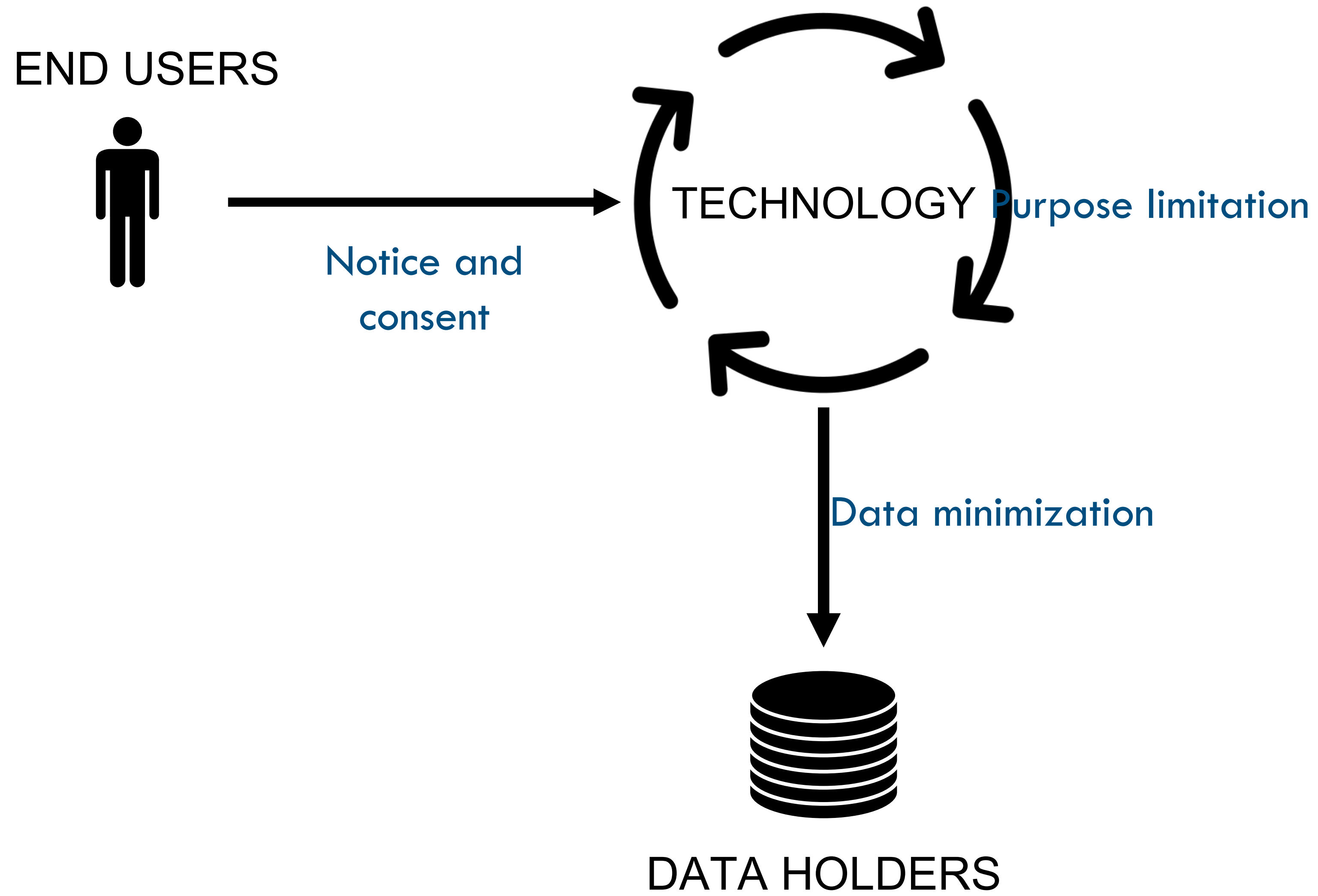


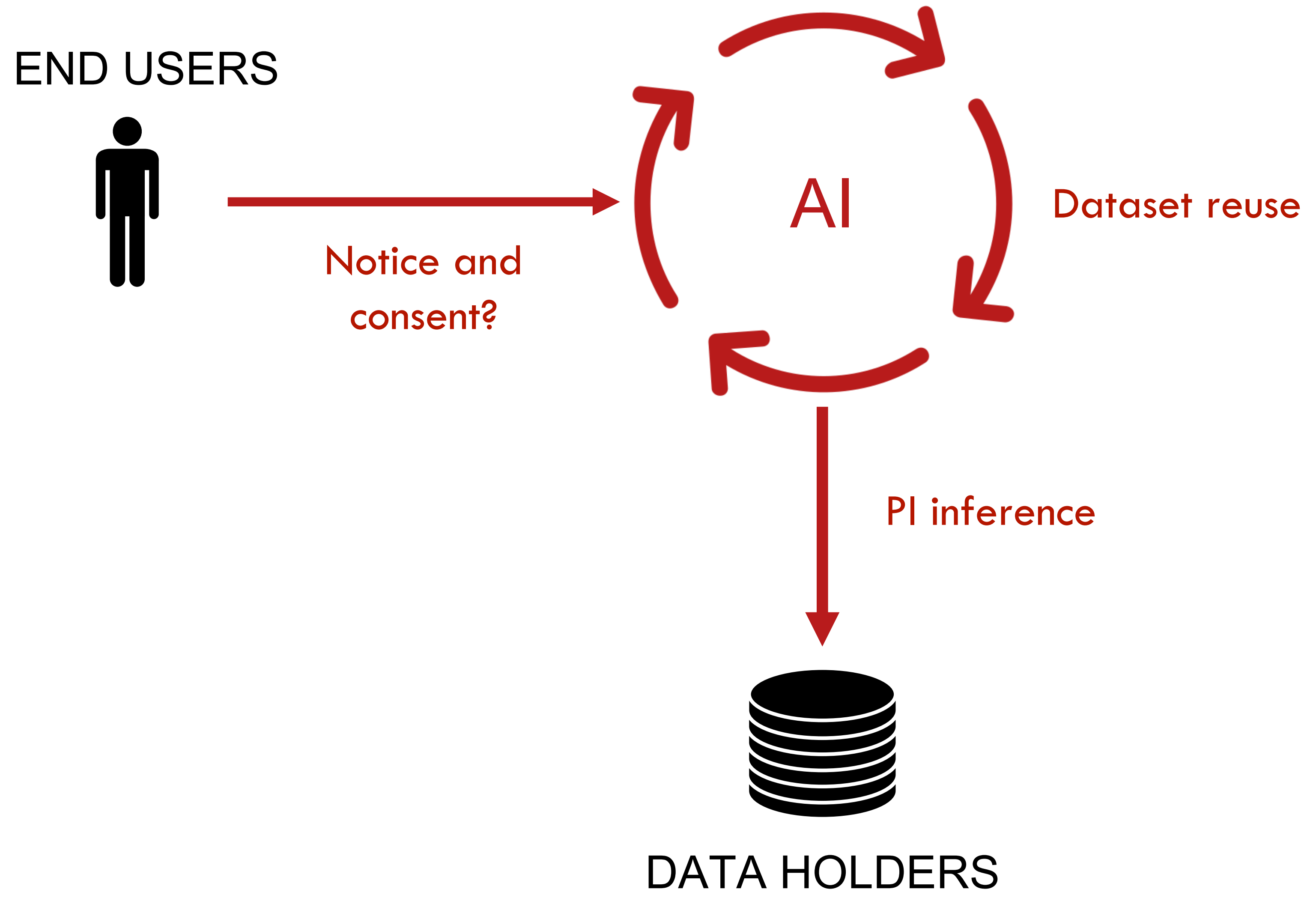
QTCinderella

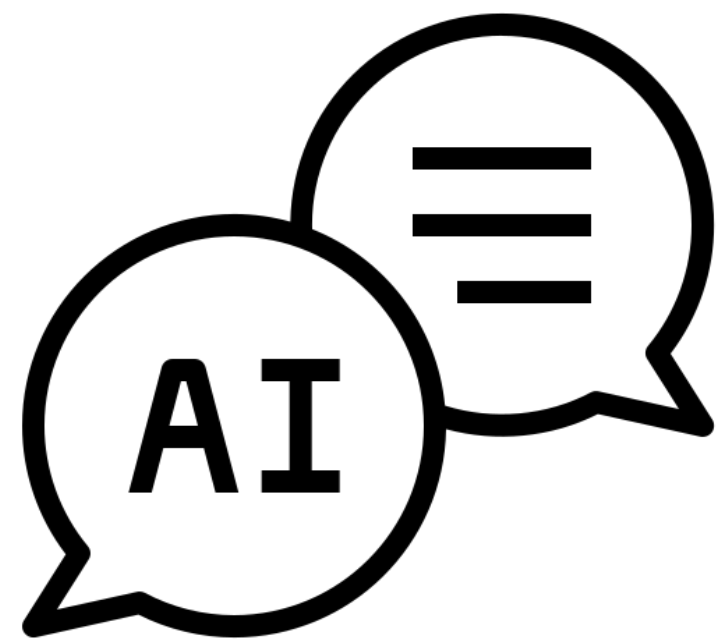
@qtcinderella · Follow

I want to scream. Stop. Everybody fucking stop. Stop spreading it. Stop advertising it. Stop. Being seen “naked” against your will should NOT BE A PART OF THIS JOB.

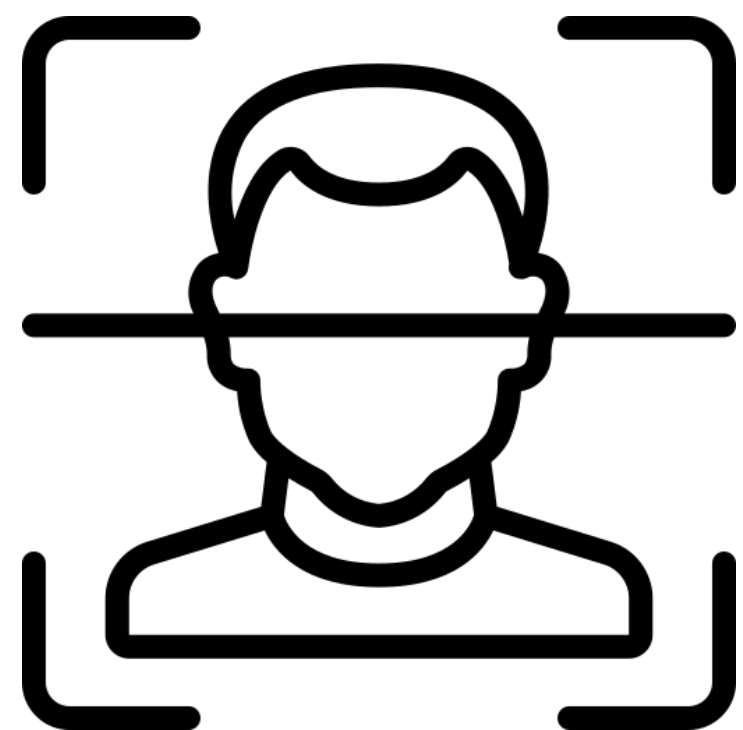
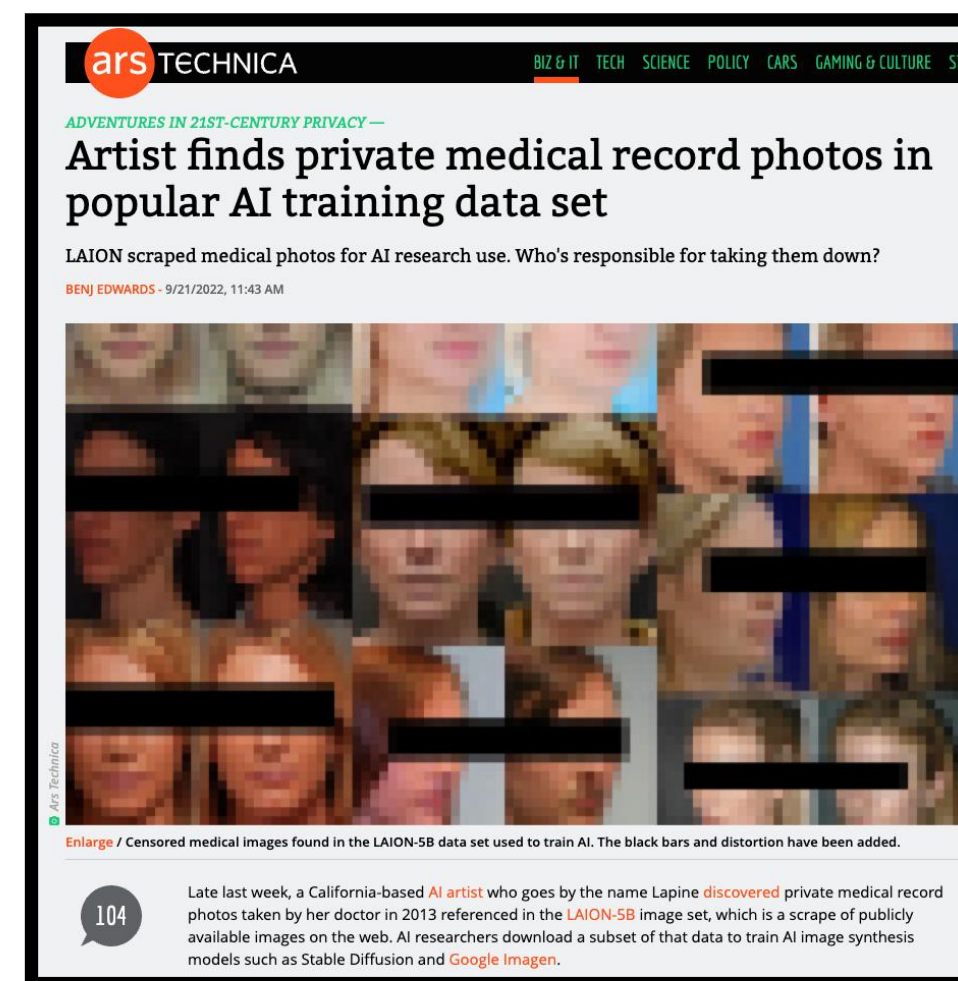
Thank you to all the male internet “journalists” reporting on this issue. Fucking losers. @LUNAR



