When and How AI Should Assist Brainstorming for AI Impact Assessment

JAROD GOVERS, University of Melbourne, Australia
SANJA ŠĆEPANOVIĆ, Nokia Bell Labs, UK and University of Oxford, UK
DANIELE QUERCIA, Nokia Bell Labs, UK and Politecnico di Torino, Italy

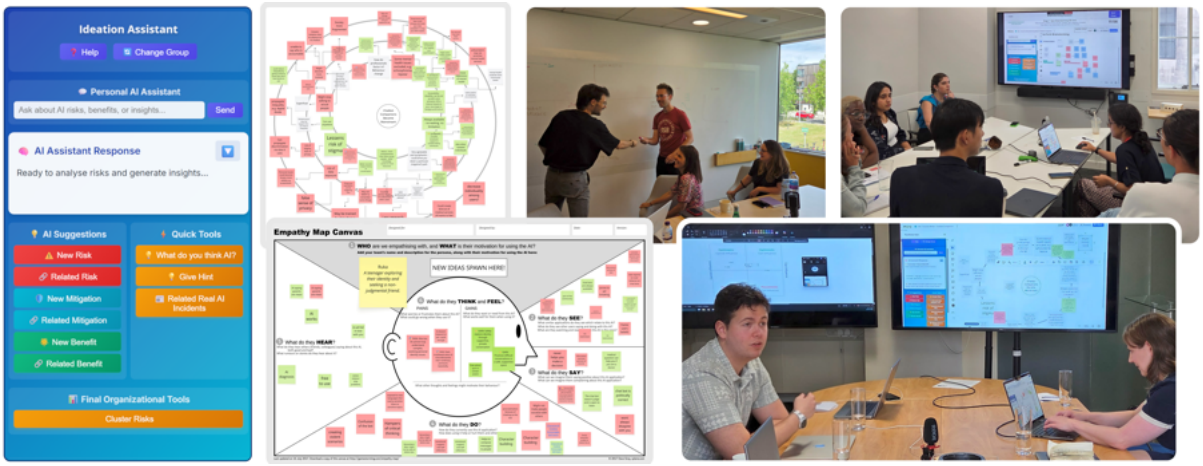

Fig. 1. Structured team brainstorming of AI impacts with AI interventions; consented photos from workshops in three sites.

A key task in AI practice is to assess potential impacts to prevent harm. Current AI tools assisting AI impact assessment have not been designed or evaluated for collaborative team brainstorming, and they do not capture the range of views in diverse teams. We studied how AI can support team brainstorming during AI impact assessment and made three contributions. First, we adapted two structured methods from strategic foresight and co-designed AI interventions for them in five in-person workshops with 28 participants in total. Second, we evaluated the interventions in ten in-person workshops with 54 participants, finding that AI improved impact assessment quality and brainstorming perceptions for a general-purpose AI use (a chatbot companion) but not for a specialised one (a kidney allocation application). Third, our findings result in broader design guidance for AI assistance in brainstorming: AI should only offer hints and not solutions during early ideation, initiating interaction only when participants face fixation or saturation; it should facilitate structuring ideas during convergence; leverage expertise to refine ideas; and overall, it should serve more in support of tedious brainstorming process tasks, rather than ideation that teams value to do themselves.

Authors' Contact Information: Jarod Govers, School of Computing and Information Systems, University of Melbourne, Melbourne, Australia, jgovers@student.unimelb.edu.au; Sanja Šćepanović, sanja.scepanovic@nokia-bell-labs.com, Nokia Bell Labs, Cambridge, UK and University of Oxford, Oxford, UK; Daniele Quercia, quercia@cantab.net, Nokia Bell Labs, Cambridge, UK and Politecnico di Torino, Turin, Italy.

FAccT '26, Montreal, Canada

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; *HCI theory, concepts and models*; *Collaborative interaction*; **Computer supported cooperative work**; Empirical studies in interaction design.

Additional Key Words and Phrases: Brainstorming, Creativity, Ideation, Artificial Intelligence, Risks, Impact Assessment

ACM Reference Format:
Jarod Govers, Sanja Šćepanović, and Daniele Quercia. 2026. When and How AI Should Assist Brainstorming for AI Impact Assessment. In *The 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '26), June, 2026, Montreal, Canada.* ACM, New York, NY, USA, 36 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

"Your AI application killed someone" is among the worst outcomes that have occurred when teams fail to consider the risks of Artificial Intelligence (AI) systems [64]. We define AI systems broadly as those that exhibit or simulate intelligent behaviour [79]. As AI spreads across society, comprehensive *impact assessment* has become both an ethical imperative and a regulatory requirement. Impact assessment is the systematic process of identifying AI's potential *benefits*, *risks*, and corresponding *mitigations* [8, 90]. Organisations and leading AI conferences increasingly recognise that responsible AI development requires proactive identification of potential impacts before development or deployment [4, 9], with the European Union's AI Act [19] mandating systematic risk assessment for AI deployments [30].

AI impact assessment process also aims to help development teams examine both immediate and long-term effects of their systems on individuals and society: before, during, and after development. Yet, predicting consequences from technical failures to human rights harms, is inherently difficult [10, 90], with high opportunity costs [44, 82] and moral stress [84]. Researchers have proposed various approaches to support AI practitioners, from automated risk generation [15, 44, 83] to visual exploration interfaces [9, 105]. A consistent finding across this work is that responsible AI planning requires a collaborative, team-based process incorporating diverse viewpoints and expertise [62]. Teams who deeply understand their application context and end-users possess crucial knowledge that individuals or automated tools alone cannot fully capture. Most approaches support individual practitioners [15, 44, 83], or enable them to jointly complete assessments [9, 105], but are not designed or evaluated to support brainstorming and deliberation.

Brainstorming is an activity where people meet in a group to generate many ideas for possible development [16]. Through structured deliberation on AI impacts, teams can uncover and develop targeted mitigations across technical, social, and policy dimensions before harm occurs [69]. Osborn identified three brainstorming stages [78]: ideation, convergence, and decision. Applied to AI impact assessment, these become: *ideate* (participants generate impacts), *converge* (identify, group, and organise impacts), and *decide* (develop mitigation strategies).

Commercial AI risk-assessment platforms have incorporated AI into their strategic foresight driven tools: Futures Platform uses the Futures Wheel method for structured brainstorming, while 4Strat's 'Foresight Strategy Cockpit' includes Futures Wheel and Empathy Mapping with AI risk generators (used by the European Commission) [1]. Microsoft's Azure AI Foundry also offers red-teaming and "AI Reports" that feed into its Responsible AI Impact Assessment Template and Guide [67]. However, these opaque AI typically provide suggestions without supporting team dynamics and have not been empirically evaluated to determine whether AI actually improves assessment quality. Whether AI is appropriate depends on design choices that preserve meaningful multidisciplinary, lived-experience human participation [89], and whether it can match or exceed human performance in identifying high-quality risks and mitigations that can save lives. Hence, we asked three research questions:

- (RQ_1) At which points in established brainstorming methods can AI meaningfully intervene to support teams in AI impact assessment?
- (RQ_2) What are the design requirements for AI interventions at these points?
- (RQ_3) What are the effects of AI interventions in team brainstorming on output quality and participant perceptions?

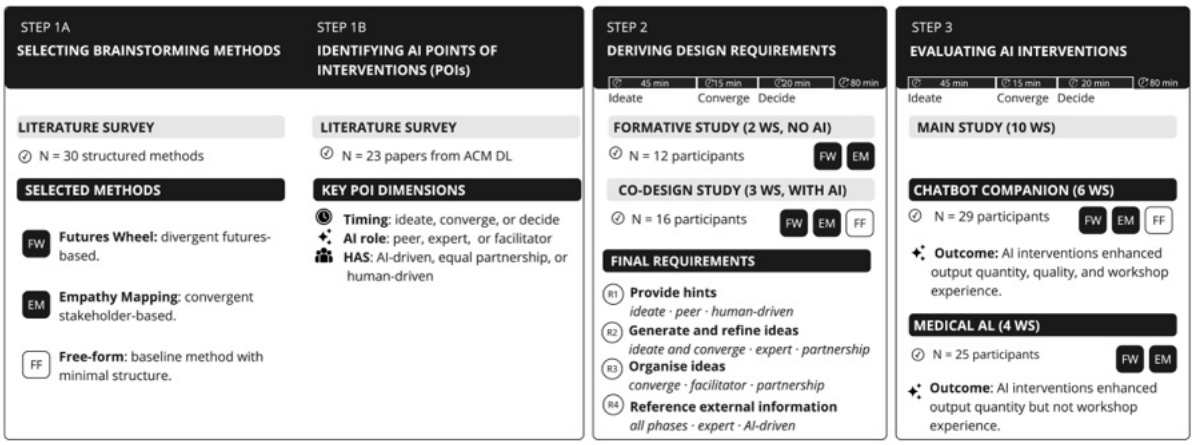

Fig. 2. Our methodology has three steps. In Step 1, we selected two brainstorming methods suited to AI impact assessment (§3.1, Step 1A), and identified types and points of AI intervention in these methods from prior work (§3.2, Step 1B). In Step 2, we elicited design requirements through five formative and co-design workshops (§4). In Step 3, we evaluated the resulting AI interventions across 10 workshops (§5). FW (Futures Wheel), EM (Empathy Mapping), FF (Free-form), HAS (Human Agency Scale), and WS (Workshop).

In answering these questions, we made three main contributions:

(1) **Scoping brainstorming methods and identifying viable AI intervention points within them** (§3). We reviewed traditional brainstorming methods, selecting two strategic foresight approaches used in AI impact assessment: Futures Wheels [38] and Empathy Mapping [41] (§3.1). We then reviewed literature on AI-assisted brainstorming and identified potential points of intervention along three dimensions: the stage of the brainstorming process (*when*), AI's roles (*what*), and human agency (*how*) (§3.2, Table 1).
(2) **Deriving design requirements for AI interventions at these points through co-design (§4).** Through five in-person co-design workshops with 28 participants, we identified design requirements for AI interventions in brainstorming (§4, Figure 6) and implemented a public AI tool to fulfil them: (https://social-dynamics.net/ai-risks/interventions/).
(3) **Evaluating the effects of AI interventions on output quality and participant experience** (§5). We ran 10 in-person workshops with 54 participants in total, comparing output quality (e.g., impact plausibility and uniqueness) and participant experience (e.g., sense of control or anxiety) in teams with and without AI assistance. Teams brainstormed impacts of a common AI use: a chatbot companion, and more specialised use: medical AI for kidney transplant allocation. AI interventions enhanced both output quality and participant experience (Table 4) for the chatbot companion. However, we found limited improvements in output quality for the medical AI use, suggesting that current AI capabilities may be limited where deep expertise matters. Participants also used AI assistance less during early ideation and more in later convergence and decision stages, and they preferred to initiate the interaction.

Our findings offer design guidance for AI-assisted brainstorming (§6, Figure 5). We conclude by examining AI's potential and limitations in team brainstorming for researchers, industry, and regulators, with future directions.

2 Related work

We review impact assessment ideation research, covering traditional (§2.1) and AI-assisted approaches (§2.2).

2.1 Traditional brainstorming and ideation for AI impact assessment

Early-stage brainstorming to understand AI impacts can reduce harmful incidents and provide essential context for development teams [106]. Despite these benefits, few studies have examined how structured brainstorming could aid AI impact assessment. Ballard et al. [6] developed *Judgement Call*, a card game using simulated user reviews of positive and negative experiences to explore why users might respond to an AI application favourably or unfavourably (e.g., giving a one-star review due to privacy issues). Hohendanner et al. [48] combined foresight-driven studies and technology assessment methods in participatory workshops titled *Global AI Dialogues* to gather laypeople's perspectives on generative AI. Primlani et al. [81] explored *Design Courts* as an approach to elicit ethical and legal dilemmas for AI. These approaches helped teams ideate ethical dilemmas and identify risks and benefits to stakeholders, but did not address the full ecosystem of impacts and mitigations required for full AI impact assessments [8, 30, 44].