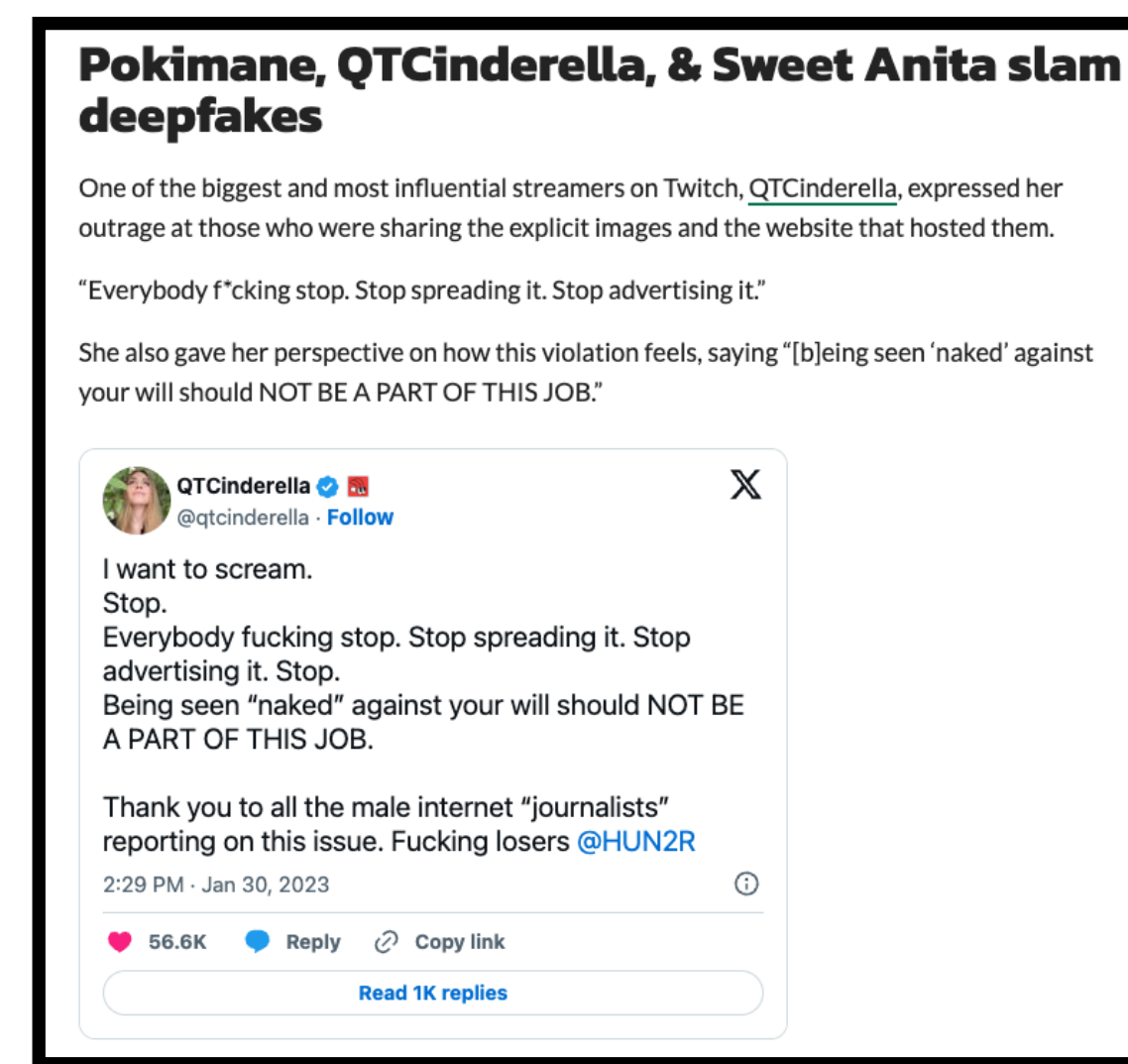
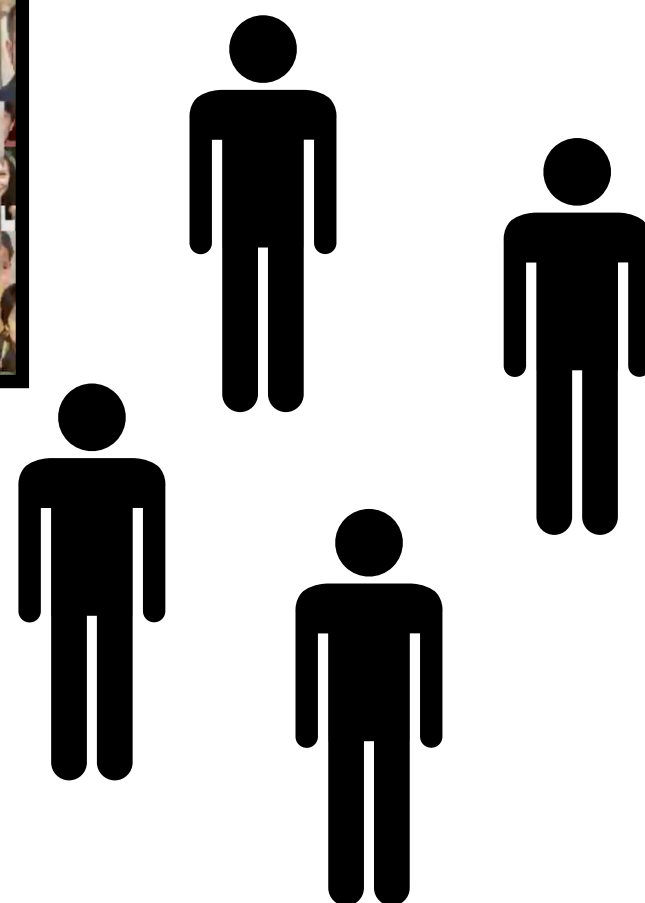




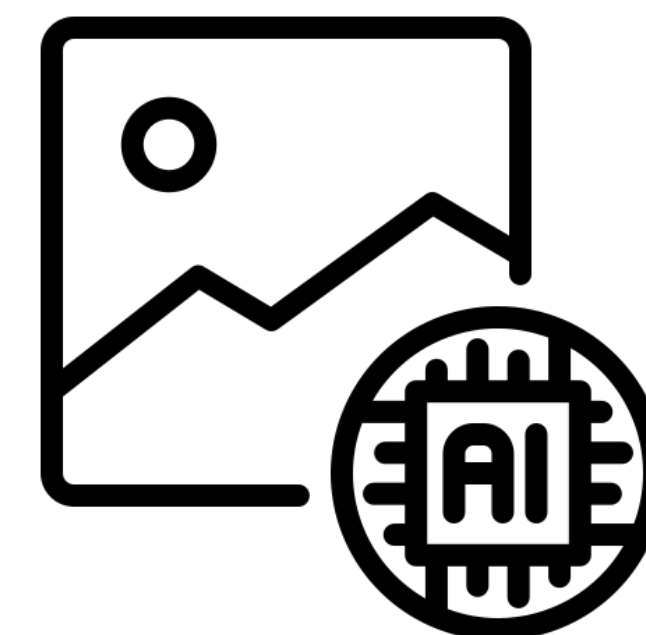
Large language model
Training models with
personal sensitive data



Facial recognition
Large-scale identity identification

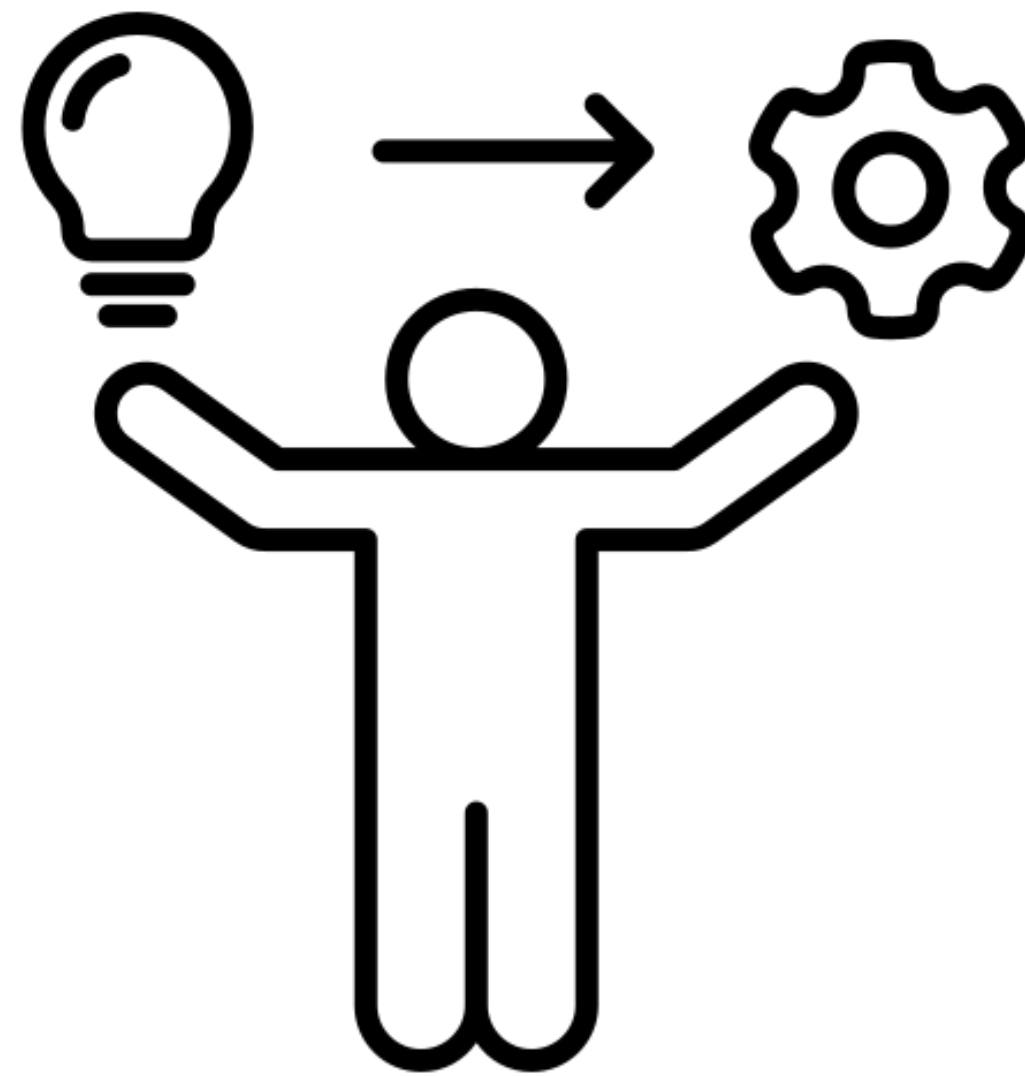


Creating realistic fake images



Diffusion-based algorithm

Product Concepts
Privacy Principles



Practices

Practitioners

How can we design tools that make it easier for practitioners to identify and mitigate privacy risks in their AI products and services?