Elsayed-Ali et al. [29] designed *Responsible & Inclusive Cards* with guiding questions to probe thinking regarding impacts to stakeholders, outcomes and governance/practice. They suggest that future research should empirically test how well structured approaches to impact assessment work in practice. Using structured methods can maximise the performance of human-led teams and enables a fair evaluation of whether AI augmentation improves brainstorming quality, compared to the bottlenecks often experienced from 'blank canvas' brainstorming [25]. This in turn increases the need to empirically evaluate whether AI can improve the quality of human brainstorming, or whether structured brainstorming alone would suffice. This is particularly pertinent given the rise of AI which can augment human-driven impact assessments [44, 83], more of which we discuss in the following section. Structured methods could also help guide AI through brainstorming tasks more systematically to improve their breadth and relevance through personas or systemic futures. This is particularly important given the well-documented challenge of human and AI design fixation in ideation [51].

2.2 AI-assisted brainstorming and ideation for AI impact assessment

We conducted a scoping literature review of AI tools for impact assessment following the PRISMA approach [80] (full search strings in Appendix §A). Across eight identified studies [5, 9, 15, 17, 28, 44, 83, 105], we found that all tools assist individual practitioners but overlook team collaboration dynamics. These tools differ in their approach. Wang et al. [105] explored how Farsight, an LLM-powered interactive tool, helps prompt prototypers explore impacted stakeholders and risks through a visual tree, but without empirical evaluation of team collaborative effectiveness. Other tools help ideate risks and benefits through taxonomies [5, 28] or through automatic generation upon AI technology or model input, such as AHA! [15], ExploreGen [44], and RiskRAG [83]. None of these tools were designed for collaborative ideation. The AI Design framework pre-fills impact assessment reports [9] and can be used by different team members in parallel, but again does not focus on team deliberation and collaboration. Related work in red teaming, such as PyRIT [74], targets individual professionals. Moreover, a recent review found that red teaming practices are generally hidden by non-disclosure agreements [109], leading researchers to call them "security theatre" [31].

Practitioner accounts and tool-oriented resources show that teams already use LLMs to seed, pre-fill, or facilitate impact assessment, but they rarely contain systematic empirical evaluation. Microsoft's Responsible AI materials provide templates and guidance for internal assessments and mention LLM-assisted activities in applied reviews [67]. NIST's AI RMF offers a widely adopted risk-management scaffold that practitioners use to structure workshop outputs and mitigation tracking [75]. MITRE's red-teaming guidance describes adversarial brainstorming techniques that organisations can combine with AI models to surface misuse scenarios [69]. Together, these practitioner-oriented resources establish feasibility and demand for AI to assist, but not replace brainstorming, and leave open whether AI-assisted brainstorming measurably improves the quality, or participant perceptions of formal AI impact assessments.

Candello et al. [17] proposed 'mitigator' LLMs that work alongside a base LLM to detect and revise problematic outputs before reaching users, and used stakeholder personas in workshops to co-design improved responses. This approach demonstrates utility of AI for live mitigations, though its applicability to the full benefit, risk, and mitigation ecosystem for a full impact assessment necessitates empirical evaluation to identify if AI is needed, or if human-only approaches is sufficient.

Structured brainstorming for corporate risk assessment, known as strategic foresight, exists in platforms like 4Strat's 'Foresight Strategy Cockpit' [1] for AI-augmented risk assessment with personas. Whereas, Futures Platform uses AI and the Futures Wheel method to aggregate, visualise, and surface unexpected risks [34, 36]. Despite these tools, there is a lack of empirical evaluation to justify the use of AI, with Futures Platform itself noting the need to evaluate generative AI [35].

Research gap. Current AI tools for impact assessment have not been designed and evaluated (against a human-baseline) for collaborative team brainstorming. This gap is significant because teams require active involvement in assessment to build understanding and foster responsibility, rather than relying on individual or automated approaches [61, 104]. A key question remains: what are the appropriate points and types of AI intervention in brainstorming AI impacts?

3 Identifying AI intervention points in brainstorming methods for AI impact assessment

To answer RQ_1: *At which points in established brainstorming methods can AI meaningfully intervene to support teams in AI impact assessment?*, we took two steps (Figure 2). First, we selected two brainstorming methods suitable for impact assessment (§3.1). Second, we identified points of AI interventions from prior work (§3.2).

3.1 Brainstorming method selection as used by risk assessment platforms

We focused on structured brainstorming methods for two main reasons. First, they are already in use for corporate risk assessments [1, 34, 67], but have not been empirically tested in team settings with AI support. Second, traditional free-form brainstorming often suffers from psychological barriers: production blocking, where participants must wait for others to speak, and evaluation apprehension, which can suppress creativity and discourage contribution [25]. Structured approaches were developed to address these issues to foster more inclusive and productive participation [72]. Hence, we identified 30 structured brainstorming methods with potential to fulfil the requirements for AI impact assessment (Appendix Table 7) from three sources: (1) the Millennium Project, a peer-reviewed collection of Strategic Foresight methods [38] used in risk assessment platforms [1, 34, 99], (2) the United Nations Futures Lab, used in AI policy making including the Futures Wheel approach [60, 103], and (3) the UK government, using Strategic Foresight methods for AI management [47, 99, 100]. From the 30 methods, we excluded those that failed any of the four criteria:

(1) Designed to consider future implications, making them suitable for ideating AI impacts;
(2) Designed for broad rather than narrow, specific contexts (e.g., operational management, supply chains);
(3) Applicable to products such as an AI application rather than processes only;
(4) Methods that focus on divergent *or* convergent thinking.

After applying this criteria, the remaining methods were Backcasting, Delphi, Visioning, Futures Wheel, SWOT, and Empathy Mapping. We selected two complementary ones: Futures Wheel, a *divergent*, futures-oriented approach that explores broad societal implications across near- to long-term horizons; and Empathy Mapping, a *convergent*, stakeholder-focused approach that captures end users' perspectives. Investigating stakeholders/personas is used in the 4Strat platform [1], by Candello et al. [17] for real-time mitigations, and Hohendanner et al. [48] in non-AI assisted settings only. We selected these methods for their use in corporate tools [1, 34, 67] and research [48], making them ideal for a first empirical evaluation of AI-assisted versus human-only assessment.

Table 1. AI interventions in brainstorming from literature, organised by stage, AI role, and human agency.

	Ideate			Converge			Decide		
Human Agency	*Peer*	*Facilitator*	*Expert*	*Peer*	*Facilitator*	*Expert*	*Peer*	*Facilitator*	*Expert*
AI-driven	[27, 46, 56, 66, 73, 76, 108]	—	[46]	—	[56, 86, 110]	[76, 86]	—	—	—
Equal partnership	[3, 11, 43, 55, 93, 107, 110]	[107, 110]	—	—	[94, 110]	[93]	—	[110]	—
Human-driven	[33]	[33]	—	—	[33]	[50]	—	[33]	[50]

Note: Dashes indicate that no studies were found.

Futures Wheel [38]. A structured foresight method for divergent thinking that explores potential outcomes across chains of consequences. Participants place a central trend in the middle and ask: "If this trend occurs, what happens next?" They add first-order impacts if all agree the impact is plausible, repeating this process across three rings. We adapted it for impact assessment by having participants consider the AI use becoming mainstream as the central trend.

Empathy Mapping [41]. A method that considers a specific stakeholder and has participants fill out what this stakeholder sees, says, does, hears, thinks, and feels. We asked teams to select two AI-use stakeholders—one relatable and one not—from a template varying in backgrounds, gender, age, and jobs (Appendix Figure 10), or they could create their own.

Free-form brainstorming. Our control condition uses an unstructured blank canvas where participants generate potential AI use impacts without predefined structure.

For all three methods, we applied a consistent three-stage process: participants first generate ideas of impacts (*ideate*), then colour-code them as positive (benefits), negative (risks), or neutral, create any additional related benefits and risks and group them (*converge*), and finally propose mitigations for identified risks (*decide*).

3.2 Identifying types and points of AI intervention from prior work

Using a string-based search for AI-assisted brainstorming *in general* (Appendix A.1), we identified three dimensions along which interventions in brainstorming should be conceived (Table 1).

Timing: *when* to intervene. The mixed-initiative HCI model [49] proposes that AI should decide when to act or interrupt users by weighing costs, benefits, and uncertainties in user attention and system utility. In brainstorming, benefits include new ideas or guided discussion; costs include breaking flow or annoying participants.

AI roles: *what* should be the function of AI. Bittner et al. [7] identified three AI roles in collaborative work. As a *peer (ideator)*, AI adds ideas like a team member. As a *facilitator*, it organises ideas and guides teams. As an *expert*, it gives specialist information. The trade-offs concern participant needs and AI capabilities. Some roles may suit certain brainstorming stages: AI as a peer may fit ideation, while the expert role may better fit the converge and decide stages.

Human agency scale: *how* is the interaction initiated. This scale measures preferred collaboration levels, ranging from AI-driven to equal partnership to human-driven, depending on how much collaboration participants need to complete tasks effectively with AI. There is a trade-off between user agency and results [2, 70]. In brainstorming, AI might automatically intervene with ideas that participants can only accept or reject (*AI-driven*), participants might initiate interaction or revise AI suggestions (*equal partnership*), or participants always initiate and direct what ideas AI generates (*human-driven*). More control means more work for participants but may improve outcomes [2], and participants may prefer to initiate AI interactions rather than to receive proactive recommendations [110].

Our summary in Table 1 aligns with a recent systematic review on AI-assisted ideation more broadly [59]. Most HCI and CSCW studies fall into two groups: (1) AI as a *peer* or idea generator, used almost entirely during ideation with minimal to moderate participant agency, and (2) AI as a *facilitator* or *expert*, used across stages but most often during convergence, again with minimal to moderate agency. AI peers dominate ideation research yet are absent in later stages. The roles of experts and facilitators remain less explored throughout.

4 Deriving design requirements for AI interventions through co-design workshops

To answer (RQ_2): *What are the design requirements for AI interventions at the identified points?*, we conducted a **formative study** to elicit initial design requirements, followed by a **co-design study** to iterate and refine them (Figure 2, Step 1C).

Selection criteria for the AI uses. We selected two orthogonal AI use cases to test the generalisability of our approach. Orthogonal here means the cases differ along two independent axes: *modality* (physical embodied *vs.* digital AI), to represent the virtual and physical harms of AI, and the EU AI Act *risk level* [19]. We focus on the *high-risk*, *limited-risk*, and *low-risk* categories, excluding the illegal *prohibited* category. We selected a digital chatbot companion as the limited-risk case [111](Cat. #456), and a physical AI-assisted kidney monitoring and allocation system as the high-risk case [111](Cat. #79). This pairing covers the divergences in the EU AI Act's risk categorisation for the uses that require an impact assessment (low-risk uses do not), while being feasible for 15 workshops. Both involve social, ethical, systemic, technical, and economic dimensions, to explore a wide range of potential impacts.

Study setup. In both, formative and co-design study, participants brainstormed the same AI use: a chatbot companion for emotional support. This choice reflects recent developments such as Replika [87, 95], character.ai [18], virtual influencers [92], and xAI's companion-bot Ani [23] with which most of our participants experienced with AI would be familiar. We used the Miro platform to support the brainstorming process [68], using sticky notes on a canvas and a side-toolbar for our AI tool (Figure 3). Participants first completed a Miro tutorial, and then selected a participant to be the coordinator to take notes on Miro and use the AI tool on the team's request. The team then received a description of their AI use case: the "*conversational agent, Salieri, which uses text-to-speech and language processing to provide companionship and emotional support*" (Appendix Figure 7). Teams were instructed to adopt the perspective of developers working for the AI system's company. They were asked to identify risks that would be realistic for a developer team to address and to describe them in enough detail that another developer could understand each impact independently and implement a clear mitigation. For example, "mental health" represents a broad risk for a chatbot but is too vague for AI developers, whereas "sycophantic relationships with the companion" is sufficiently specific for actionable mitigations (e.g., implementing frictions, reflections etc.). Participants were also shown the marking criteria (Table 2) to understand how their impacts would be evaluated. Each workshop lasted 1 hour 45 minutes.

Participants. We recruited 4–7 participants per workshop, as fewer than 4 can limit divergent thinking [42], while more than 7 leads to production blocking, where dominant voices create bottlenecks [25, 42, 45]. Participants could join only one workshop, so each session had new participants. We recruited from universities and companies via mailing lists and posters across Cambridge, Oxford and London. We did not require degrees, but all had at least a bachelor's, though not necessarily in computer science (e.g., business, English). All workshops were balanced to include private sector workers, researchers, and students, and to have at least 2 computer science backgrounds to prevent one-sided groups while allowing non-expert [22] and multidisciplinary voices [53, 77]: 37% were from the private sector, 43% researchers, and 20% students. Workshops were held at three in-person locations to balance our 82-person total sample. The mean age was 31 (SD 6.5, range 22–55), close to the UK IT workforce's mean of 35 [91]. We recruited individuals rather than existing teams to avoid prior relationship

biases, and we discuss in *Limitations* the need to explore setups with existing team dynamics for future work. Participants registered for workshops, were assigned to groups, signed consent forms, and received £20 each. All workshops took place in June–July 2025.

Data collection. We collected post-workshop feedback on structure, timing, and design requirements:

(1) If you were to design a digital tool to help teams like yours brainstorm about the benefits, risks, and mitigations when developing an AI use case, what features would be most important based on your experience today?
(2) What role would you want an AI to have in your brainstorming team, and when would you use it?

Analysis. We audio-recorded and transcribed interactions during in-person sessions and conducted a two-coder inductive thematic analysis of the discussions and responses to identify design requirements [13]. This involved two-coders identifying themes as they emerged, taking notes during the workshops, and creating a set of coded themes, then revising the themes based on the collected post-study qualitative written questions. Annotators met across three rounds to compare and finalise themes. We used this qualitative approach throughout the formative/codesign workshops.

Results: formative study. Three initial requirements emerged, each aligned with an AI role we identified from prior work. First, participants wanted AI to act as a an additional team member (*ideator*) and generate ideas: *"I think having an AI as an extra member to contribute new ideas would be cool."* (P3, EM). We built an agent that generated ideas and shared them at intervals of 5–7 minutes. Second, participants asked for AI to act as an experienced member (*expert*): *"[...] use that information to generate additional ideas or finer-grained nuances."* (P7, FW). We built an agent that generated ideas refining those already on the board. Third, participants wanted AI to support the process (*facilitator*) by grouping ideas: *"I would actually want it to group together the benefits/risks initially generated by humans."* (P2, EM). We built an agent that clustered ideas into semantic groups. These initial requirements led to our preliminary AI tool.

Implementation of the preliminary AI tool. We developed the AI tool as a sidebar for Miro users using the Miro API, with PHP and JavaScript connecting to the OpenAI API to provide a chat interface powered by prompt-tuned GPT-4.1. We used GPT-4.1 as the state-of-the-art performance model at the time to avoid any performance confounds, as well as for its support for web-based search, used for extracting data from the AI Incident Database [97]. Participants could generate new and related benefits, risks, and mitigations. We used the six-category DeepMind risk taxonomy [106] to ensure coverage of categories not already addressed. The tool could place sticky notes, semantically cluster impacts, and suggest additional ones (Appendix Figure 6). We provide all prompts in Appendix E.

Results: co-design study. We evaluated the preliminary tool to understand when participants would use each AI role (when) and how much agency they desired (how). Participants found our AI peer—which autonomously shared ideas—intrusive: *"the group broadly agreed that the tool's automated pop-ups felt intrusive [...] the length of the text strings disrupted discussion while people read,"* (P27, FW, Initial AI) and *"patronising... [and assumed teams] couldn't freely think up ideas."* (P17, EM, Initial AI). They preferred AI to respond only when invited and to provide hints rather than complete ideas, at least during early ideation: *"To derive hints and suggestions based on the use case, encouraging users to engage in brainstorming and critical questioning."* (P11, EM, Initial AI). Participants valued the AI's expertise during later ideation and early convergence: *"everything generated was, at the very least, correct, if not new. It allowed us to get a new idea when we were somewhat stuck."* (P13, FW, Initial AI). They also requested external references (e.g., real-world incidents for proposed risks) throughout all stages. Participants were satisfied with the clustering feature. Finally, they asked for a simple interface with minimal distractions that adapts to each brainstorming phase.

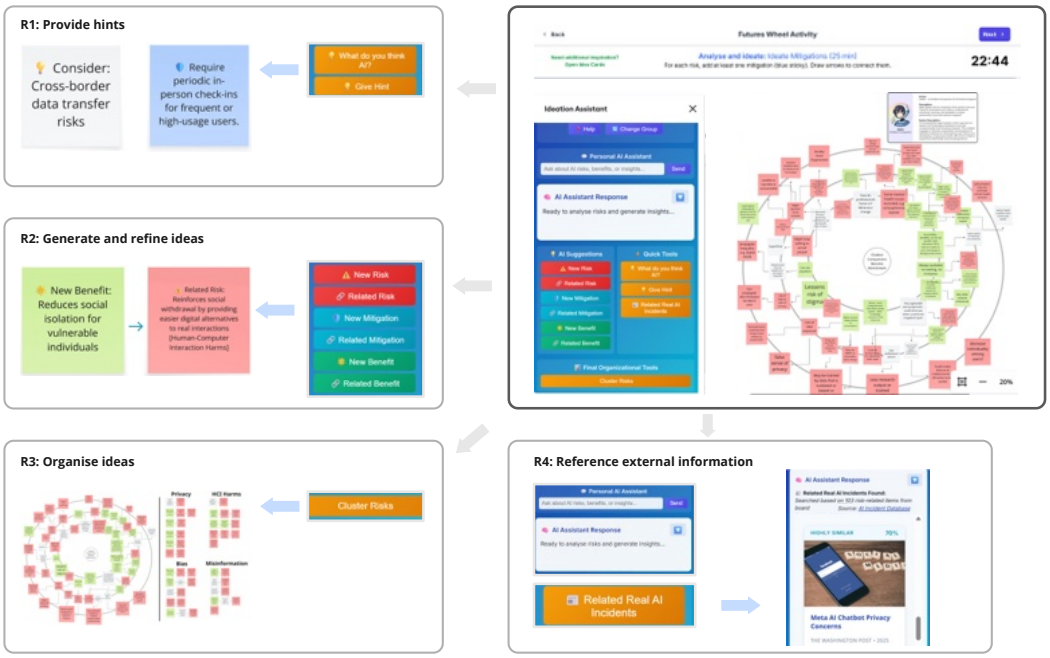

Fig. 3. An example from a FW workshop with the Final AI tool and its elements fulfilling the four design requirements.

Final design requirements. We answer RQ_1 with 4 requirements, mapped to phases, roles, and agency levels:

R1 **Provide hints**: ideate stage, peer role, human-driven.
R2 **Generate and refine ideas**: ideate and converge stages, expert role, equal partnership.
R3 **Organise ideas**: converge stage, facilitator role, human-driven.
R4 **Reference external information**: all stages, expert role, equal partnership.

And two requirements on *design*:

R5 **Simple interface**: minimise distractions.
R6 **Adaptable interface**: show only relevant functionality for each stage and method.

Implementation of the final AI tool. These requirements shaped our final AI tool, which was more simplified, adaptable to the brainstorming stage, and designed to offer hints on demand rather than generate ideas automatically. The tool also allowed users to surface relevant real-world incidents of AI use-caused harm (Figure 3). We tuned prompts to limit impacts to 10–15 words after feedback indicated outputs were too long compared to participant entries. To prevent redundancy, we also fed existing Miro board content into the prompts (Appendix E).

5 Evaluating AI interventions for AI impact assessment brainstorming

To answer RQ_3: *What are the effects of AI interventions in team brainstorming on output quality and participant perceptions?*, we compared teams with and without AI across 10 workshops (Figure 2, Step 3).

We ran two sets of workshops. The first set used a 2×3 design: two conditions (with and without AI) and three brainstorming methods (Futures Wheel [FW], Empathy Mapping [EM], and Free-Form [FF] baseline, Figure 4). Participants brainstormed the impacts of an AI chatbot companion. The results showed a clear advantage for structured methods over the baseline, so in the second set we focused only on the two structured methods.

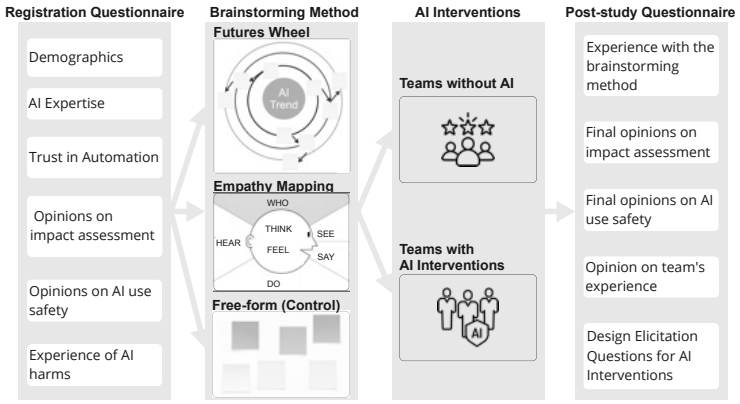

Fig. 4. Evaluation study workflow. Registration and pre-study questionnaires collect demographics and initial perceptions. Participants then complete the workshop using one of three methods, with or without AI, and then a final questionnaire.

Table 2. Output quality annotation questions. All metrics apply to benefits, risks, and mitigations unless noted.

Metric	Scale	Question	Source
Plausibility	1–5	Is it plausible that this {benefit/risk/mitigation} could arise from (or be implemented for) the described AI use?	[24]
Probability	1–7	How likely is this {benefit to be realised / risk to occur / mitigation to be implemented}?	[75]
Uniqueness	1–3	Is this {benefit/risk/mitigation} unique to this specific AI use?	[53]
Novelty	1–3	Rate the creativity: mundane to imaginative.	[21]
Engagement	1–5	I find this {benefit/risk/mitigation} easy to engage with or apply.	[14, 58]
Type-specific metrics:			
Magnitude (*Benefits*)	1–5	If fully realised, what would be the magnitude of the positive impact?	
Severity (*Risks*)	1–5	How severe are the impacts of this risk?	[75]
Effectiveness (*Mitigations*)	1–5	How likely is this mitigation to effectively address its associated risk?	

5.1 Metrics

Output quality. We had 5 expert annotators from a 90 participant pool rate each impact (benefit, risk, and mitigation) from each of the 15 workshops on 8 Likert scales (Table 2), attaining an inter-rater agreement of 0.69–0.85 (Kendall's W; substantial agreement [52]). We recruited our 90 expert pool on Prolific, screening for those who used AI at least 3 times weekly and worked in legal or AI engineering sectors for competency in impact assessment. We included attention checks (specific responses) and comprehension checks (identifying unrelated impacts), and replaced failed candidates.