Design tools that help practitioners address privacy risks of their AI product ideas

Privacy impact assessments

-  demand expert knowledge
-  slow to adapt to new challenges

LLM-powered interactive interfaces to surface AI ethics

-  privacy specific
-  end-to-end risk mitigation

Formative Study

Privy beta

Formative study with 11 AI and privacy practitioners using the beta prototype:

The screenshot shows the 'EXPLORE' phase of the Privy AI Privacy Risk Assistant. The interface is divided into several sections:

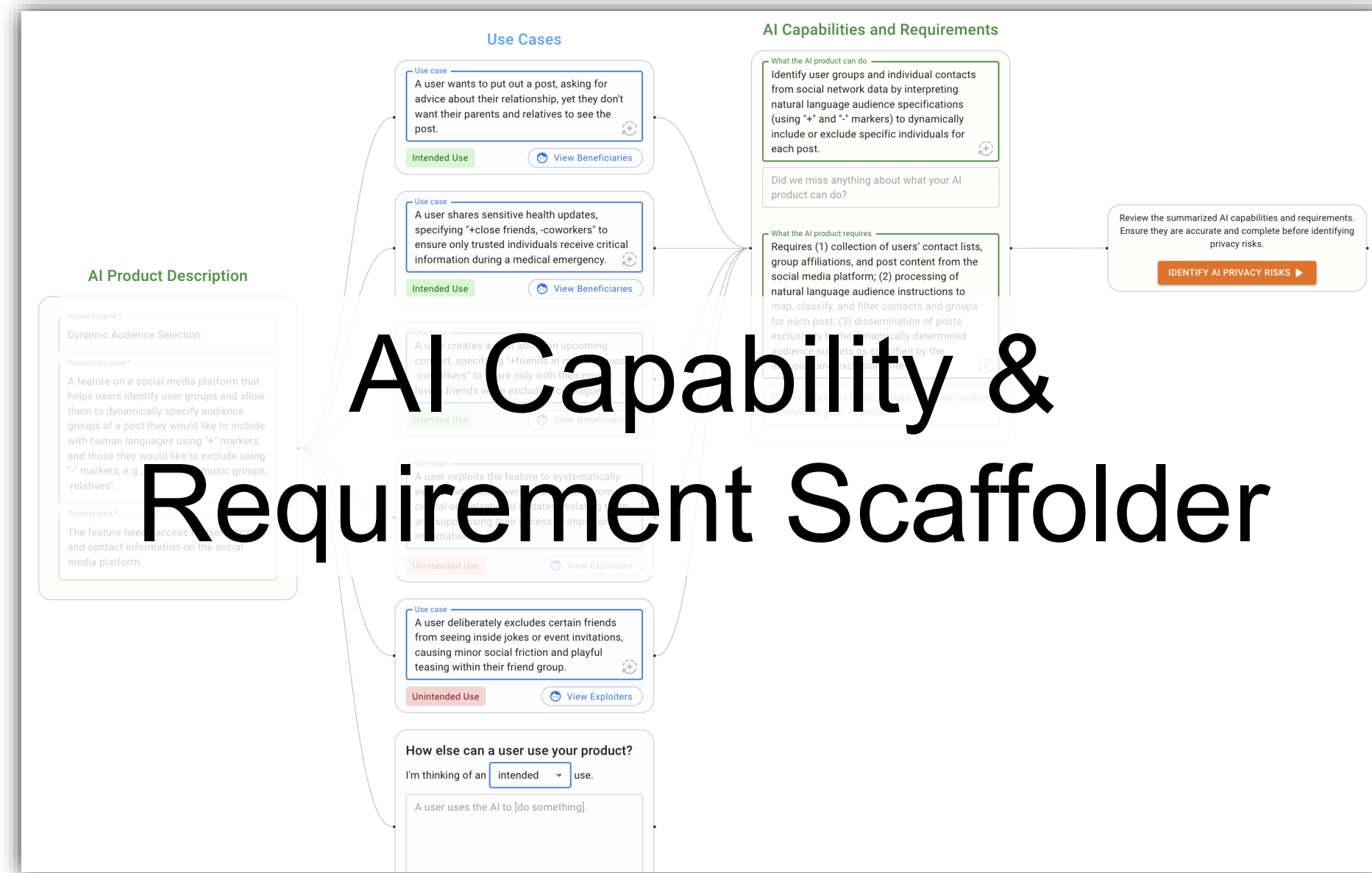
- Product Information:** Product Name (Dynamic Audience Selection) and Product Description (A feature on a social media platform that allows users to dynamically specify audience groups of a post they would like to include using "+" markers, and those they would like to exclude using "-" markers, e.g., "+friends in music groups, -relatives. The feature needs access to users' contact information and their interactions.').
- AI Capabilities and Requirements:** A section with a 'Capability Description' (Enables users to dynamically specify audience groups for social media posts by including or excluding specific contacts) and a 'Requirement Description' (Requires access to users' contact information and data on their interactions within the social media platform).
- Use Cases:** A central section with three use cases: 1) 'A user utilizes the AI to target specific audience groups from a user's social media contacts for a fundraising campaign, ensuring that the message reaches individuals most likely to contribute, thereby maximizing financial support for a critical health initiative.' (Intended). 2) 'A user exploits the AI to create targeted social media posts that incite hatred or violence against specific individuals or groups, leading to real-world harassment and threats.' (Unintended). 3) 'How else can someone use your tool?' (Intended).
- Impacted Stakeholders:** A section with three stakeholders: 1) 'Individuals whose contact information is accessed without their consent' (Intended). 2) 'Contacts who are excluded from the fundraising campaign and may feel alienated.' (Unintended). 3) 'Who else might be impacted in terms of privacy?' (Unintended).
- Privacy Risks:** A section with five risks: 1) 'Surveillance' (Individuals whose contact information is accessed without their consent may be subjected to surveillance as their interactions and preferences are monitored to tailor social media posts). 2) 'Insecurity' (Individuals whose contact information is accessed without their consent may have their personal data inadequately secured, leading to unauthorized access or leaks that could expose their private information to unwanted solicitation or harassment). 3) 'Aggregation' (Individuals whose contact information is accessed without their consent may have their social connections and interaction patterns aggregated, leading to inferences about their interests, financial capabilities, and willingness to support certain causes). 4) 'Exposure' (Individuals whose contact information is accessed without their consent may have their personal relationships and social connections exposed, revealing sensitive details about their social circles and financial situations). 5) 'Are there any other privacy risks your product entails?' (A risk type selector and a text input field for a privacy risk).

privacy risk explorer

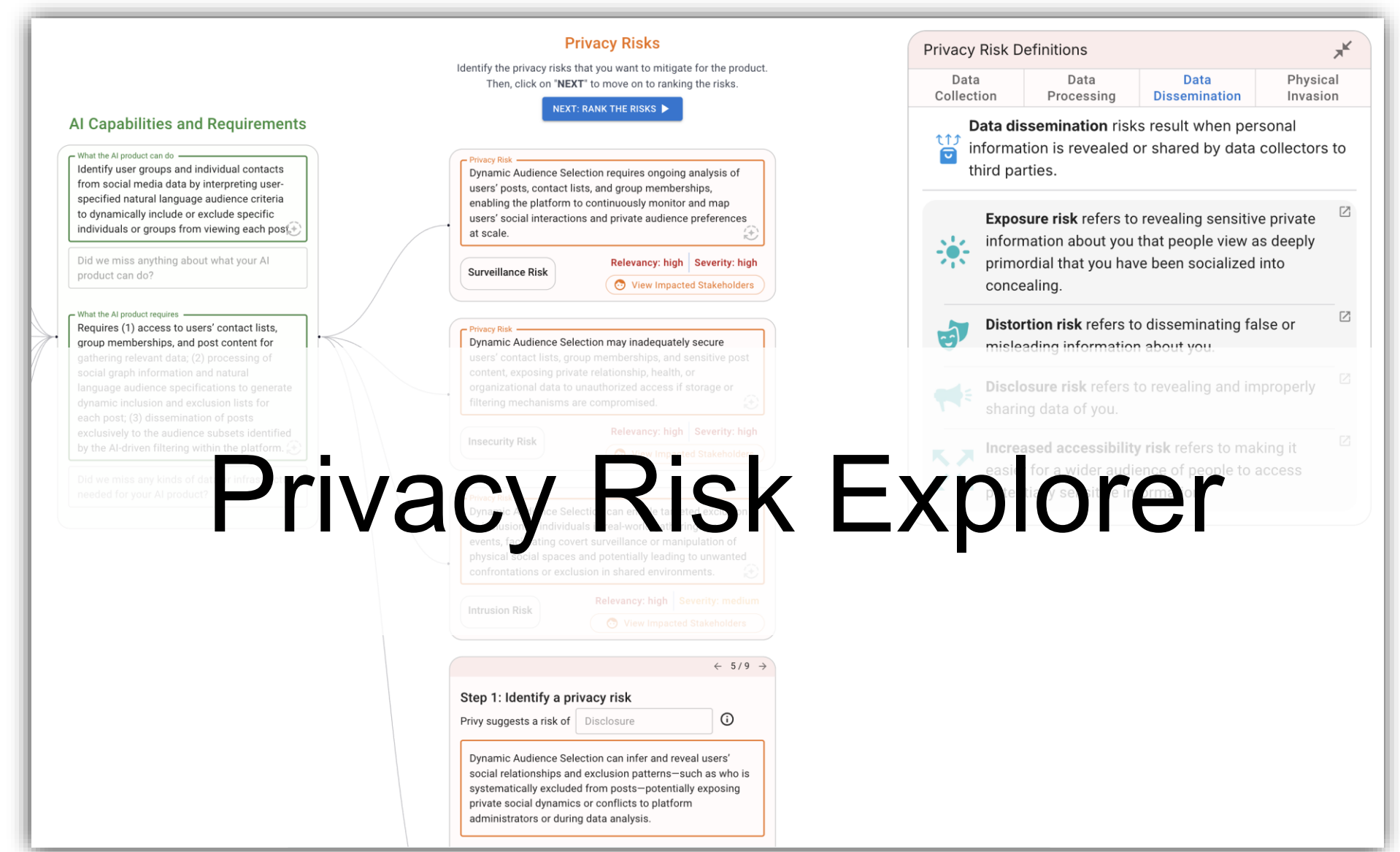
The screenshot shows the 'SUMMARIZE' phase of the Privy AI Privacy Risk Assistant. The interface displays a summary of the 'Targeted Fundraising Campaigns' use case, including impacted stakeholders and privacy risks.

- Filter:** A dropdown menu with 'Intended' and 'Misuse' options.
- Targeted Fundraising Campaigns:** A green header section with a 'View Beneficiaries' button. The summary text reads: 'A user employs the AI to target specific audience groups for a fundraising campaign on social media, ensuring that the message reaches individuals most likely to contribute, thereby maximizing financial support for a critical health initiative.'
- Impacted Stakeholders and Privacy Risks:** A section with a '1' indicator. It lists three risks: 1) 'Surveillance Risk' (Potential donors who may feel uncomfortable with targeted messaging could be subjected to increased surveillance as their interactions and preferences are analyzed to refine audience targeting). 2) 'Insecurity Risk' (Potential donors who may feel uncomfortable with targeted messaging could have their contact information and interaction data inadequately secured, leading to unauthorized access or leaks that expose their personal preferences and social connections). 3) 'Disclosure Risk' (Potential donors who may feel uncomfortable with targeted messaging could have their contact information and interaction data disclosed to unauthorized parties, leading to unwanted solicitations). 4) 'Aggregation Risk' (Potential donors who may feel uncomfortable with targeted messaging could have their contact information and interaction data aggregated, leading to inferences about their interests, financial capabilities, and willingness to support certain causes).
- Privacy Incidents Relevant to the Product and Use Case:** A section with two incidents: 1) 'Meta under fire for decision to train generative AI on user content' (with links to BBC and SWGL). 2) 'Meta admits farming Australians' Facebook photos to train AI' (with links to The Conversation and The Guardian).