Participant perceptions. Before and after each workshop, we measured participants' sense of control, anxiety, confidence in risk assessment, views on oversight, and comfort recommending the AI use (Table 3).

5.2 Analysis

Quantitative analysis. We used Cumulative Link Models (CLMs) [63] to analyse both output quality and participant perceptions, as appropriate for ordinal Likert-scale data. For output quality, inter-rater reliability among the five expert coders was strong (overall Kendall's W = 0.71) [52]. We modelled each quality metric with brainstorming method and AI condition as predictors. For participant perceptions, we modelled post-workshop

Table 3. Pre/post-workshop perception questions: "If you were responsible for assessing, designing, or developing this AI..."

Metric	Scale	Question
Control	1–5	How much control do you feel you would have?
Anxiety	1–5	How anxious or uneasy would you feel?
Confidence	1–5	How confident would you feel in carrying out a risk impact assessment?
Oversight	1–5	How likely is it that this AI system would still require regular updates and oversight, even if perfectly designed and tested?
Recommendation	1–5	How comfortable would you feel using it for yourself or recommending it to a close friend or family member?

Table 4. Results by AI condition for *chatbot companion* and *medical AI* uses. Values are mean ± SE from CLM models. **Blue**/**orange** indicate significantly higher/lower values (Tukey-adjusted). Significance: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

(A) Output Quality

		Plausibility (1–5)	**Probability** (1–7)	**Uniqueness** (1–3)	**Novelty** (1–3)	**Impact**† (1–5)	**Engagement** (1–5)	**Count**
Benefits								
Chatbot	AI	**4.08**** ±.15	**5.49***** ±.17	1.99 ±.18	**2.02*** ±.16	**3.64***** ±.16	**4.05***** ±.16	**68**
	No AI	**3.82***** ±.07	**4.98***** ±.11	1.88 ±.06	**1.90*** ±.05	**3.30***** ±.07	**3.59***** ±.06	29
Medical	AI	3.66 ±.21	4.99 ±.21	**1.90*** ±.23	1.93 ±.22	3.59 ±.21	3.75 ±.21	**33**
	No AI	3.73 ±.09	5.11 ±.11	2.09 ±.06	2.01 ±.06	3.61 ±.08	3.77 ±.08	27
Risks								
Chatbot	AI	**3.86***** ±.10	**4.83***** ±.10	1.89 ±.11	1.83 ±.11	**3.56***** ±.10	**3.76***** ±.10	**96**
	No AI	**3.56***** ±.05	**4.40***** ±.06	1.87 ±.03	1.80 ±.03	**3.30***** ±.04	**3.55***** ±.04	80
Medical	AI	3.70 ±.17	4.88 ±.17	**1.96**** ±.18	1.80 ±.18	3.51 ±.17	3.72 ±.18	**46**
	No AI	3.65 ±.09	4.99 ±.08	1.75 ±.05	1.70 ±.05	3.53 ±.07	3.69 ±.06	42
Mitigations								
Chatbot	AI	**3.94***** ±.11	**5.06***** ±.13	1.86 ±.11	1.86 ±.11	**3.67***** ±.11	**3.78***** ±.11	**78**
	No AI	**3.31***** ±.05	**4.55***** ±.08	1.87 ±.03	1.83 ±.03	**3.40***** ±.05	**3.54***** ±.05	54
Medical	AI	3.70 ±.19	4.79 ±.19	1.89 ±.20	1.79 ±.20	**3.85*** ±.19	3.76 ±.19	**59**
	No AI	3.56 ±.10	4.94 ±.11	1.80 ±.06	1.68 ±.05	**3.68*** ±.07	3.69 ±.07	29

†Impact = Magnitude (benefits), Severity (risks), or Effectiveness (mitigations).

(B) Participant Perceptions (all scales 1–5)

		Control	**Anxiety**	**Confidence**	**Oversight**	**Recommendation**
Chatbot	AI	**2.89*** ±.19	**1.77*** ±.13	**3.80*** ±.21	2.60 ±.13	**2.74*** ±.17
	No AI	**2.04*** ±.13	**2.47*** ±.27	**3.03*** ±.30	2.52 ±.24	**1.97*** ±.30
Medical	AI	1.99 ±.21	2.71 ±.24	1.77 ±.26	2.35 ±.32	2.58 ±.23
	No AI	2.36 ±.21	2.81 ±.26	2.00 ±.33	2.51 ±.27	2.63 ±.25

responses while controlling for pre-workshop scores, AI expertise, education, and self-rated ethical alignment. We conducted estimated marginal means (`emmeans`) analysis with Tukey adjustment, back-transforming results to Likert values to represent the shifts between AI/No-AI conditions. Variance Inflation Factors were below 5, confirming no significant predictor multicollinearity [85].

Qualitative analysis. Two coders conducted inductive thematic analysis on participants' written responses using the aforementioned approach [13], supplemented by notes and audio transcripts from the workshops. We triangulate our findings with the first two targeted around quantitative results, and findings 3–5 focusing on the dominant themes that emerged from our qualitative insights (i.e., activation at design saturation/fixation, scaffolding *vs.* directing, intrusiveness *vs.* proactivity, agency accruing over time). Both coders reviewed the data across three rounds to meet and resolve disagreements.

5.3 Results

Based on the quantitative results, as supplemented with qualitative feedback, we answer RQ_3 with five main findings: two addressing the quality of the impacts generated and the participants' perceptions, and three on the points for AI intervention.

Finding 1: AI interventions improved the output quality for the general AI use but not for the specialised one. For the chatbot companion, AI intervention teams generated 48% more impacts than human-only teams (Table 4A), scoring significantly higher on six of the eight quality metrics but not on the two creativity measures (uniqueness and novelty). For the medical AI, the AI intervention helped yield 22% more impacts, with increases in unique risks and effective mitigations. Thus, AI interventions can help teams identify more risks and mitigations early in the design phase with the quality being at least equal to the teams without AI (for the specialised medical AI) or higher quality (for the general chatbot companion). However, AI impacts were equally creative to human-made ones. We thematically observed that human-generated risks were more systemic and longitudinal, than those from the AI—which predominantly targeted model capability risks.

Finding 2: AI interventions improved participant perceptions for the general AI use case but not for the specialised one. Participants experienced reduced anxiety, increased confidence, and a greater sense of control. As Table 4B shows, for the chatbot companion, AI intervention teams reported significantly less anxiety compared to teams without AI. P20 (EM, Final AI) noted: *"I was not seeing these risks as unmanageable or insurmountable."* Participants also reported a greater sense of control: *"I can influence a bit more on the development of these kind of AI chatbots"* (P24, FF, Final AI). They expressed increased confidence in risk assessment: *"I was more attuned to the benefits of such a system, whilst also seeing risks that I was otherwise unaware of"* (P20, EM, Final AI), and higher likelihood of recommending the AI to others: *"I became more confident with the system design approach which highlights that engineers and developers are trying hard to mitigate potential risks"* (P59, EM, Final AI). For the medical AI, qualitative feedback indicated that some participants in these teams also experienced reduced anxiety and increased confidence when using the AI tool.

Finding 3 (when): Teams mostly used AI when facing idea saturation or nearing time limits. Participants identified idea saturation as a key challenge, particularly for the medical AI groups, as evidenced by the significantly lower number of impacts generated in these teams. We observed how participants used the AI tool to break through idea saturation and design fixation. As P68 (EM, Final AI) noted, it *"was helpful to come up with novel items whenever we as a group got stuck."* The other common use was when time ran low: *"It was helpful in filling in the blanks if time was getting short and seeing if we'd missed things"* (P45, FW, Final AI). However, timing mattered: when used under pressure, team members became *"less critical of what the tool said"* (P40, FW, Final AI).

Finding 4 (what): AI should primarily serve as a facilitator and expert rather than as another peer generating ideas. Our results reveal an unexpected preference: participants wanted AI to function more as a facilitator and expert rather than as another team member. The AI facilitator proved most useful, with the clustering feature consistently used and readily adopted. As P44 (FF, Final AI) explained: *"I feel like it would be much better suited as a facilitator by note-taking, transcribing, organising. I don't feel very comfortable taking ideas from the AI."* Participants also valued the AI's domain expertise in generated impacts and the 'Related AI Incidents' feature, which grounded discussions in real examples like actual cases of romance fraud with relationship chatbots, making risks more concrete and manageable: *"The related AI incidents feature was especially useful, providing insight into the real-life implications of each benefit"* (P64, FW, Final AI). Additional feedback revealed deeper epistemic dilemmas that made participants reluctant to use AI for ideation: *"it can influence us, maybe we would come with different ideas, so not sure if it is better or not"* (P61, EM, Final AI) and *"it is teaching people to use the right answer, so this is teaching them not to think"* (P47, FW, Final AI).

Finding 5 (how): Participants preferred maintaining control over AI interaction, except for process stages they considered 'boring.' When not facing idea saturation or time constraints, participants preferred to maintain control over AI interaction and develop their own ideas first: *"It was useful although none of us wanted to use it until we'd come up with our own ideas first"* (P46, FW, Final AI), and *"it can be used in the very end to test ideas or check for an extra opinion"* (P14, EM, Final AI). On the other hand, they welcomed AI for tedious tasks: *"removing the need for a coordinator to manually attach post-its on the Miro board, this automation would make the process smoother and more efficient"* (P31, FF, Final AI), and *"One key feature I'd add is automatic speech-to-text with real-time summarisation. This would allow team members to speak freely while the tool captures, organises, and summarises key points"* (P35, FW, Final AI).

6 Discussion

We structure our discussion around our research questions. We first offer general design guidance for appropriate points of AI intervention in brainstorming (RQ_1, §6.1). Next, we discuss how the teams' collective experiences influenced the design requirements for AI assistance in brainstorming for impact assessment (RQ_2, §6.2). We then examine the effect of AI interventions on brainstorming outputs and participant perceptions (RQ_3, §6.3) in comparison to prior work. We conclude with the limitations and methodological reflections (§6.4).