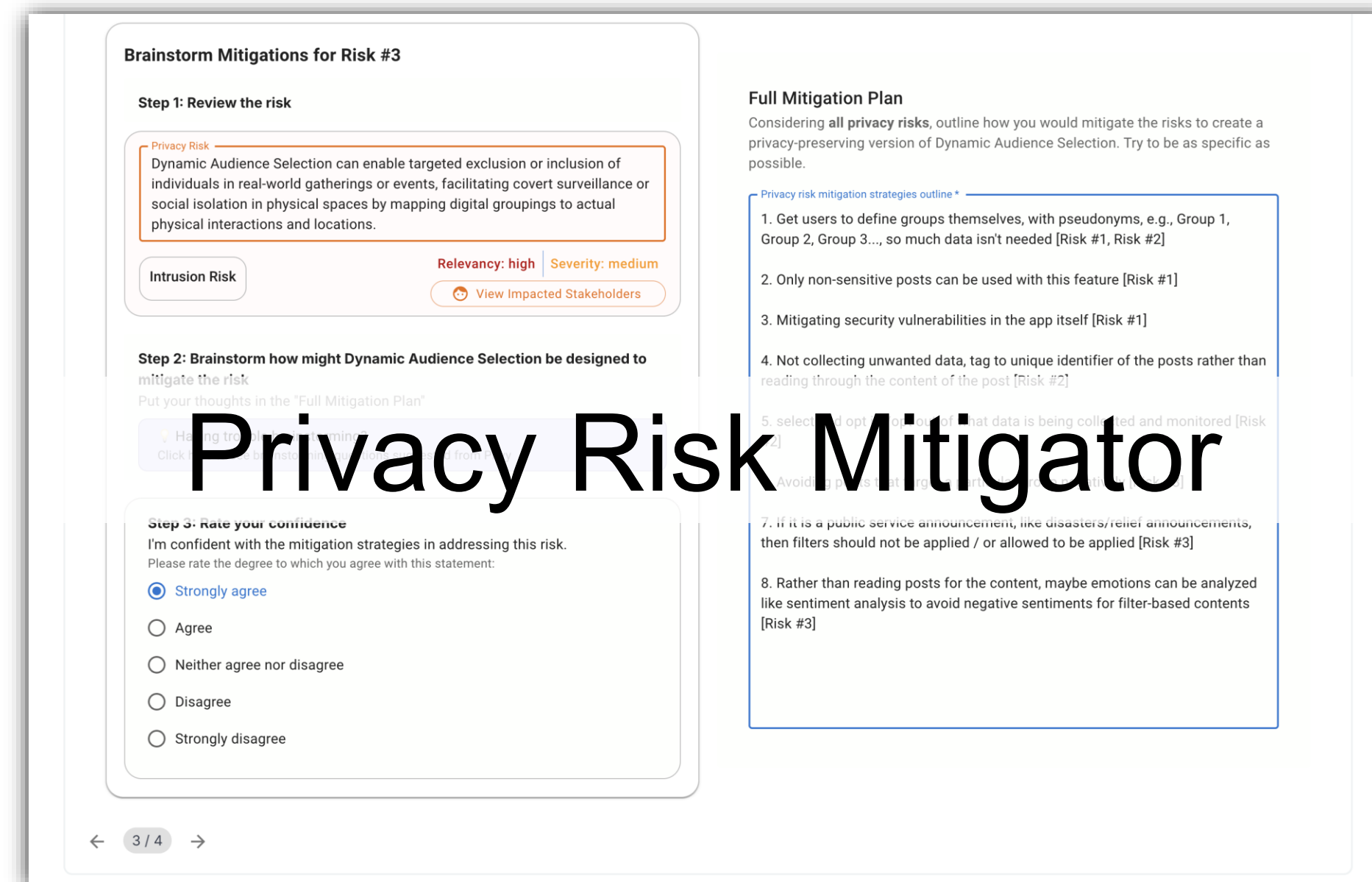
privacy risk summarizer



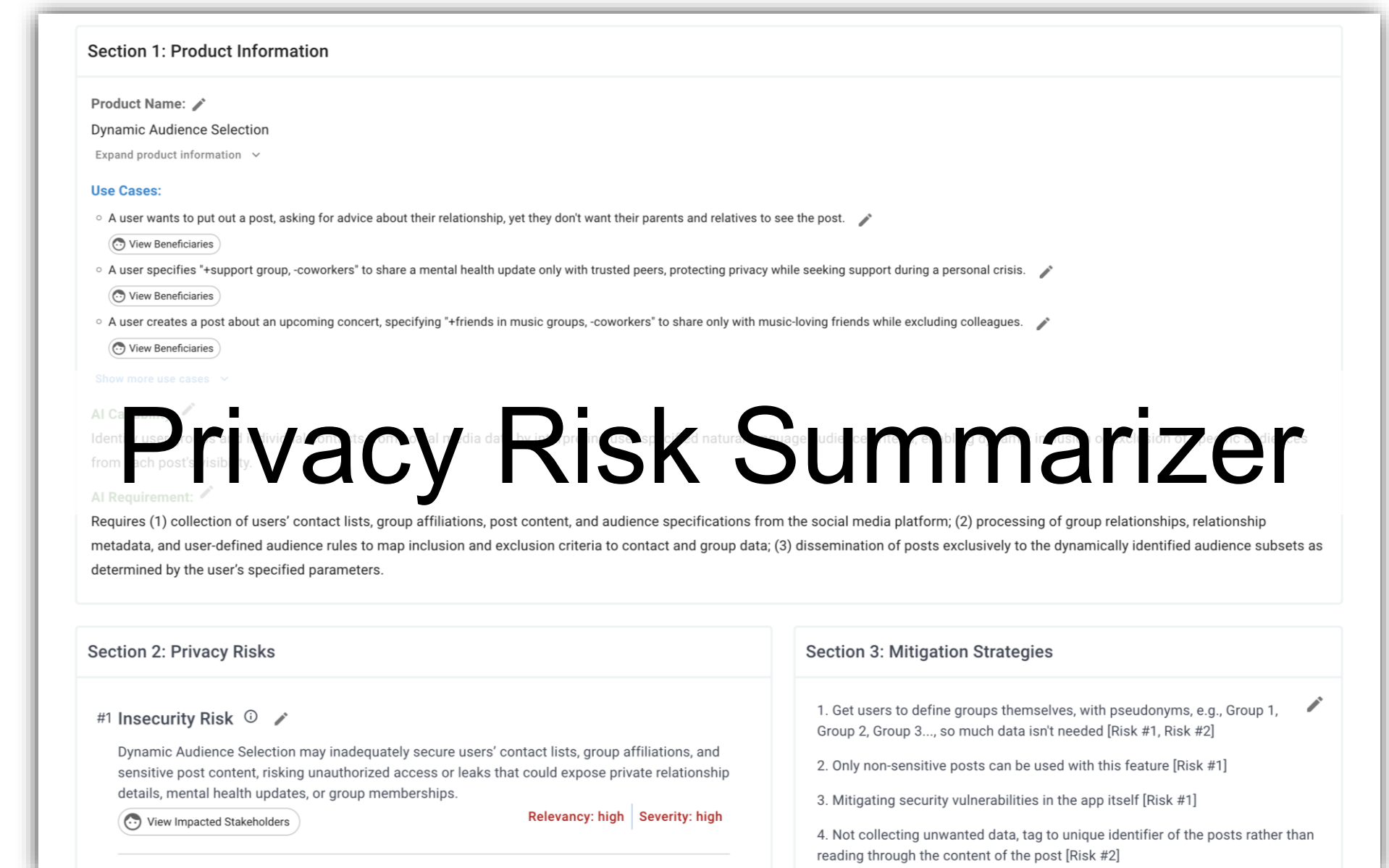
AI Capability & Requirement Scaffolder



Privacy Risk Explorer



Privacy Risk Mitigator



Privacy Risk Summarizer

Design Goal #1:

Help practitioners elicit the AI capabilities and requirements that entail privacy risks

"This is helpful for seeing what types of data... initially [I] didn't really think about the capabilities, but the underlying requirements and how that really requires sensitive user data." (U8)

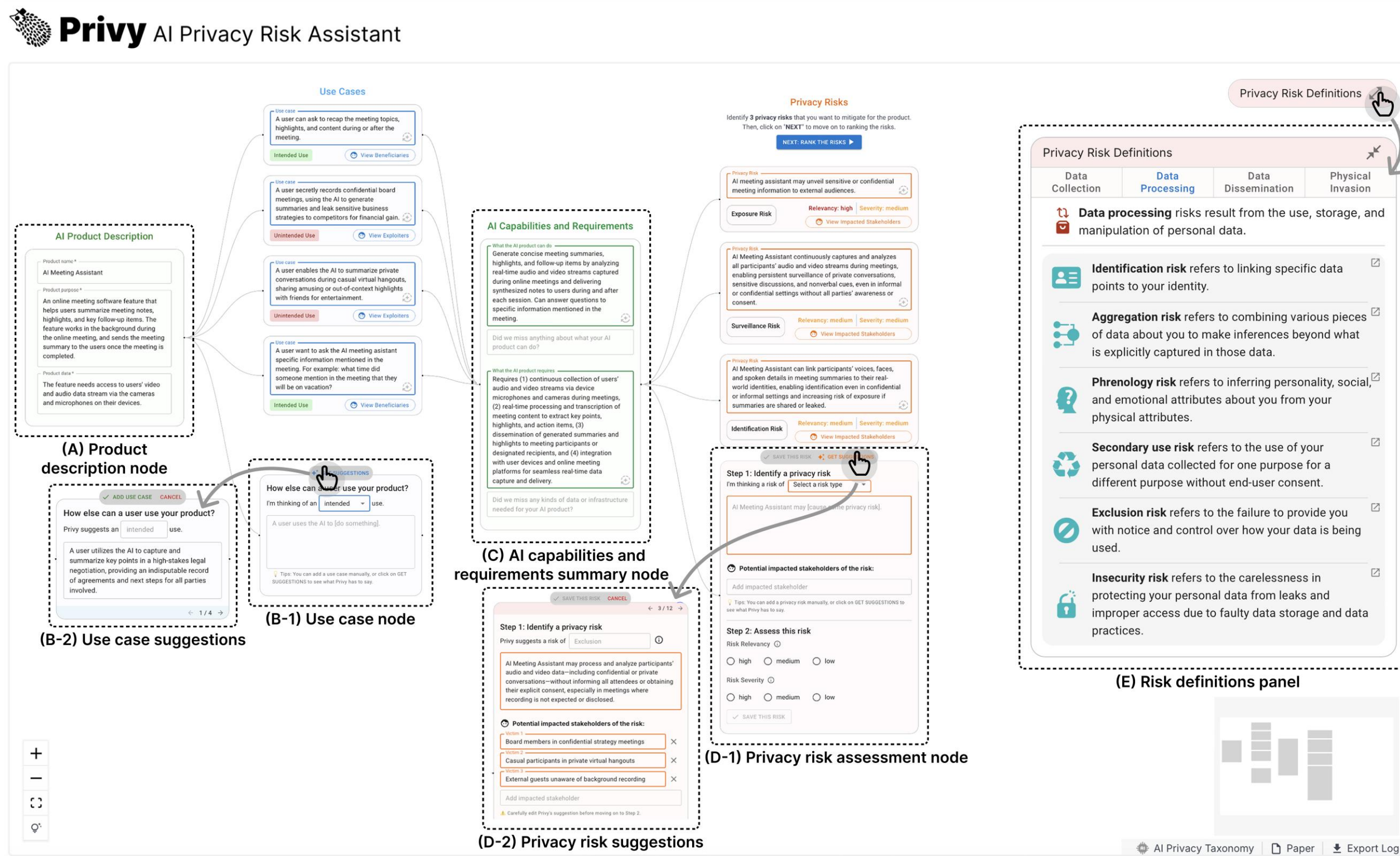
AI Capability & Requirement Scaffold

Guides users through (1) brainstorming use cases, and (2) summarizing resulting AI capabilities and requirements



Summative Study

Two Privy Variants



AI Privacy Template

Product Name:

Product Purpose:

Product Data:

Section 1: Product Information

Potential Use Cases and Users: 1) What potential intended or unintended use cases could arise with this product, and 2) who are the entities or individuals that will use and benefit from the product?

- **Use Case 1 | Select Type** : A user uses the product to...
 - User(s):
- **Use Case 2 | Select Type** :
- User(s):
- **Use Case 3 | Select Type** :
- User(s):
- **Use Case 4 | Select Type** :
- User(s):

AI Capabilities: What can the AI do?

-
-

AI Requirements: What data or infrastructure is needed for this AI to work?

-
-

Section 2: Privacy Risks

Step 1 Identify: Identify privacy risks entailed by the product (1.1), their associated impacted stakeholders (1.2), and their relevance and severity (1.3).
Step 2 Rank: Rank the order (1st - 3rd) in which you think these risks should be addressed.

Relevancy: Considering the product purpose, this risk is relevant to the AI user experience:

- **High:** The risk is highly relevant and closely tied to the use cases and product.
- **Medium:** The risk is somewhat relevant to the use cases and product, though the relevance might be indirect and speculative in certain situations.
- **Low:** The risk has minimal direct relevance to the use cases and product. In rare instances, it may have indirect relevance (e.g., speculated downstream impact).

Severity: Considering the product purpose, the risk poses a significant impact on society and/or individuals due to the AI:

- **High:** The risk is substantial and directly harmful on an individual and/or societal scale.
- **Medium:** The risk is noticeable and harmful at the individual level and/or societal level.
- **Low:** The risk is negligible at both the individual and societal scales.

Privacy Risk #	Step 1.1: Describe the privacy risks you think might arise from this product.	Step 1.2: List the potential impacted stakeholders of the risk.	Step 1.3: Rate the relevancy and severity of the risk.	Step 2: In what order do you think these risks should be addressed?
#1	Risk type: <i>Please select</i> Risk description:		Relevancy score: <i>Please select</i> Severity score: <i>Please select</i>	<i>Please select</i>
#2	Risk type: <i>Please select</i> Risk description:		Relevancy score: <i>Please select</i> Severity score: <i>Please select</i>	<i>Please select</i>
#3	Risk type: <i>Please select</i> Risk description:		Relevancy score: <i>Please select</i> Severity score: <i>Please select</i>	<i>Please select</i>

Section 3: Mitigation Strategies

Outline mitigation strategies associated with the three privacy risks you identified (3.1), and rate how confident you are these strategies address those risks (3.2).

Step 3.1: Considering all three privacy risks, outline how you would mitigate the risks to create a privacy-preserving version of the product. Try to be as specific as possible.

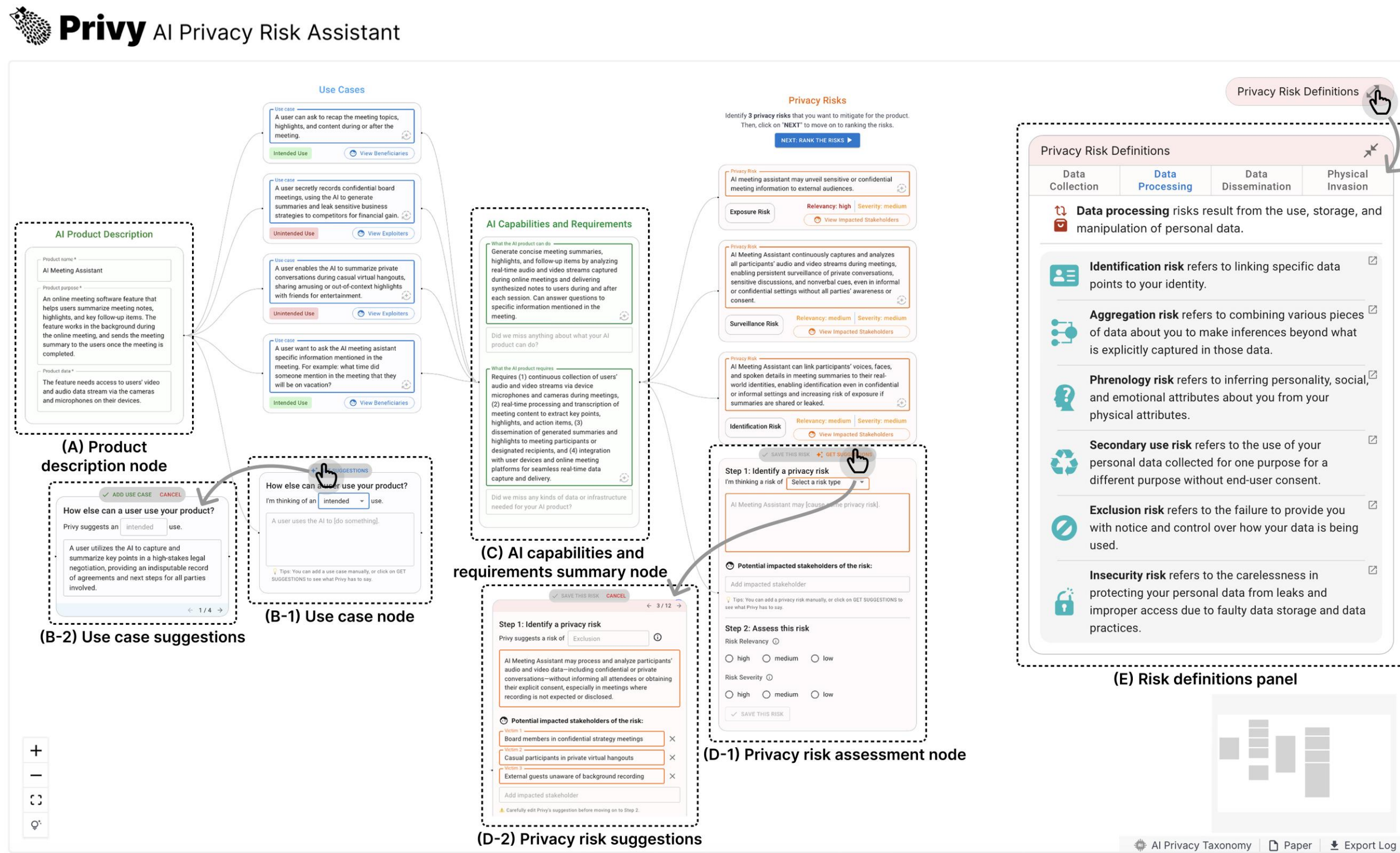
Step 3.2: I'm confident in the proposed mitigation strategies for addressing this risk.
 Please rate the degree to which you agree with this statement for each identified risk.

- For each risk, you might think about:
- How can the product be designed to enhance users' awareness of the risk?
 - How can the product be designed to enhance users' motivation to address the risk?
 - How can the product be designed to provide users with the ability to manage the risk?

Privacy Risk #	
#1	<i>Please select</i>
#2	<i>Please select</i>
#3	<i>Please select</i>

Summative Study

Two Privy Variants



AI Privacy Template

Product Name: _____

Product Purpose: _____

Product Data: _____

Section 1: Product Information

Potential Use Cases and Users: 1) What potential *intended* or *unintended* use cases could arise with this product, and 2) who are the entities or individuals that will use and benefit from the product?

- Use Case 1 | Select Type: A user uses the product to...
 - User(s):
- Use Case 2 | Select Type:
 - User(s):
- Use Case 3 | Select Type:
 - User(s):
- Use Case 4 | Select Type:
 - User(s):

AI Capabilities: What can the AI do?

-
-

AI Requirements: What data or infrastructure is needed for this AI to work?

-
-

Section 2: Privacy Risks

Step 1 Identify: Identify privacy risks entailed by the product (1.1), their associated impacted stakeholders (1.2), and their relevance and severity (1.3).
 Step 2 Rank: Rank the order (1st - 3rd) in which you think these risks should be addressed.

Relevancy: Considering the product purpose, this risk is relevant to the AI user experience:

- High: The risk is highly relevant and closely tied to the use cases and product.
- Medium: The risk is somewhat relevant to the use cases and product, though the relevance might be indirect and speculative in certain situations.
- Low: The risk has minimal direct relevance to the use cases and product. In rare instances, it may have indirect relevance (e.g., speculated downstream impact).

Severity: Considering the product purpose, the risk poses a significant impact on society and/or individuals due to the AI:

- High: The risk is substantial and directly harmful on an individual and/or societal scale.
- Medium: The risk is noticeable and harmful at the individual level and/or societal level.
- Low: The risk is negligible at both the individual and societal scales.

Privacy Risk #	Step 1.1: Describe the privacy risks you think might arise from this product.	Step 1.2: List the potential impacted stakeholders of the risk.	Step 1.3: Rate the relevancy and severity of the risk.	Step 2: In what order do you think these risks should be addressed?
#1	Risk type: <i>Please select</i> Risk description:		Relevancy score: <i>Please select</i> Severity score: <i>Please select</i>	<i>Please select</i>
#2	Risk type: <i>Please select</i> Risk description:		Relevancy score: <i>Please select</i> Severity score: <i>Please select</i>	<i>Please select</i>
#3	Risk type: <i>Please select</i> Risk description:		Relevancy score: <i>Please select</i> Severity score: <i>Please select</i>	<i>Please select</i>

Section 3: Mitigation Strategies

Outline mitigation strategies associated with the three privacy risks you identified (3.1), and rate how confident you are these strategies address those risks (3.2).

Step 3.1: Considering all three privacy risks, outline how you would mitigate the risks to create a privacy-preserving version of the product. Try to be as specific as possible.

Step 3.2: I'm confident in the proposed mitigation strategies for addressing this risk.
 Please rate the degree to which you agree with this statement for each identified risk.

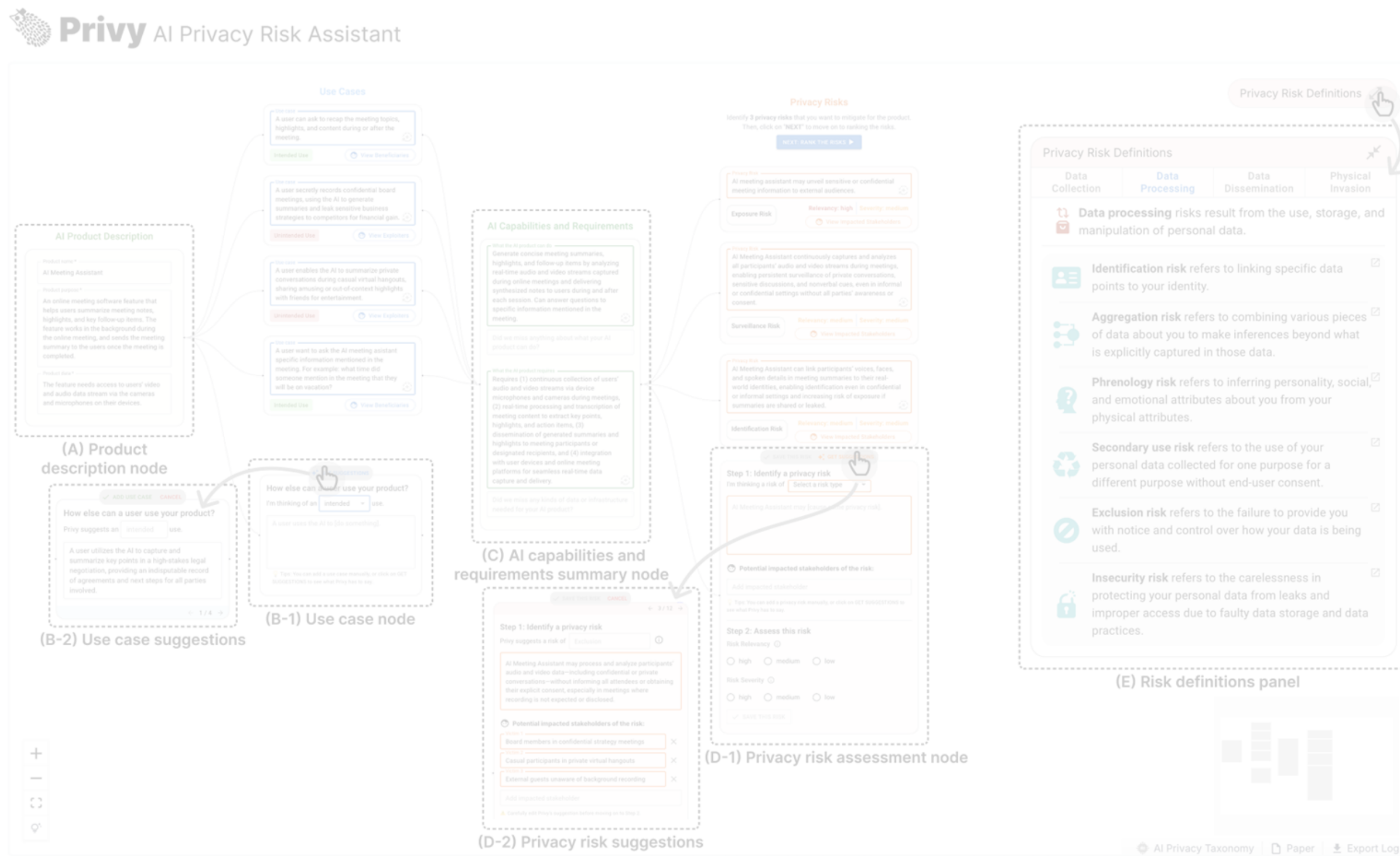
Privacy Risk #	Confidence
#1	<i>Please select</i>
#2	<i>Please select</i>
#3	<i>Please select</i>

For each risk, you might think about:

- How can the product be designed to enhance users' awareness of the risk?
- How can the product be designed to enhance users' motivation to address the risk?
- How can the product be designed to provide users with the ability to manage the risk?

Summative Study

Two Privy Variants



AI Privacy Template

Product Name:

Product Purpose:

Product Data:

Section 1: Product Information

Potential Use Cases and Users: 1) What potential intended or unintended use cases could arise with this product, and 2) who are the entities or individuals that will use and benefit from the product?

- **Use Case 1 | Select Type** : A user uses the product to...
 - User(s):
- **Use Case 2 | Select Type** :
- User(s):
- **Use Case 3 | Select Type** :
- User(s):
- **Use Case 4 | Select Type** :
- User(s):

AI Capabilities: What can the AI do?

-
-

AI Requirements: What data or infrastructure is needed for this AI to work?

-
-

Section 2: Privacy Risks

Step 1 Identify: Identify privacy risks entailed by the product (1.1), their associated impacted stakeholders (1.2), and their relevance and severity (1.3).
Step 2 Rank: Rank the order (1st - 3rd) in which you think these risks should be addressed.

Relevancy: Considering the product purpose, this risk is relevant to the AI user experience:

- **High:** The risk is highly relevant and closely tied to the use cases and product.
- **Medium:** The risk is somewhat relevant to the use cases and product, though the relevance might be indirect and speculative in certain situations.
- **Low:** The risk has minimal direct relevance to the use cases and product. In rare instances, it may have indirect relevance (e.g., speculated downstream impact).

Severity: Considering the product purpose, the risk poses a significant impact on society and/or individuals due to the AI:

- **High:** The risk is substantial and directly harmful on an individual and/or societal scale.
- **Medium:** The risk is noticeable and harmful at the individual level and/or societal level.
- **Low:** The risk is negligible at both the individual and societal scales.

Privacy Risk #	Step 1.1: Describe the privacy risks you think might arise from this product.	Step 1.2: List the potential impacted stakeholders of the risk.	Step 1.3: Rate the relevancy and severity of the risk.	Step 2: In what order do you think these risks should be addressed?
#1	Risk type: Please select Risk description:		Relevancy score: Please select Severity score: Please select	Please select
#2	Risk type: Please select Risk description:		Relevancy score: Please select Severity score: Please select	Please select
#3	Risk type: Please select Risk description:		Relevancy score: Please select Severity score: Please select	Please select

Section 3: Mitigation Strategies

Outline mitigation strategies associated with the three privacy risks you identified (3.1), and rate how confident you are these strategies address those risks (3.2).

Step 3.1: Considering all three privacy risks, outline how you would mitigate the risks to create a privacy-preserving version of the product. Try to be as specific as possible.

For each risk, you might think about:

- How can the product be designed to enhance users' awareness of the risk?
- How can the product be designed to enhance users' motivation to address the risk?
- How can the product be designed to provide users with the ability to manage the risk?

Step 3.2: I'm confident in the proposed mitigation strategies for addressing this risk.