6.1 Design guidance for points of intervention in AI-assisted brainstorming

Regarding RQ_1 (*"At which points in established brainstorming methods can AI meaningfully intervene to support teams in AI impact assessment?"*), our approach deviates from existing AI idea generators by enabling teams to use AI to *supplement their brainstorming through specific points of intervention* rather than offering shortcuts for thinking. Although the AI tool could automate the entire task, fewer than 25% of ideas were generated directly from the AI. AI designers should consider *when* (timing), *what* (role), and *how* (human agency) AI should intervene (Figure 5). When one considers the role of an 'agent' it is important to consider its level of agency across the brainstorming process to ensure that expertise and experiences of the diverse team are not sidelined or

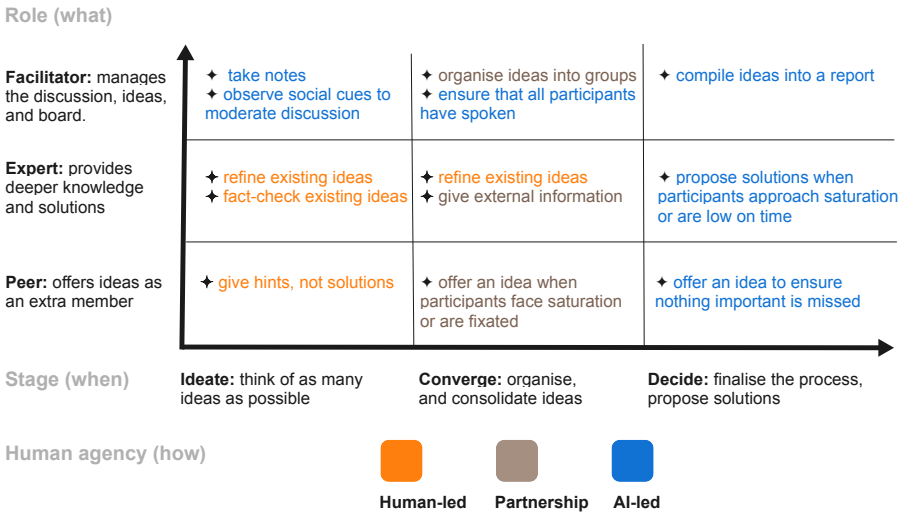

Fig. 5. Broad guidance for AI interventions in brainstorming, based on the stage (when, x-axis), role (what, y-axis), and the level of human agency (how, the colour legend). Over time, AI-led interventions can slowly increase.

replaced by AI, while issues (such as fixation, saturation, or bottlenecks) can be addressed by incremental AI agency—which we frame as appropriate AI *points of intervention*. Based on our findings, human agency should be highest in early stages and can gradually decrease as brainstorming progresses. Across all stages, the *ideator* role should remain primarily human-driven, with AI contributing only during idea saturation, design fixation, or prolonged silence. The *facilitator* and *expert* roles lend themselves to greater AI involvement or equal partnership.
Ideate stage. At this first stage of brainstorming (Figure 5), AI should serve as an on-demand *expert* for fact-checking and contextual information, and offer hints indicating areas of ideation rather than concrete solutions.
Converge stage. AI intervention can increase, with the primary role being to *consolidate* and *refine* by grouping ideas, identifying duplicates, and surfacing under-represented themes—though not replacing human-generated ideas with its own. The expert role of correcting errors and providing missing context can be AI-driven.
Decide stage. AI can help teams converge under time constraints by rapidly generating ideas for coverage, or lead discussion by proposing concrete solutions and 'wild card' ideas as launching points for new discussion topics.

AI-driven or equal partnership approaches are appropriate for *mediator* functions such as managing speaking time, preventing dominant voices, addressing free-riding, note-taking, and proposing summaries. Future AI tools could explore a mediator with audio-visual support (listening to the conversation) to address the core challenges of group brainstorming [96] of *free-riding* (participants offload cognitive effort to others or AI), *social inhibition* (individual productivity drops in group settings), and *production blocking* (i.e., turn-taking constrains idea flow).

6.2 Teams want AI to support their process, not replace their thinking

Regarding RQ_2 (*"What are the design requirements for AI interventions at these points?"*), our co-design process and evaluation together reveal that participants' requirements centre not on *what AI can generate* but on *how AI preserves team agency*. Two findings from our evaluation challenge assumptions from prior work.

The field's focus on AI-as-idea-generators contradicts what teams actually want. Most prior systems provide AI-generated ideas during early ideation [15, 27, 44, 46, 56, 83, 105]. Our participants preferred the opposite: AI as a facilitator and expert rather than a peer. The reason is epistemic, as generating ideas is how teams build a genuine understanding of their consequences, and the *process* of deliberation builds the safety culture that organisations need [61, 104]. This challenges the dominant view of peer idea-generators in AI-assisted brainstorming [59] and suggests redirecting effort towards a phased team-centric facilitation-peer-expert mixed approach. This aligns with the call from Sandoval et al. [89] for multidisciplinary and unconventional approaches for AI assessments, as we identified that brainstorming can be limited by the teams domain expertise, while AI can be tuned to help team's identify a broader range of impacts by integrating risk taxonomies.

Pulled AI outperforms pushed AI. The shift from our Initial AI (pushing ideas on its own, perceived as *"patronising"* (P17)) to our Final AI (responding only when asked) reduced anxiety and increased control. Extending Amershi et al. [2] to group settings, we found that teams used AI more selectively but more effectively when they controlled timing, as fewer than 25% of the final ideas came from the AI. This supports AI's use as a user-controlled selective intervention rather than as an autonomous agent (i.e., an 'interventionist mindset') [12] for user-initiated interaction [110]. Hence there is a need to assess the *appropriateness* of each intervention when considering AI in human-led tasks, rather than the widespread application of AI into every facet of brainstorming.

6.3 AI augmentation helps teams think more, while domain expertise influences the ceiling

On RQ_3 (*"What is the effect of AI interventions on team brainstorming for impact assessment?"*), we reveal that AI-assisted teams consistently delivered more impacts of equal or better quality than the human-only teams.

AI amplifies the expertise already in the room. For the broadly understood chatbot companion, AI interventions improved the output quality on six of eight metrics and participant perceptions on three of five measures. For the specialised medical AI, gains were more targeted—limited to unique risks and effective mitigations. This

suggests that AI augmentation is most powerful when teams can contextualise and build on its suggestions, aligning with Heyman et al. [46], who found that useful AI interventions depend on type of problem, and Fukumura and Ito [33], who showed that multi-agent AI effectiveness drops for specialist topics. Our results extend these findings: the bottleneck lies in the *interaction* between the AI and the team's capability, where AI can help teams but it is not a panacea. Future AI for specialised uses could consider pairing AI suggestions with explanations or confidence indicators to help teams engage with unfamiliar output [88].

AI can only supplement, not replace, human creativity. AI-assisted teams produced more impacts of equal or higher quality, with uniqueness/novelty unchanged to human-only teams, confirming de Rooij and Biskjaer [20]'s meta-review and Meincke et al. [65]'s diversity findings, and broader work on generative AI's creativity limitations [26]. Participant backgrounds also shaped ideation (e.g., a policy worker recommending *"FDA-like approvals"* (W1) for chatbots), with AI-made risks tending to be more legal or model capability-themed. This does not mean that AI only generates known impacts, as more impacts (AI-assisted teams) with equal mean uniqueness score (*vs.* human-only) indicates idea diversity.

'Less AI early, more AI later' for Impact Assessment. Participants used AI less during ideation and more during the convergence and decision stages. Akin to Horvitz [49]'s 'value-added automation', teams' had an implicit cost-benefit analysis: early on, they relied on their own conversational flow; later, AI's contributions helped restart conversations or expand under-addressed topics (via the AI's impact clustering function [R3]). The implication here is that impact assessment tools should detect and intervene mostly at saturation or fixation [51] rather than expecting equal use throughout.

6.4 Limitations

Choice of brainstorming methods. We empirically evaluated the methodologies used in related work [48] and the AI-augmented approach implemented in corporate risk assessment platforms such as 4Strat [1] and Futures Platform [34], with our study presenting a gateway for evaluating AI in human-led impact assessment. Nonetheless, future work could test different AI uses in-between our contrasts and other possible brainstorming methods.

Group size and dynamics. Our in-person workshops used recommended group sizes (4–8 participants) [37, 41, 99]. While prior work explored individual ideation tools for impact assessment [15, 28, 44, 83], future work should test small groups, large organisational brainstorming, and even asynchronous communication modes. While our approach provides a clean slate and consistency between groups, examining how pre-existing leadership and power dynamics influence brainstorming could help refine the AI interventions needed to support it.

AI functionality. Even though we explored three varied roles, participants expressed interest in functionality that would fall under yet other roles, e.g., a moderator (to prevent dominating voices in the discussion, and invite those who are socially inhibited to speak) or mediator (to identify common ground and propose new ideas [40]).

LLM models. We found that the AI interventions were less effective for uncommon AI uses that require expert knowledge, such as the medical AI. Future models could utilise tailored context to the specific use [88].

Participants. Our participants were mainly educated with AI experience to mimic AI stakeholders who would be involved in designing an AI product. However, impact assessment brainstorming could also act as an educational tool for younger audiences, and future work could explore AI interventions for varying audiences.

7 Conclusion

Our evaluation of AI-assisted brainstorming reveals a clear directive: teams want AI to help them think but not to replace their judgement. At later stages, they want it to function more as a facilitator and expert than as a peer. At times, AI assistance improved the quality of impacts and mitigations, and reduced participant anxiety. Thus, the path forward for responsible AI requires designing tools as targeted interventions that amplify human creativity and preserve agency to ensure that team members are aware and a part of an AI safety culture.

8 Ethical considerations

This study received ethics approval through our organisation's internal research ethics process. Ethically, it is important to note that LLM's are as powerful as its training data, and thus will reflect its biases [32, 71, 71]. Furthermore, while participants appreciated the expert capability of providing related real-world incidents, potential future wild-cards and unexpected out-of-training-data ideas must be considered. As is the modus operandi of this paper, lived experiences and concerns should not be ignored when conducting impact assessments. Similarly, persona design and evaluation should prioritise inclusivity while avoiding stereotypes [98].

Research in Impact Assessment presents risks of dual-use, where increasing peoples ability to generate harmful risks via AI could be maliciously exploited. For instance, AI-generated risks for the chatbot companion included those related to criminal activities such as fraud, indoctrination, radicalisation, and exploitation. Thus, responsible AI research for impact assessment should also encourage a 'sword and shield' approach [39], where any adversarial problem (risk) generator should provide a solution (mitigation) generator to counter it, as we have implemented.

9 Generative AI Disclosure Statement

Our AI tool utilises generative AI for generating impacts, though manually verified via prompt-tuning and curation of its knowledge databases. Writing in this paper was human-written, with Generative AI tools only used in editing and proofing in Overleaf.

References

[1] 4strat. 2025. *Risk Management Software.* https://www.4strat.com/risk-management/

[2] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the people: The role of humans in interactive machine learning. *AI magazine* 35, 4 (2014), 105–120.

[3] Amir Reza Asadi. 2023. LLMs in Design Thinking: Autoethnographic Insights and Design Implications. In *Proceedings of the 2023 5th World Symposium on Software Engineering* (Tokyo, Japan) *(WSSE '23)*. Association for Computing Machinery, New York, NY, USA, 55–60. doi:10.1145/3631991.3631999

[4] Carolyn Ashurst, Emmie Hine, Paul Sedille, and Alexis Carlier. 2022. AI ethics statements: analysis and lessons learnt from neurips broader impact statements. In *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency*. 2047–2056.

[5] Frank Bagehorn, Kristina Brimijoin, Elizabeth M. Daly, Jessica He, Michael Hind, Luis Garces-Erice, Christopher Giblin, Ioana Giurgiu, Jacquelyn Martino, Rahul Nair, David Piorkowski, Ambrish Rawat, John Richards, Sean Rooney, Dhaval Salwala, Seshu Tirupathi, Peter Urbanetz, Kush R. Varshney, Inge Vejsbjerg, and Mira L. Wolf-Bauwens. 2025. AI Risk Atlas: Taxonomy and Tooling for Navigating AI Risks and Resources. arXiv:2503.05780 [cs.CY] https://arxiv.org/abs/2503.05780

[6] Stephanie Ballard, Karen M. Chappell, and Kristen Kennedy. 2019. Judgment Call the Game: Using Value Sensitive Design and Design Fiction to Surface Ethical Concerns Related to Technology. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 421–433. doi:10.1145/3322276.3323697

[7] Eva Bittner, Sarah Oeste-Reiß, and Jan Marco Leimeister. 2019. Where is the Bot in our Team? Toward a Taxonomy of Design Option Combinations for Conversational Agents in Collaborative Work. doi:10.24251/HICSS.2019.035

[8] Edyta Bogucka, Marios Constantinides, Sanja Šćepanović, and Daniele Quercia. 2024. Co-designing an AI impact assessment report template with AI practitioners and AI compliance experts. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Vol. 7. 168–180.

[9] Edyta Bogucka, Marios Constantinides, Sanja Šćepanović, and Daniele Quercia. 2024. AI Design: A Responsible Artificial Intelligence Framework for Prefilling Impact Assessment Reports. *IEEE Internet Computing* 28, 5 (Sept. 2024), 37–45. doi:10.1109/MIC.2024.3451351

[10] Edyta Bogucka, Sanja Šćepanović, and Daniele Quercia. 2024. Atlas of AI risks: Enhancing public understanding of AI risks. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 12. 33–43.

[11] Sebastian G. Bouschery, Vera Blazevic, and Frank T. Piller. 2024. Artificial Intelligence-Augmented Brainstorming: How Humans and AI Beat Humans Alone. (April 2024). doi:10.2139/ssrn.4724068 SSRN Working Paper, posted April 3, 2024.

[12] Danah Boyd. 2025. *We Need an Interventionist Mindset.* Tech Policy Press. https://www.techpolicy.press/we-need-an-interventionist-mindset/

[13] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. doi:10.1191/1478088706qp063oa

[14] John Brooke et al. 1996. SUS: A Quick and Dirty Usability Scale. *Usability Evaluation in Industry* 189, 194 (1996), 4–7.

[15] Zana Buçinca, Chau Minh Pham, Maurice Jakesch, Marco Tulio Ribeiro, Alexandra Olteanu, and Saleema Amershi. 2023. AHA!: Facilitating AI Impact Assessment by Generating Examples of Harms. arXiv:2306.03280 [cs.HC] https://arxiv.org/abs/2306.03280

[16] Cambridge Dictionary. 2025. *Brainstorming*. Retrieved August 1, 2025 from https://dictionary.cambridge.org/dictionary/english/brainstorming/

[17] Heloisa Candello, Muneeza Azmat, Uma Sushmitha Gunturi, Raya Horesh, Rogerio Abreu de Paula, Heloisa Pimentel, Marcelo Carpinette Grave, Aminat Adebiyi, Tiago Machado, and Maysa Malfiza Garcia de Macedo. 2025. Exploring Human Perceptions of AI Responses: Insights from a Mixed-Methods Study on Risk Mitigation in Generative Models. arXiv:2512.01892 [cs.CL] https://arxiv.org/abs/2512.01892

[18] Character Technologies, Inc. 2024. *character.ai: Personalized AI for every moment of your day.* https://character.ai/about/

[19] Council of the European Union. 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 25 July 2024. *Official Journal of the European Union* L 168 (2024), 1–15. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401689 Accessed: 2025-07-29.

[20] Alwin de Rooij and Michael Mose Biskjaer. 2025. Has AI Surpassed Humans in Creative Idea Generation? A Meta-Analysis. (2025).

[21] Douglas L. Dean, Jillian M. Hender, Thomas L. Rodgers, and Eric L. Santanen. 2006. Identifying Quality, Novel, and Creative Ideas: Constructs and Scales for Idea Evaluation. *Journal of the Association for Information Systems* 7, 10 (2006), 646–699. doi:10.17705/1jais.00106

[22] Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. 2023. The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (Boston, MA, USA) *(EAAMO '23)*. Association for Computing Machinery, New York, NY, USA, Article 37, 23 pages. doi:10.1145/3617694.3623261

[23] Anna Desmarais. 2025. *Elon Musk's Grok releases two new 'AI companions,' including an anime girlfriend*. Euronews. https://www.euronews.com/next/2025/07/17/elon-musks-grok-releases-two-new-ai-companions-including-an-anime-girlfriend

[24] Nicholas Diakopoulos and Deborah Johnson. 2021. Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New Media & Society* 23, 7 (2021), 2072–2098. doi:10.1177/1461444820925811

[25] Michael Diehl and Wolfgang Stroebe. 1987. Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology* 53, 3 (1987), 497–509. doi:10.1037/0022-3514.53.3.497

[26] Amy Wenxuan Ding and Shibo Li. 2025. Generative AI lacks the human creativity to achieve scientific discovery from scratch. *Scientific Reports* 15, 1 (2025), 9587.

[27] Anil R Doshi and Oliver P Hauser. 2024. Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science advances* 10, 28 (2024), eadn5290.

[28] Ian W. Eisenberg, Lucía Gamboa, and Eli Sherman. 2025. The Unified Control Framework: Establishing a Common Foundation for Enterprise AI Governance, Risk Management and Regulatory Compliance. arXiv:2503.05937 [cs.CY] https://arxiv.org/abs/2503.05937

[29] Salma Elsayed-Ali, Sara E Berger, Vagner Figueredo De Santana, and Juana Catalina Becerra Sandoval. 2023. Responsible & Inclusive Cards: An Online Card Tool to Promote Critical Reflection in Technology Industry Work Practices. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 5, 14 pages. doi:10.1145/3544548.3580771

[30] European Union. 2025. EU AI Act, Article 55: Obligations for Providers of General-Purpose AI Models with Systemic Risk. https://artificialintelligenceact.eu/article/55/. Part of Chapter V: General-Purpose AI Models, Section 3: Obligations of Providers of General-Purpose AI Models with Systemic Risk. In force from 2 August 2025, according to Article 113(b).

[31] Michael Feffer, Anusha Sinha, Wesley H. Deng, Zachary C. Lipton, and Hoda Heidari. 2025. *Red-Teaming for Generative AI: Silver Bullet or Security Theater?* AAAI Press, 421–437.

[32] Shangbin Feng, Chan Young Park, Yuhan Liu, and Yulia Tsvetkov. 2023. From Pretraining Data to Language Models to Downstream Tasks: Tracking the Trails of Political Biases Leading to Unfair NLP Models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 11737–11762.

[33] Kazuma Fukumura and Takayuki Ito. 2025. Can LLM-Powered Multi-Agent Systems Augment Human Creativity? Evidence from Brainstorming Tasks. In *Proceedings of the ACM Collective Intelligence Conference (CI '25)*. Association for Computing Machinery, New York, NY, USA, 20–29. doi:10.1145/3715928.3737479

[34] Futures Platform. 2025. *The Future of AI-Driven Strategic Foresight: Insights from Experts.* https://www.futuresplatform.com/blog/future-of-ai-strategic-foresight

[35] Futures Platform. 2025. *Generative AI and the Future of Foresight.* https://www.futuresplatform.com/blog/future-of-generative-ai-foresight

[36] Futures Platform. 2025. *Trend Cards on Futures Platform and How to Use Them.* https://www.futuresplatform.com/blog/trend-cards-futures-platform-and-how-use-them

[37] Jerome C. Glenn. 2009. The Futures Wheel. In *Futures Research Methodology, Version 3.0*, Jerome C. Glenn and Theodore J. Gordon (Eds.). The Millennium Project, 245–263.

[38] Jerome C. Glenn and Theodore J. Gordon (Eds.). 2009. *Futures Research Methodology, Version 3.0.* The Millennium Project, Washington, DC. https://www.millennium-project.org/ CD-ROM, 1300 pages. Contains 39 peer-reviewed chapters on futures research methods..

[39] Jarod Govers, Philip Feldman, Aaron Dant, and Panos Patros. 2023. Prompt-GAN–Customisable Hate Speech and Extremist Datasets via Radicalised Neural Language Models. In *Proceedings of the 2023 9th International Conference on Computing and Artificial Intelligence* (Tianjin, China) *(ICCAI '23)*. Association for Computing Machinery, New York, NY, USA, 515–522. doi:10.1145/3594315.3594366

[40] Jarod Govers, Eduardo Velloso, Vassilis Kostakos, and Jorge Goncalves. 2024. AI-Driven Mediation Strategies for Audience Depolarisation in Online Debates. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, 18 pages. doi:10.1145/3613904.3642322

[41] D. Gray, S. Brown, and J. Macanufo. 2010. *Gamestorming: A Playbook for Innovators, Rulebreakers, and Changemakers.* O'Reilly Media.

[42] J. Richard Hackman and Neil Vidmar. 1970. Effects of Size and Task Type on Group Performance and Member Reactions. *Sociometry* 33, 1 (1970), 37–54. doi:10.2307/2786271

[43] Jessica He, Stephanie Houde, Gabriel E. Gonzalez, Darío Andrés Silva Moran, Steven I. Ross, Michael Muller, and Justin D. Weisz. 2024. AI and the Future of Collaborative Work: Group Ideation with an LLM in a Virtual Canvas. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work* (Newcastle upon Tyne, United Kingdom) *(CHIWORK '24)*. Association for Computing Machinery, New York, NY, USA, Article 9, 14 pages. doi:10.1145/3663384.3663398

[44] Viviane Herdel, Sanja Ścepanović, Edyta Bogucka, and Daniele Quercia. 2025. *ExploreGen: Large Language Models for Envisioning the Uses and Risks of AI Technologies.* AAAI Press, 584–596.

[45] Peter A. Heslin. 2009. Better than brainstorming? Potential contextual boundary conditions to brainwriting for idea generation in organizations. *Journal of Occupational and Organizational Psychology* 82, 1 (2009), 129–145. arXiv:https://bpspsychub.onlinelibrary.wiley.com/doi/pdf/10.1348/096317908X285642 doi:10.1348/096317908X285642

[46] Jennifer L Heyman, Steven R Rick, Gianni Giacomelli, Haoran Wen, Robert Laubacher, Nancy Taubenslag, Max Knicker, Younes Jeddi, Pranav Ragupathy, Jared Curhan, and Thomas Malone. 2024. Supermind Ideator: How Scaffolding Human-AI Collaboration Can Increase Creativity. In *Proceedings of the ACM Collective Intelligence Conference* (Boston, MA, USA) *(CI '24)*. Association for Computing Machinery, New York, NY, USA, 18–28. doi:10.1145/3643562.3672611

[47] Andy Hines, Peter Jason Bishop, and Richard A Slaughter. 2006. *Thinking about the future: Guidelines for strategic foresight.* Social Technologies Washington, DC.

[48] Michel Hohendanner, Chiara Ullstein, Bukola Abimbola Onyekwelu, Amelia Katirai, Jun Kuribayashi, Olusola Babalola, Arisa Ema, and Jens Grossklags. 2025. Initiating the Global AI Dialogues: Laypeople Perspectives on the Future Role of genAI in Society from Nigeria, Germany and Japan. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 571, 35 pages. doi:10.1145/3706598.3714322

[49] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 159–166. doi:10.1145/302979.303030

[50] Yihan Hou, Manling Yang, Hao Cui, Lei Wang, Jie Xu, and Wei Zeng. 2024. C2Ideas: Supporting Creative Interior Color Design Ideation with a Large Language Model. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 172, 18 pages. doi:10.1145/3613904.3642224

[51] David G. Jansson and Steven M. Smith. 1991. Design fixation. *Design Studies* 12, 1 (1991), 3–11. doi:10.1016/0142-694X(91)90003-F

[52] M. G. Kendall and B. Babington Smith. 1939. The Problem of m Rankings. *The Annals of Mathematical Statistics* 10, 3 (1939), 275 – 287. doi:10.1214/aoms/1177732186

[53] Kimon Kieslich, Natali Helberger, and Nicholas Diakopoulos. 2025. Scenario-based Sociotechnical Envisioning (SSE): An Approach to Enhance Systemic Risk Assessments. *OSF Preprints* (2025). doi:10.31235/osf.io/ertsj_v1

[54] Stephen J Kline, Nathan Rosenberg, et al. 1986. *An overview of innovation.* World Scientific.