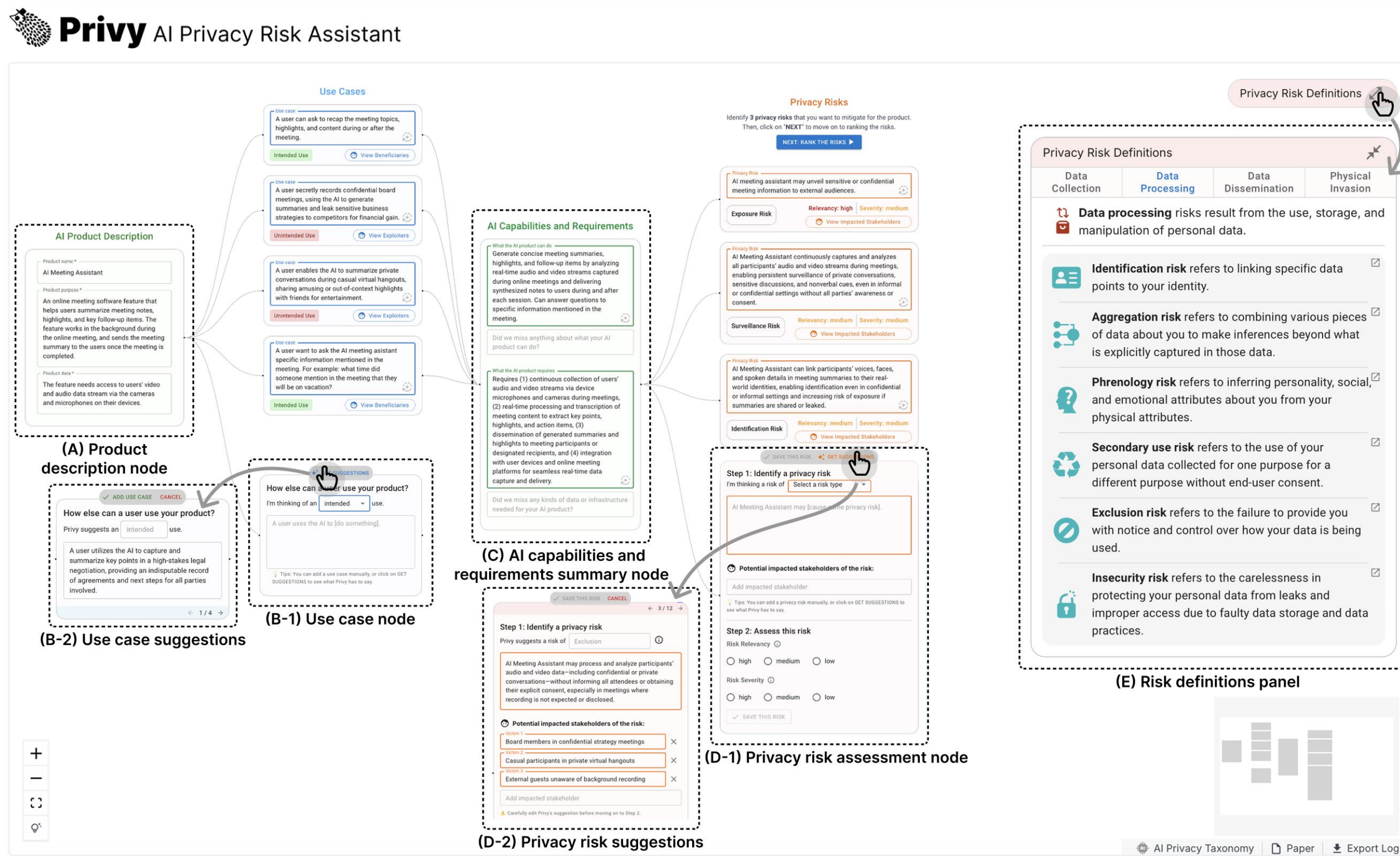
Please rate the degree to which you agree with this statement for each identified risk.

Privacy Risk #

#1	Please select
#2	Please select
#3	Please select

Summative Study

Two Privy Variants



AI Privacy Template

Product Name:

Product Purpose:

Product Data:

Section 1: Product Information

Potential Use Cases and Users: 1) What potential intended or unintended use cases could arise with this product, and 2) who are the entities or individuals that will use and benefit from the product?

- **Use Case 1 | Select Type** : A user uses the product to...
 - User(s):
- **Use Case 2 | Select Type** :
- User(s):
- **Use Case 3 | Select Type** :
- User(s):
- **Use Case 4 | Select Type** :
- User(s):

AI Capabilities: What can the AI do?

-
-

AI Requirements: What data or infrastructure is needed for this AI to work?

-
-

Section 2: Privacy Risks

Step 1 Identify: Identify privacy risks entailed by the product (1.1), their associated impacted stakeholders (1.2), and their relevance and severity (1.3).

Step 2 Rank: Rank the order (1st - 3rd) in which you think these risks should be addressed.

Relevancy: Considering the product purpose, this risk is relevant to the AI user experience:

- **High:** The risk is highly relevant and closely tied to the use cases and product.
- **Medium:** The risk is somewhat relevant to the use cases and product, though the relevance might be indirect and speculative in certain situations.
- **Low:** The risk has minimal direct relevance to the use cases and product. In rare instances, it may have indirect relevance (e.g., speculated downstream impact).

Severity: Considering the product purpose, the risk poses a significant impact on society and/or individuals due to the AI:

- **High:** The risk is substantial and directly harmful on an individual and/or societal scale.
- **Medium:** The risk is noticeable and harmful at the individual level and/or societal level.
- **Low:** The risk is negligible at both the individual and societal scales.

Privacy Risk #	Step 1.1: Describe the privacy risks you think might arise from this product.	Step 1.2: List the potential impacted stakeholders of the risk.	Step 1.3: Rate the relevancy and severity of the risk.	Step 2: In what order do you think these risks should be addressed?
#1	Risk type: <i>Please select</i> Risk description:		Relevancy score: <i>Please select</i> Severity score: <i>Please select</i>	<i>Please select</i>
#2	Risk type: <i>Please select</i> Risk description:		Relevancy score: <i>Please select</i> Severity score: <i>Please select</i>	<i>Please select</i>
#3	Risk type: <i>Please select</i> Risk description:		Relevancy score: <i>Please select</i> Severity score: <i>Please select</i>	<i>Please select</i>

Section 3: Mitigation Strategies

Outline mitigation strategies associated with the three privacy risks you identified (3.1), and rate how confident you are these strategies address those risks (3.2).

Step 3.1: Considering all three privacy risks, outline how you would mitigate the risks to create a privacy-preserving version of the product. Try to be as specific as possible.

For each risk, you might think about:

- How can the product be designed to enhance users' awareness of the risk?
- How can the product be designed to enhance users' motivation to address the risk?
- How can the product be designed to provide users with the ability to manage the risk?

Step 3.2: I'm confident in the proposed mitigation strategies for addressing this risk.

Please rate the degree to which you agree with this statement for each identified risk.

Privacy Risk #

#1	<i>Please select</i>
#2	<i>Please select</i>
#3	<i>Please select</i>

Summative Study

Study Design and Data Analysis

Participants

- **24 AI practitioners** to complete privacy impact assessment (PIA) using Privy (evenly assigned to Privy-LLM or Privy-Template)
- **13 privacy experts** to evaluate the quality of the generated PIA reports

Data analysis

- Qualitative: think-aloud, interview
- Quantitative: self-report tool experience, expert-rated report quality

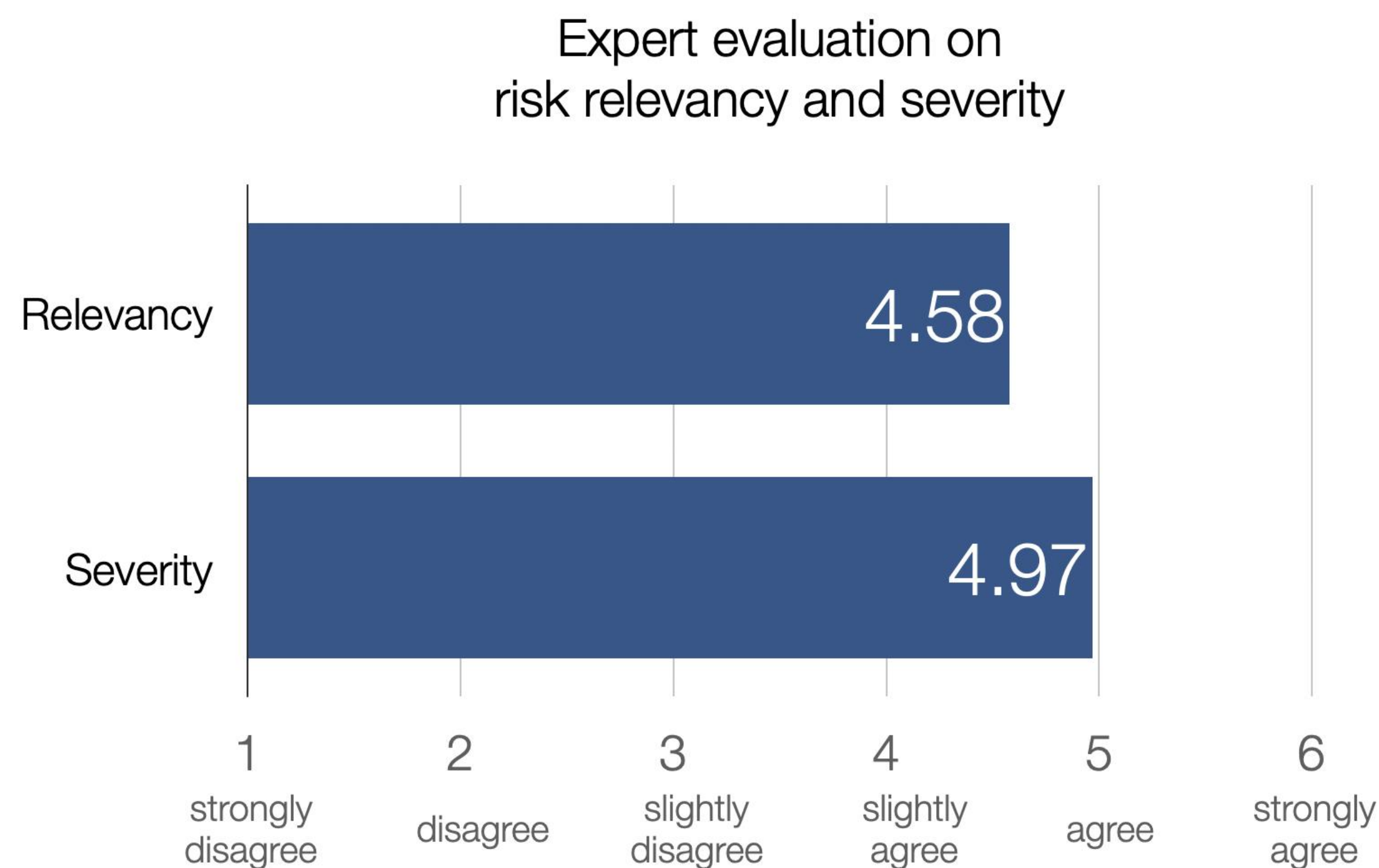
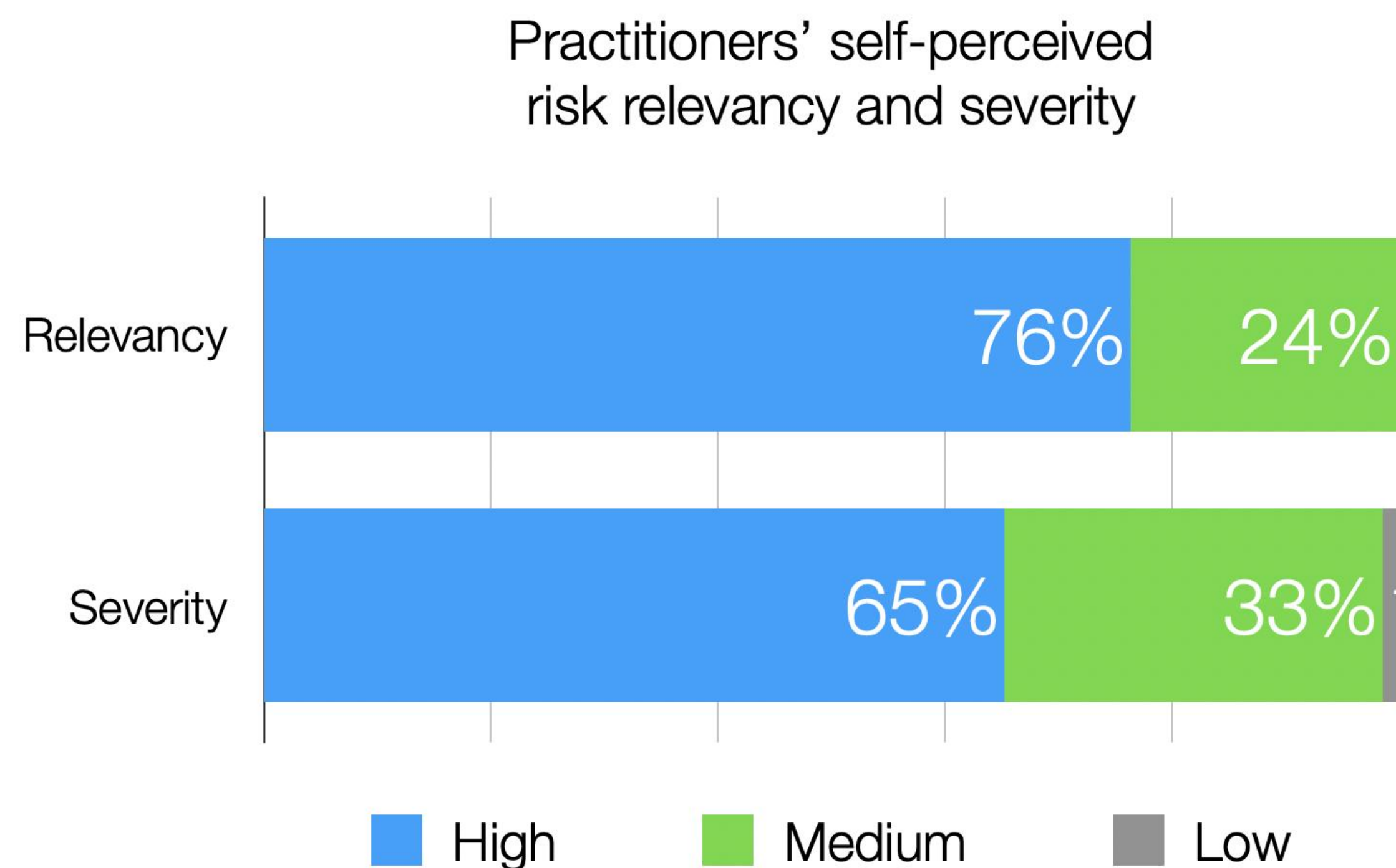
RQ1: How does Privy help practitioners identify and mitigate privacy risks with AI product concepts?