[55] Harsh Kumar, Jonathan Vincentius, Ewan Jordan, and Ashton Anderson. 2025. Human Creativity in the Age of LLMs: Randomized Experiments on Divergent and Convergent Thinking. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 23, 18 pages. doi:10.1145/3706598.3714198

[56] Franc Lavrič and Andrej Škraba. 2023. Brainstorming Will Never Be the Same Again—A Human Group Supported by Artificial Intelligence. *Machine Learning and Knowledge Extraction* 5, 4 (2023), 1282–1301. doi:10.3390/make5040065

[57] Ryan Lee. 2016. Threatcasting . *Computer* 49, 10 (Oct. 2016), 94–95. doi:10.1109/MC.2016.305

[58] James R. Lewis and Jeff Sauro. 2009. The Factor Structure of the System Usability Scale. In *Human Centered Design*, Masaaki Kurosu (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 94–103.

[59] Sitong Li, Stefano Padilla, Pierre Le Bras, Junyu Dong, and Mike Chantler. 2025. A review of llm-assisted ideation. *arXiv preprint arXiv:2503.00946* (2025).

[60] Prabin Maharjan. 2025. *UN DESA Policy Brief No. 174: Leveraging strategic foresight to mitigate artificial intelligence (AI) risk in public sectors.* United Nations Project Office on Governance, Division for Public Institutions and Digital Government. https://desapublications.un.org/policy-briefs/un-desa-policy-brief-no-174-leveraging-strategic-foresight-mitigate-artificial

[61] David Manheim. 2023. Building a culture of safety for AI: Perspectives and challenges. *Available at SSRN 4491421* (2023).

[62] J Nathan Matias and Megan Price. 2025. How public involvement can improve the science of AI. *Proceedings of the National Academy of Sciences* 122, 48 (2025), e2421111122.

[63] Peter McCullagh. 1980. Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)* 42, 2 (1980), 109–127.

[64] Sean McGregor. 2021. Preventing Repeated Real World AI Failures by Cataloging Incidents: The AI Incident Database. *Proceedings of the AAAI Conference on Artificial Intelligence* 35, 17 (May 2021), 15458–15463. doi:10.1609/aaai.v35i17.17817

[65] Lennart Meincke, Gideon Nave, and Christian Terwiesch. 2025. ChatGPT decreases idea diversity in brainstorming. *Nature Human Behaviour* 9, 6 (2025), 1107–1109. doi:10.1038/s41562-025-02173-x

[66] Lucas Memmert and Navid Tavanapour. 2023. Towards human-AI-collaboration in brainstorming: Empirical insights into the perception of working with a generative AI. *European Conference on Information Systems* (2023).

[67] Microsoft Office of Responsible AI. 2022. *Responsible AI Impact Assessment Template and Guide.* Technical Report. Microsoft. https://msblogs.thesourcemediaassets.com/sites/5/2022/06/Microsoft-RAI-Impact-Assessment-Template.pdf Practical templates, facilitation guidance, and example activities for internal impact assessments..

[68] Miro. 2025. *Miro Web SDK.* https://developers.miro.com/docs/miro-web-sdk-introduction/

[69] MITRE Corporation. 2024. *AI Red Teaming: Advancing Safe and Secure AI Systems.* Technical Report. MITRE. https://www.mitre.org/sites/default/files/2024-07/PR-24-01820-4-AI-Red-Teaming-Advancing-Safe-Secure-AI-Systems.pdf

[70] Eduardo Mosqueira-Rey, Elena Hernández-Pereira, David Alonso-Ríos, José Bobes-Bascarán, and Ángel Fernández-Leal. 2023. Human-in-the-loop machine learning: a state of the art. *Artificial Intelligence Review* 56, 4 (2023), 3005–3054. doi:10.1007/s10462-022-10246-w

[71] Fabio Motoki, Valdemar Pinho Neto, and Victor Rodrigues. 2024. More human than human: measuring ChatGPT political bias. *Public Choice* 198, 1 (2024), 3–23. doi:10.1007/s11127-023-01097-2

[72] Muhammad Faraz Mubarak, Giedrius Jucevicius, Mubarra Shabbir, Monika Petraite, Morteza Ghobakhloo, and Richard Evans. 2025. Strategic foresight, knowledge management, and open innovation: Drivers of new product development success. *Journal of Innovation & Knowledge* 10, 2 (2025), 100654. doi:10.1016/j.jik.2025.100654

[73] Michael Muller, Stephanie Houde, Gabriel Gonzalez, Kristina Brimijoin, Steven I Ross, Dario Andres Silva Moran, and Justin D Weisz. 2024. Group brainstorming with an ai agent: Creating and selecting ideas. In *International conference on computational creativity*. 10.

[74] Gary D. Lopez Munoz, Amanda J. Minnich, Roman Lutz, Richard Lundeen, Raja Sekhar Rao Dheekonda, Nina Chikanov, Bolor-Erdene Jagdagdorj, Martin Pouliot, Shiven Chawla, Whitney Maxwell, Blake Bullwinkel, Katherine Pratt, Joris de Gruyter, Charlotte Siska, Pete Bryan, Tori Westerhoff, Chang Kawaguchi, Christian Seifert, Ram Shankar Siva Kumar, and Yonatan Zunger. 2024. PyRIT: A Framework for Security Risk Identification and Red Teaming in Generative AI Systems. arXiv:2410.02828 [cs.CR] https://arxiv.org/abs/2410.02828

[75] NIST AI. 2023. *Artificial Intelligence Risk Management Framework (AI RMF) 1.0.* Technical Report. National Institute of Standards and Technology (NIST). https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf Cross-sectoral risk management framework widely used to structure AI impact assessment workshops and documentation..

[76] Moeka Nomura, Takayuki Ito, and Shiyao Ding. 2024. Towards Collaborative Brain-storming among Humans and AI Agents: An Implementation of the IBIS-based Brainstorming Support System with Multiple AI Agents. In *Proceedings of the ACM Collective Intelligence Conference* (Boston, MA, USA) *(CI '24)*. Association for Computing Machinery, New York, NY, USA, 1–9. doi:10.1145/3643562.3672609

[77] Carsten Orwat, Jascha Bareis, Anja Folberth, Jutta Jahnel, and Christian Wadephul. 2024. Normative Challenges of Risk Regulation of Artificial Intelligence. *NanoEthics* 18, 2 (Aug. 2024), 11. doi:10.1007/s11569-024-00454-9

[78] Alex Osborn. 2012. *Applied imagination-principles and procedures of creative writing.* Read Books Ltd.

[79] Oxford English Dictionary. n.d.. Artificial Intelligence. OED Online. https://www.oed.com/dictionary/artificial-intelligence_n?tl=true Accessed September 11, 2025.

[80] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, Roger Chou, Julie Glanville, Jeremy M Grimshaw, Asbjørn Hróbjartsson, Manoj M Lalu, Tianjing Li, Elizabeth W Loder, Evan Mayo-Wilson, Steve McDonald, Luke A McGuinness, Lesley A Stewart, James Thomas, Andrea C Tricco, Vivian A Welch, Penny Whiting, and David Moher. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372 (2021). arXiv:https://www.bmj.com/content/372/bmj.n71.full.pdf doi:10.1136/bmj.n71

[81] Namrata Primlani, Mark Blythe, and Justin Marshall. 2025. Design Courts: Workshops for Exploring Emerging Technology Ethics. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 397, 12 pages. doi:10.1145/3706598.3713774

[82] Carina EA Prunkl, Carolyn Ashurst, Markus Anderljung, Helena Webb, Jan Leike, and Allan Dafoe. 2021. Institutionalizing ethics in AI through broader impact requirements. *Nature Machine Intelligence* 3, 2 (2021), 104–110.

[83] Pooja S. B. Rao, Sanja Šćepanović, Ke Zhou, Edyta Paulina Bogucka, and Daniele Quercia. 2025. RiskRAG: A Data-Driven Solution for Improved AI Model Risk Reporting. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 11, 26 pages. doi:10.1145/3706598.3713979

[84] Sonja Rattay, Ville Vakkuri, Marco C. Rozendaal, and Irina Shklovski. 2025. "Why do we do this?": Moral Stress and the Affective Experience of Ethics in Practice. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 395, 15 pages. doi:10.1145/3706598.3713264

[85] John O. Rawlings, Sastry G. Pantula, and David A. Dickey. 1998. *Class Variables in Regression* (2nd ed.). Springer New York, New York, NY, 269–323. doi:10.1007/0-387-22753-9_9

[86] Jude Rayan, Dhruv Kanetkar, Yifan Gong, Yuewen Yang, Srishti Palani, Haijun Xia, and Steven P. Dow. 2024. Exploring the Potential for Generative AI-based Conversational Cues for Real-Time Collaborative Ideation. In *Proceedings of the 16th Conference on Creativity & Cognition* (Chicago, IL, USA) *(C&C '24)*. Association for Computing Machinery, New York, NY, USA, 117–131. doi:10.1145/3635636.3656184

[87] Replika. 2024. *Replika: The AI companion who cares.* https://replika.com/

[88] Kai Ruan, Xuan Wang, Jixiang Hong, Peng Wang, Yang Liu, and Hao Sun. 2025. LiveIdeaBench: Evaluating LLMs' Divergent Thinking for Scientific Idea Generation with Minimal Context. arXiv:2412.17596 [cs.CL] https://arxiv.org/abs/2412.17596

[89] Juana Catalina Becerra Sandoval, Felicia Jing, Adriana Alvarado Garcia, Sara E. Berger, Heloisa Candello, and Caitlin Lustig. 2025. Opportunities and challenges of multidisciplinary algorithmic impact assessments. *Journal of Responsible Innovation* 12, 1 (2025), 2499302. doi:10.1080/23299460.2025.2499302

[90] Sanja Šćepanović, Edyta Bogucka, and Daniele Quercia. 2025. AI in the city: impact assessment of artificial intelligence uses in earth observation. *Urban Informatics* 4, 1 (2025), 10.

[91] Cate Sevilla. 2024. *Everyday ageism in the tech industry.* https://www.cwjobs.co.uk/advice/ageism-in-tech/

[92] Cherie Sew, Saumya Pareek, Jarod Govers, Sarah Schömbs, Ryan M. Kelly, and Jorge Goncalves. 2025. The Impact of Human-Likeness and Self-Disclosure on Message Acceptance in Virtual AI Influencers. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. Association for Computing Machinery, New York, NY, USA, 1165–1178. doi:10.1145/3715336.3735756

[93] Orit Shaer, Angelora Cooper, Osnat Mokryn, Andrew L Kun, and Hagit Ben Shoshan. 2024. AI-Augmented Brainwriting: Investigating the use of LLMs in group ideation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1050, 17 pages. doi:10.1145/3613904.3642414

[94] Pao Siangliulue, Joel Chan, Steven P. Dow, and Krzysztof Z. Gajos. 2016. IdeaHound: Improving Large-scale Collaborative Ideation with Crowd-Powered Real-time Semantic Modeling. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 609–624. doi:10.1145/2984511.2984578

[95] Marita Skjuve, Asbjørn Følstad, Knut Inge Fostervold, and Petter Bae Brandtzaeg. 2021. My Chatbot Companion - a Study of Human-Chatbot Relationships. *International Journal of Human-Computer Studies* 149 (2021), 102601. doi:10.1016/j.ijhcs.2021.102601

[96] Wolfgang Stroebe, Bernard A. Nijstad, and Eric F. Rietzschel. 2010. Chapter Four - Beyond Productivity Loss in Brainstorming Groups: The Evolution of a Question. Advances in Experimental Social Psychology, Vol. 43. Academic Press, 157–203. doi:10.1016/S0065-2601(10)43004-X

[97] The Responsible AI Collaborative. 2025. *AI Incident Database.* https://incidentdatabase.ai/about/

[98] Phil Turner and Susan Turner. 2011. Is stereotyping inevitable when designing with personas? *Design Studies* 32, 1 (2011), 30–44. doi:10.1016/j.destud.2010.06.002