RQ2: How does the use of LLMs affect the quality of privacy impact assessments produced by Privy?

RQ3: To what extent does Privy address practitioners' challenges in privacy risk envisioning and mitigation?

Finding #1

Privy helps practitioners envision privacy risks that are relevant and severe



Finding #1

Privy helps practitioners envision privacy risks that are relevant and severe

Grounding risk identification with use cases:

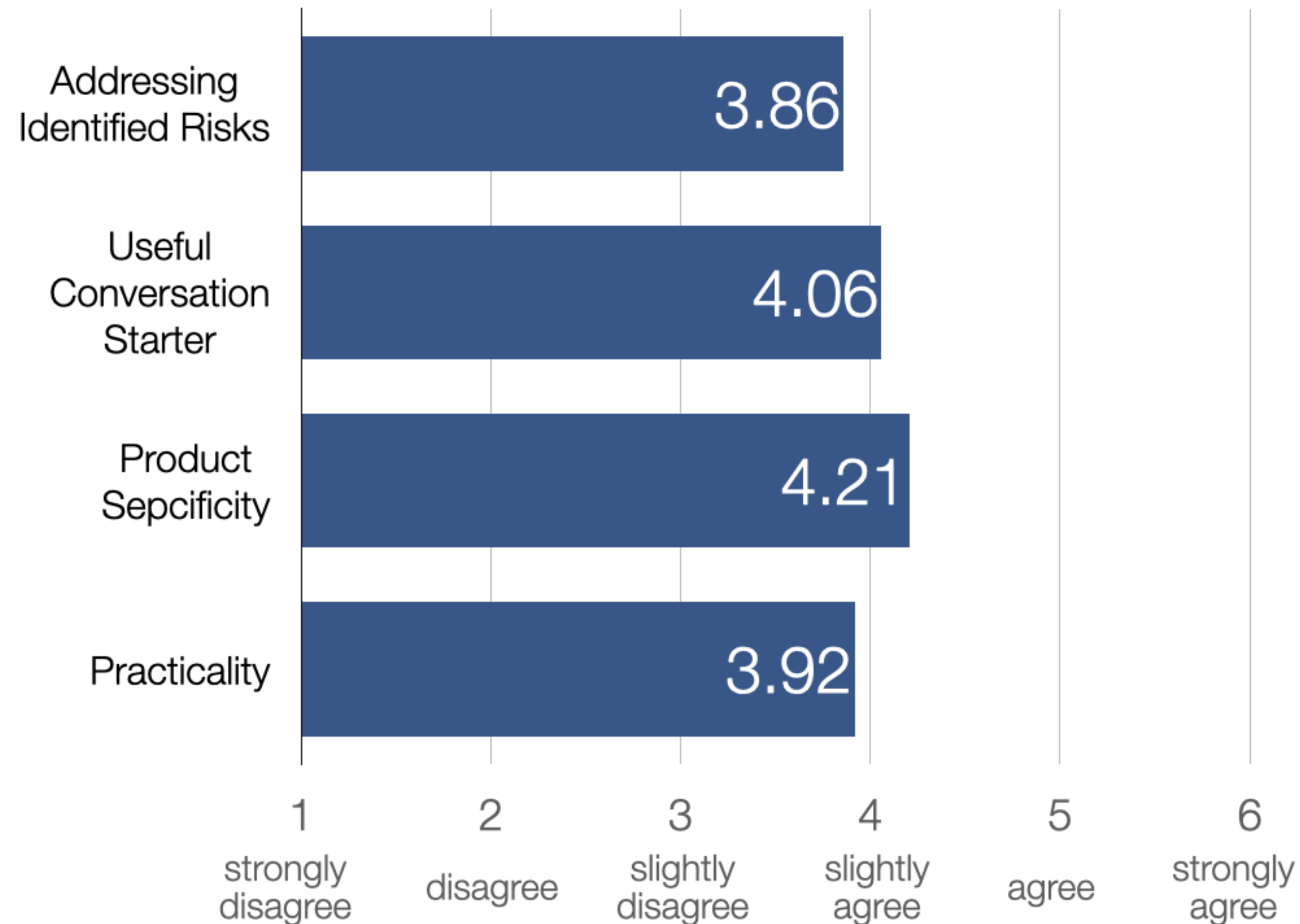
“managers can misuse this tool... to inaccurately assess webinar... the second [use case] that I mentioned here is... checking the portion of the meeting participant dialog to assess how people engaged in the meetings.” (P18)

Grounding risk identification with envisioned AI capabilities and requirements:

“infer the [users’] personality based on how they interact with posts... and then compare that with their physical image.” (P8)

Finding #2

Privy helps practitioners brainstorm risk mitigation plans that are effective and appropriate



Finding #2

Privy helps practitioners brainstorm risk mitigation plans that are effective and appropriate

Grounding risk mitigation in end-user perspectives:

“give the option to the user to choose the topics that can be used to target them... [and] give an option to choose the people with whom they want to share their [AI-inferred] information.” (P20)

Grounding risk mitigation in the utility-privacy tradeoff:

“you need audience graphs, social interaction, everything if you want to make the groups dynamically, if you want AI to generate the groups. But if you’re defining groups [to reduce the risk], I don’t think you need it [the product].” (P7)

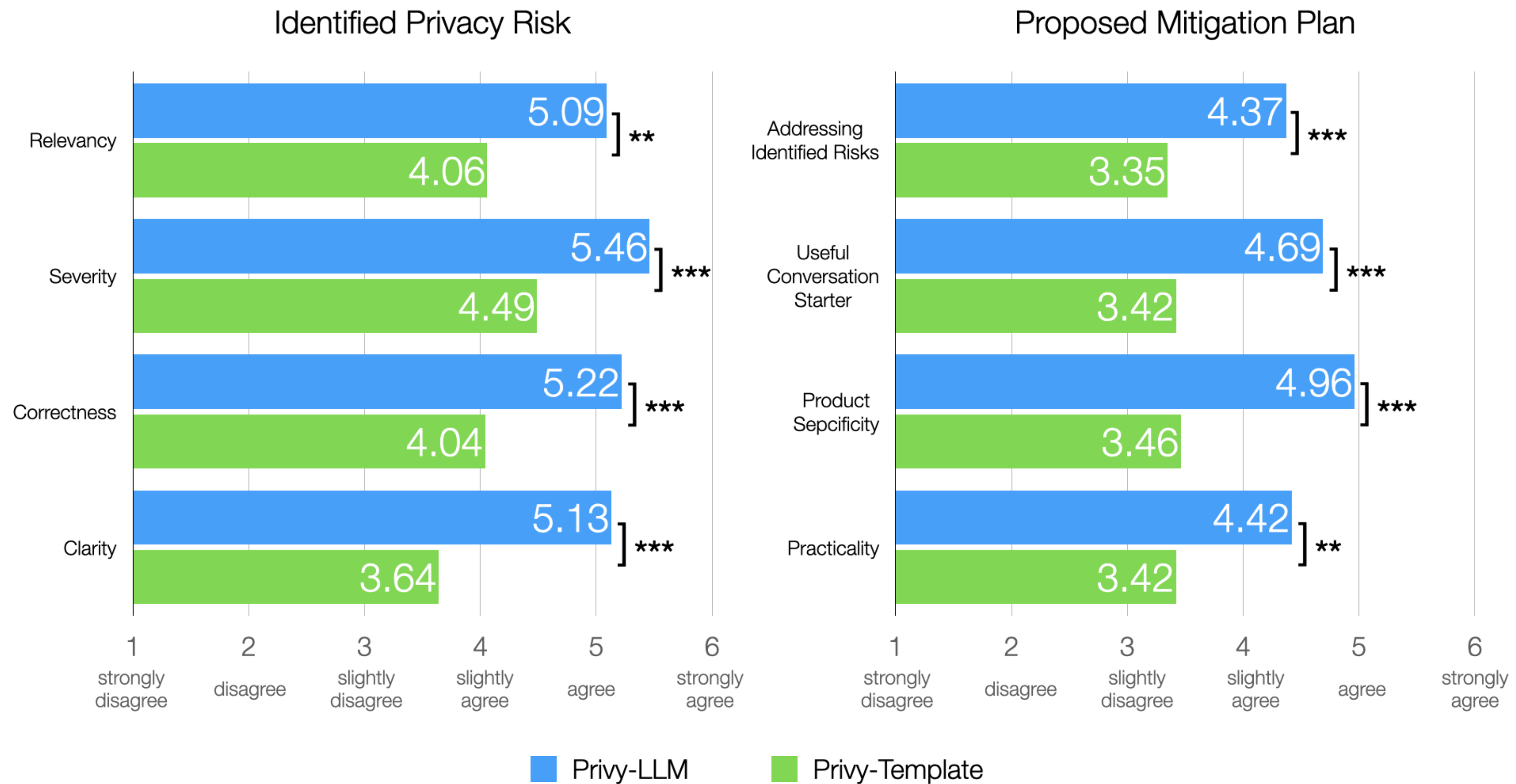
RQ1: How does Privy help practitioners identify and mitigate privacy risks with AI product concepts?

RQ2: How does the use of LLMs affect the quality of privacy impact assessments produced by Privy?

RQ3: To what extent does Privy address practitioners' challenges in privacy risk envisioning and mitigation?

Finding #3

The use of LLMs improves the quality of privacy impact assessments produced by Privy



Finding #5

Use patterns with human-AI collaboration in privacy risk envisioning and mitigation

Exploring the unknown unknowns:

“stuff like surveillance, risk intrusion, risk insecurity, I know when I see them... but it wouldn’t come up [on my own], but I know when I see them, what it is, and it’s actually risky.” (P7)

Solidifying and validating existing privacy knowledge:

“it kind of gives you at least a little bit more confidence or, yeah, like a sanity check.” (P3)

Integrating LLM-generated ideas:

“a user adds a minus marker, thinking it won’t share with this group, but doesn’t consider other parties and share it with them.” (P8)

RQ1: How does Privy help practitioners identify and mitigate privacy risks with AI product concepts?

RQ2: How does the use of LLMs affect the quality of privacy impact assessments produced by Privy?

RQ3: To what extent does Privy address practitioners' challenges in privacy risk envisioning and mitigation?

RQ3: To what extent does Privy address practitioners' challenges in privacy risk envisioning and mitigation?

Awareness

Motivation

Ability

Addressing awareness barriers: scaffolding structured privacy risk identification

“comparing [risks] was really fun... I started to realize, oh, maybe surveillance is not the big issue. Insecurity is the underlying cause.” (P17)

Addressing motivation barriers: fostering reflection and engagement

“it’s like the hacker mentality, right? You’re trying to break things... in a privacy context. You can just brainstorm and come up with a lot of these examples. You can spend like hours on this. I think it’s enjoyable to me personally.” (P20)

Addressing ability barriers: kickstarting self-efficacious privacy practice

“more confident that we have a solution here and we can move forward with a project idea... gives you more control over the problem.”(P24)

“kickoff meetings or an early conceptual phase of different ways we can use a particular AI tool.” (P8)

RQ3: To what extent does Privy address practitioners' challenges in privacy risk envisioning and mitigation?

Awareness

Motivation

Ability

Addressing awareness barriers: scaffolding structured privacy risk identification

“comparing [risks] was really fun... I started to realize, oh, maybe surveillance is not the big issue. Insecurity is the underlying cause.” (P17)

Addressing motivation barriers: fostering reflection and engagement

“it’s like the hacker mentality, right? You’re trying to break things... in a privacy context. You can just brainstorm and come up with a lot of these examples. You can spend like hours on this. I think it’s enjoyable to me personally.” (P20)

Addressing ability barriers: kickstarting self-efficacious privacy practice

“more confident that we have a solution here and we can move forward with a project idea... gives you more control over the problem.”(P24)

“kickoff meetings or an early concepting phase of different ways we can use a particular AI tool.” (P8)

RQ3: To what extent does Privy address practitioners' challenges in privacy risk envisioning and mitigation?

Awareness

Motivation

Ability

Addressing awareness barriers: scaffolding structured privacy risk identification

"comparing [risks] was really fun... I started to realize, oh, maybe surveillance is not the big issue. Insecurity is the underlying cause." (P17)

Addressing motivation barriers: fostering reflection and engagement

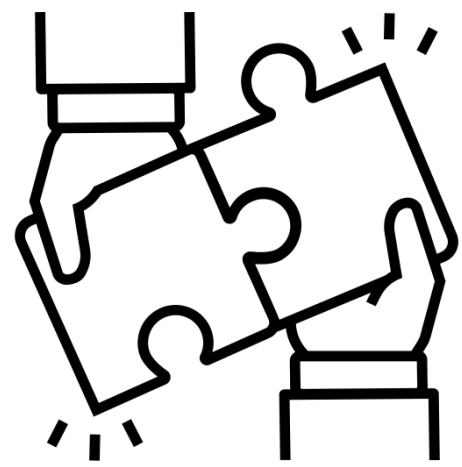
"it's like the hacker mentality, right? You're trying to break things... in a privacy context. You can just brainstorm and come up with a lot of these examples. You can spend like hours on this. I think it's enjoyable to me personally." (P20)

Addressing ability barriers: kickstarting self-efficacious privacy practice

"more confident that we have a solution here and we can move forward with a project idea... gives you more control over the problem."(P24)

"kickoff meetings or an early concepting phase of different ways we can use a particular AI tool." (P8)

AI-assisted privacy risk envisioning for privacy policy



Bridge principles →
product decisions

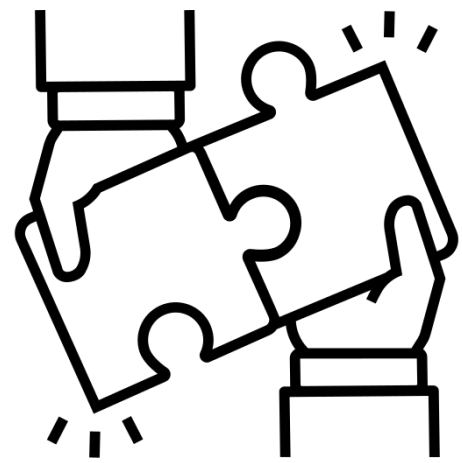


Reduce information
asymmetry in the AI
supply chain



From PIAs to policy
evidence

AI-assisted privacy risk envisioning for privacy policy



Bridge principles →
product decisions

- Help teams translate principles into design product decisions early.
- Guide teams to articulate data, risks, stakeholders, and mitigations.

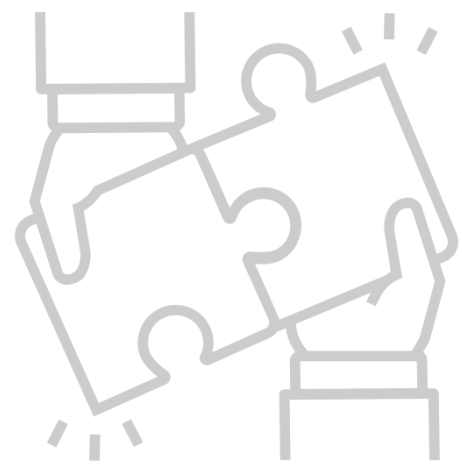


Reduce information
asymmetry in the AI
supply chain



From PIAs to policy
evidence

AI-assisted privacy risk envisioning for privacy policy



Bridge principles →
product decisions



Reduce information
asymmetry in the AI
supply chain

- Reduce informational asymmetries for non-privacy-experts.
- Make privacy claims more contestable and negotiable.



From PIAs to policy
evidence

AI-assisted privacy risk envisioning for privacy policy



Bridge principles →
product decisions



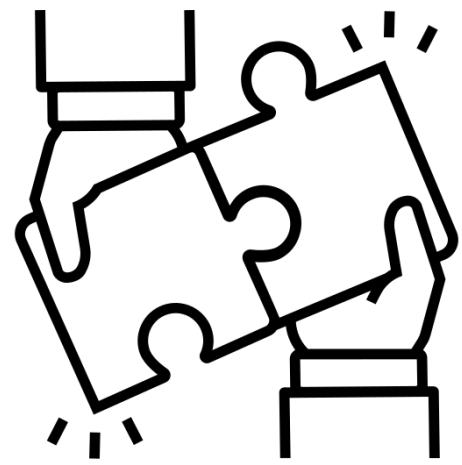
Reduce information
asymmetry in the AI
supply chain



From PIAs to policy
evidence

- Aggregate PIAs to reveal recurring risks and mitigation gaps.
- Inform targeted guidance and oversight from real-world patterns.

AI-assisted privacy risk envisioning for privacy policy



Bridge principles →
product decisions

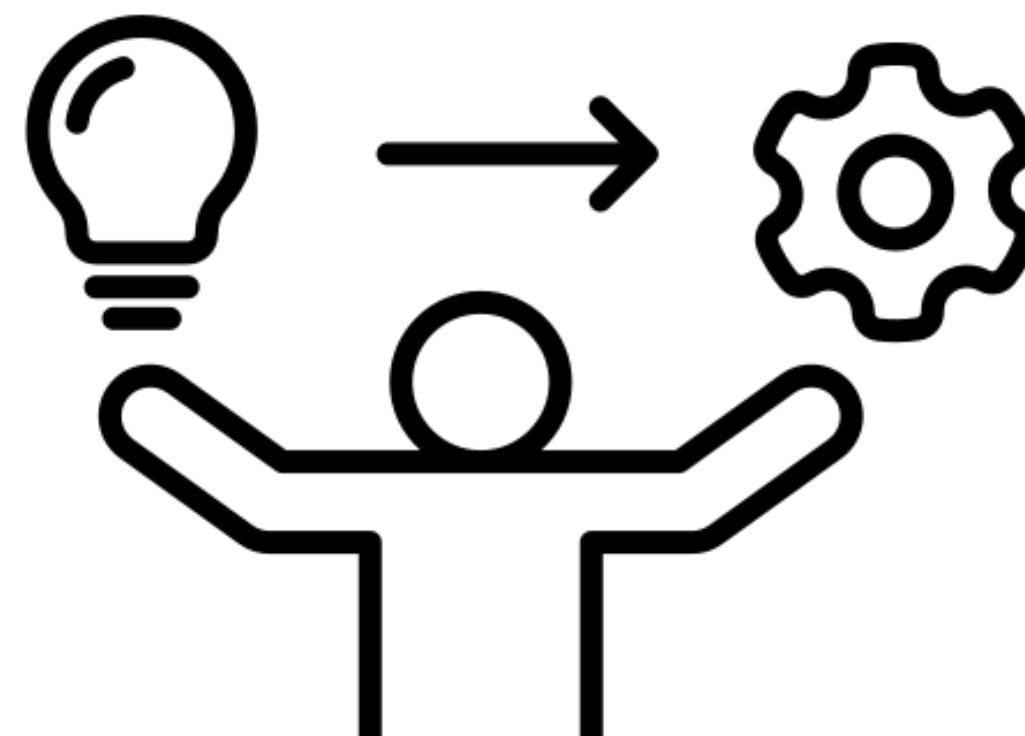


Reduce information
asymmetry in the AI
supply chain



From PIAs to policy
evidence

Product Concepts
Privacy Principles

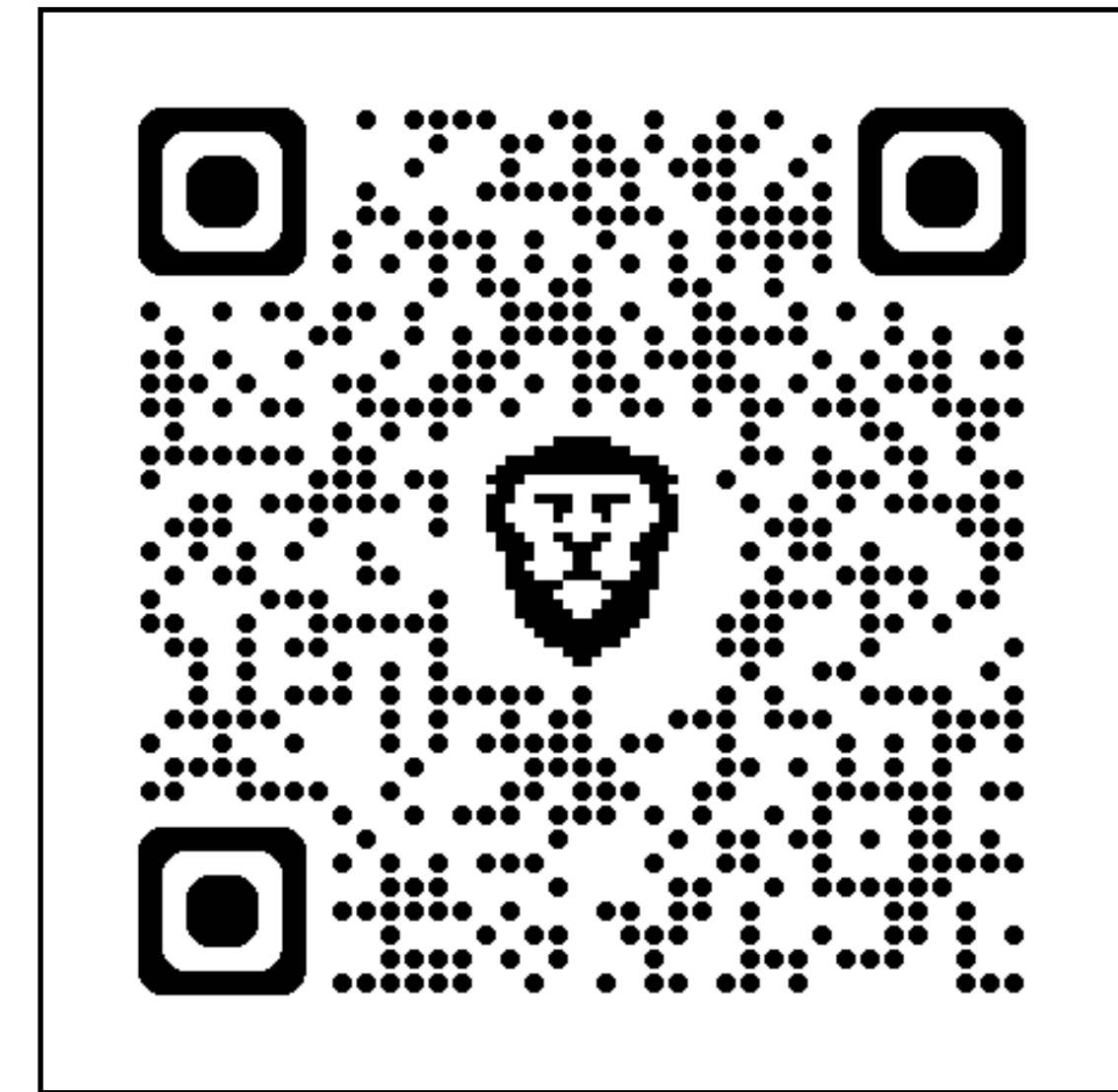


Practices

STEP 1: IDENTIFY STEP 2: RANK STEP 3: MITIGATE STEP 4: SUMMARIZE STEP 5: EXPORT

The interface is divided into several sections:

- AI Product Description:** Includes 'AI Meeting Assistant' and 'Product purpose' (an online meeting software feature that helps users summarize meeting notes, highlights, and key follow-up items).
- Use Cases:** Lists various user scenarios, such as 'A user can ask to recap the meeting topics, highlights, and content during or after the meeting'.
- AI Capabilities and Requirements:** Details what the AI product can do (e.g., generate concise meeting summaries) and what it requires (e.g., continuous collection of users' audio and video streams).
- Privacy Risks:** A central area where risks are identified and ranked. It shows a list of risks like 'Exposure Risk', 'Surveillance Risk', and 'Identification Risk' with their respective relevance and severity levels.
- Privacy Risk Definitions:** A sidebar providing definitions for various risk types:
 - Data Processing:** Data processing risks result from the use, storage, and manipulation of personal data.
 - Identification risk:** refers to linking specific data points to your identity.
 - Aggregation risk:** refers to combining various pieces of data about you to make inferences beyond what is explicitly captured in those data.
 - Phrenology risk:** refers to inferring personality, social, and emotional attributes about you from your physical attributes.
 - Secondary use risk:** refers to the use of your personal data collected for one purpose for a different purpose without end-user consent.
 - Exclusion risk:** refers to the failure to provide you with notice and control over how your data is being used.
 - Insecurity risk:** refers to the carelessness in protecting your personal data from leaks and improper access due to faulty data storage and data practices.



[👉 check out our CHI'26 paper](#)

PRIVY: a tool that helps practitioners envision and mitigate privacy risks for AI product concepts

PRIVY: Operationalizing Privacy Policy with AI-Assisted Privacy Impact Assessment Workflow



Hao-Ping (Hank) Lee

✉ haopingl@cs.cmu.edu

✂ @hankhplee

in hankhplee

Appendix

Concerns with human-AI collaboration in privacy risk envisioning and mitigation

Overestimating the level of automated assistance provided by Privy-LLM:

“I think the brainstorming mitigation tasks, I came up with them... I don’t think I took so much help from Privy.” (P7)

Risk of overreliance on its outputs:

“you may actually just be like, sure, this looks good, and accept.” (P9)