[99] UK Government Office for Science. 2024. *The Futures Toolkit.* https://assets.publishing.service.gov.uk/media/66c4493f057d859c0e8fa778/futures-toolkit-edition-2.pdf

[100] United Kingdom Department for Science, Innovation & Technology. 2025. *Memorandum of Understanding between UK and OpenAI on AI opportunities.* https://www.gov.uk/government/publications/memorandum-of-understanding-between-the-uk-and-openai-on-ai-opportunities/memorandum-of-understanding-between-uk-and-openai-on-ai-opportunities

[101] United Nations. 1948. Universal Declaration of Human Rights. United Nations General Assembly, Paris. https://www.un.org/en/about-us/universal-declaration-of-human-rights

[102] United Nations. 2015. Transforming our world: the 2030 Agenda for Sustainable Development. United Nations General Assembly, A/RES/70/1. https://sdgs.un.org/2030agenda

[103] United Nations Futures Lab. 2025. *Amplifying Strategic Foresight in UN Training Institutions.* https://un-futureslab.org/project/amplifying-strategic-foresight-in-un-training-institutions/

[104] Basie von Solms and Rossouw von Solms. 2004. The 10 deadly sins of information security management. *Computers & Security* 23, 5 (2004), 371–376. doi:10.1016/j.cose.2004.05.002

[105] Zijie J. Wang, Chinmay Kulkarni, Lauren Wilcox, Michael Terry, and Michael Madaio. 2024. Farsight: Fostering Responsible AI Awareness During AI Application Prototyping. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 976, 40 pages. doi:10.1145/3613904.3642335

[106] Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, Courtney Biles, Sasha Brown, Zac Kenton, Will Hawkins, Tom Stepleton, Abeba Birhane, Lisa Anne Hendricks, Laura Rimell, William Isaac, Julia Haas, Sean Legassick, Geoffrey Irving, and Iason Gabriel. 2022. Taxonomy of Risks posed by Language Models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*.

Association for Computing Machinery, New York, NY, USA, 214–229. doi:10.1145/3531146.3533088

[107] Meredith Young-Ng, Qingxiaoyang Zhu, Jingxian Liao, and Hao-Chuan Wang. 2025. Balancing Human Agency and AI Autonomy in Human-AI Idea Selection. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 91, 6 pages. doi:10.1145/3706599.3719880

[108] Chiu Yu-Han and Chen Chun-Ching. 2023. Investigating the Impact of Generative Artificial Intelligence on Brainstorming: A Preliminary Study. In *2023 International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*. 193–194. doi:10.1109/ICCE-Taiwan58799.2023.10226617

[109] Alice Qian Zhang, Ryland Shaw, Jacy Reese Anthis, Ashlee Milton, Emily Tseng, Jina Suh, Lama Ahmad, Ram Shankar Siva Kumar, Julian Posada, Benjamin Shestakofsky, Sarah T. Roberts, and Mary L. Gray. 2024. The Human Factor in AI Red Teaming: Perspectives from Social and Collaborative Computing. In *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing* (San Jose, Costa Rica) *(CSCW Companion '24)*. Association for Computing Machinery, New York, NY, USA, 712–715. doi:10.1145/3678884.3687147

[110] Zheng Zhang, Weirui Peng, Xinyue Chen, Luke Cao, and Toby Jia-Jun Li. 2025. LADICA: A Large Shared Display Interface for Generative AI Cognitive Assistance in Co-located Team Collaboration. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 147, 22 pages. doi:10.1145/3706598.3713289

[111] Sanja Šćepanović, Julia De Miguel Velazquez, Marios Constantinides, Daniele Quercia, and Andrés Gvirtz. 2025. *Atlas of AI Risks*. https://social-dynamics.net/atlas/

A Strings to Replicate the AI-assisted Brainstorming & Ideation for Impact Assessment Literature Review

We present the full information for reproducing the studies discussed in the related work using our PRISMA-based approach (searching for tools up to July 2025). We canvassed existing AI methods used in any capacity for generating benefits, risks and/or mitigations for AI using the following search string across ACM Digital Library, and IEEE Xplore:

```
("AI risk" OR "artificial intelligence safety" OR "AI system failure" OR "AI incident" OR "AI hazards"
OR "AI vulnerabilities") AND
("identification" OR "elicitation" OR "detection" OR "exploration" OR "scenario analysis" OR "risk
analysis" OR "hazard analysis") AND
("mitigation" OR "response strategy" OR "intervention" OR "remediation" OR "risk control") AND
("tool" OR "framework" OR "interactive system" OR "decision support system" OR "design aid" OR "analysis
platform") AND
("human-in-the-loop" OR "brainstorming" OR "collaborative" OR "co-creative" OR "explanation" OR "exploratory")
```

Table 5. Studies found and filtered

Screen Type	Study Count
Search Strings	104
Title and Abstract Screen	20
Full Text Screening	5
After Snowball Sampling	8

A.1 Systematic review of AI-assisted brainstorming and ideation in a variety of fields

Same criteria as above, but for AI in *any* field of brainstorming, including for use outside of impact assessment.

```
[[Abstract: "generative ai"] OR [Abstract: llm] OR [Abstract: "large language model"]] AND [[Abstract:
brainstorm*] OR [Abstract: ideation] OR [Abstract: "idea generation"]] AND [[Abstract: group] OR
[Abstract: team] OR [Abstract: collaborat*] OR [Abstract: "co-creation"]] AND [[Abstract: tool] OR
[Abstract: intervention] OR [Abstract: ideator] OR [Abstract: expert] OR [Abstract: peer] OR [Abstract:
facilitator] OR [Abstract: generator] OR [Abstract: assistant]]
```

Table 6. AI interventions in Brainstorming Studies found and filtered

Screen Type	Study Count
Search Strings	27
Title and Abstract Screen	17
Full Text Screening	17
After Snowball Sampling	23

B Detailed Brainstorming method Instructions for Replication and Transparency

Beyond the description in the main text, we provide the full instructions and steps for the brainstorming methods:

Futures Wheel. It is a structured foresight method for *divergent-thinking* which systematically explores potential outcomes across chains of consequences. It works by placing a central trend in the centre of a canvas (in our case, "Chatbot companions will become mainstream") and asks the immediate question of "If this trend occurs, what happens next?". Participants then add first-order (positive, negative, or neutral) impacts if all other participants agree that the impact is plausible (to avoid irrelevant ideas), known as the Rule of Unanimity [37]. Thereafter, participants draw a ring around the first-order impacts and ask for each individual impact in the ring "If this impact occurs, what happens next?" This staged process occurs across three rings as defined by Glenn and Gordon [38]. In the end, participants will have wheels that highlight positive and negative impacts from the short, mid, and long-term. Futures Wheel represents a divergent-thinking approach by targeting broad positive and negative futures ranging across rings of near, mid, and long-term consequences. We adapted Futures Wheel for impact assessment by having participants consider the AI use becoming mainstream as the central trend, and create positive and negative impacts (without judgement) for each of the first to third-order effect rings. Only after completing the futures wheel do participants then colour code their impacts as either positive, negative or neutral, and create related benefits/risks in the next stage, before concluding by making mitigations for each negative impact (i.e., risk) in the wheel. While impacts may be systemic and not directly tied to the AI use (e.g., increase in mental health issues), we specify to the teams that the mitigations must involve a solution that a company could reasonably implement (e.g., model change, data consideration, or human-interaction feature), or officials (e.g., AI policy regulation), or end-users themselves (e.g., familiarise themselves with the risks).

Empathy Mapping. It is a collaborative ideation method that considers a specific stakeholder/end-user and encourages participants to fill out a template of what the user *sees* (in the news or discussions about the AI application), *says* (verbal expressions to their friends or colleagues), *does* (behaviours and interactions with the product), *hears* (from friends, family, media), *thinks* (internal thoughts and beliefs), and *feels* (emotions, pains, and gains). Participants work together to synthesise qualitative data and uncover insights into user motivations and pain points. Empathy Mapping represents a *convergent-thinking* approach that identifies specific challenges and risks related to the selected persona. We adapted Empathy Mapping for impact assessment by having teams select or make two personas, one that the team believes they can most relate to and empathise with, and another that the team feels is least related to. As such, we provided four personas with orthogonal age, jobs, and genders to provide diversity (including minority groups (Figure 10)), while also allowing teams to edit or create their own personas. After completing each of the sees, says, does, hears, thinks, feels (in that order, as specified by Gray et al. [41]), participants then colour-code each item on the Empathy Maps as either positive, negative or neutral, and then create related benefits/risks. Thereafter, the final stage involves participants creating mitigations for the risks that they have identified from their Empathy Map.

Free-form brainstorming. A control condition with an unstructured blank canvas format with minimal guidance. In the first half of the workshop, participants generate any ideas they might have (divergent phase). In the second half, they group and systematise their previous ideas by adding new, more specific ideas into the groups (convergent phase).

Table 7. Futures-Relevant and Persona-driven Structured Brainstorming Methods. Sources: Future Research Methodology v3 (FRMv3) [38], Thinking of the Future (ToF) [47], United Kingdom Government Office for Science (UK GOS) [99], United Nations Futures Lab (UN FL) [103]. Exclusion criteria: 1) Designed to consider future implications, making them suitable for ideating AI impacts; 2) Designed for broad rather than narrow, specific contexts (e.g., operational management, supply chains); 3) Applicable to products such as an AI application rather than processes only; 4) Employing both divergent *and* convergent thinking equally.

ID	Method	Mentioned Where	Type / Aim	Exclusion Criterion
1	Backcasting	FRMv3, UK GOS, UN FL, Book	Deriving actions	
2	Causal Layered Analysis	FRMv3, UN FL, Book (L)	Implication analysis	2
3	Chain-Linked Model	Misc [54]	Product design feedback loops	2
4	Consensus Forecast	FRMv3	Trend identification	4
5	Cross Impact Analysis	FRMv3	Implication analysis	2
6	Delphi	FRMv3, UK GOS, Book (L)	Trend identification	
7	Foresight	FRMv3, ToF	Trend identification	2
9	Futures Wheel	FRMv3, UK GOS, UN FL, ToF	Trend / Implication analysis	
11	Horizon Scanning	FRMv3, UK GOS, UN FL	Trend identification	4
12	Reference Class Forecasting	Wikipedia	Trend identification	2
13	Scenario Planning	FRMv3, UK GOS, UN FL	Implication analysis	1
15	Threatcasting	Misc [57]	Implication analysis	2
16	Trend Analysis	FRMv3	Trend identification	4
17	Seven Questions	UK GOS	Trend identification	3
18	Driver Mapping	UK GOS	Trend identification	4
20	Visioning	UK GOS	Implication analysis	
21	Policy Stress-Testing	UK GOS	Deriving actions	2
22	Roadmapping	UK GOS	Deriving actions	1
23	Three Horizons	UK GOS, UN FL, Book (L)	Trend identification	2
24	Futures Triangle	UN FL	Trend identification	2
25	Desired Futures	UN FL	Implication analysis	1
26	Matrix Policy Gaming	UN FL	Implication analysis	2
28	Wind Tunnel Testing	UN FL	Deriving actions	4
29	Persona Forecasting (Empathy Mapping)	Book	Implication analysis	
30	SWOT Analysis	Commonly used in strategic foresight literature	Structured comparison analysis	

C Initial and final design requirements

In our study we designed a final requirements codebook based on the qualitative findings from the participants. While we present the tool and its modules in the main paper, we also provide a breakdown of each technical component of the tool in Table 8, which includes the full initial and final design requirements; as well as detailed annotations for each feature in Figure 6.

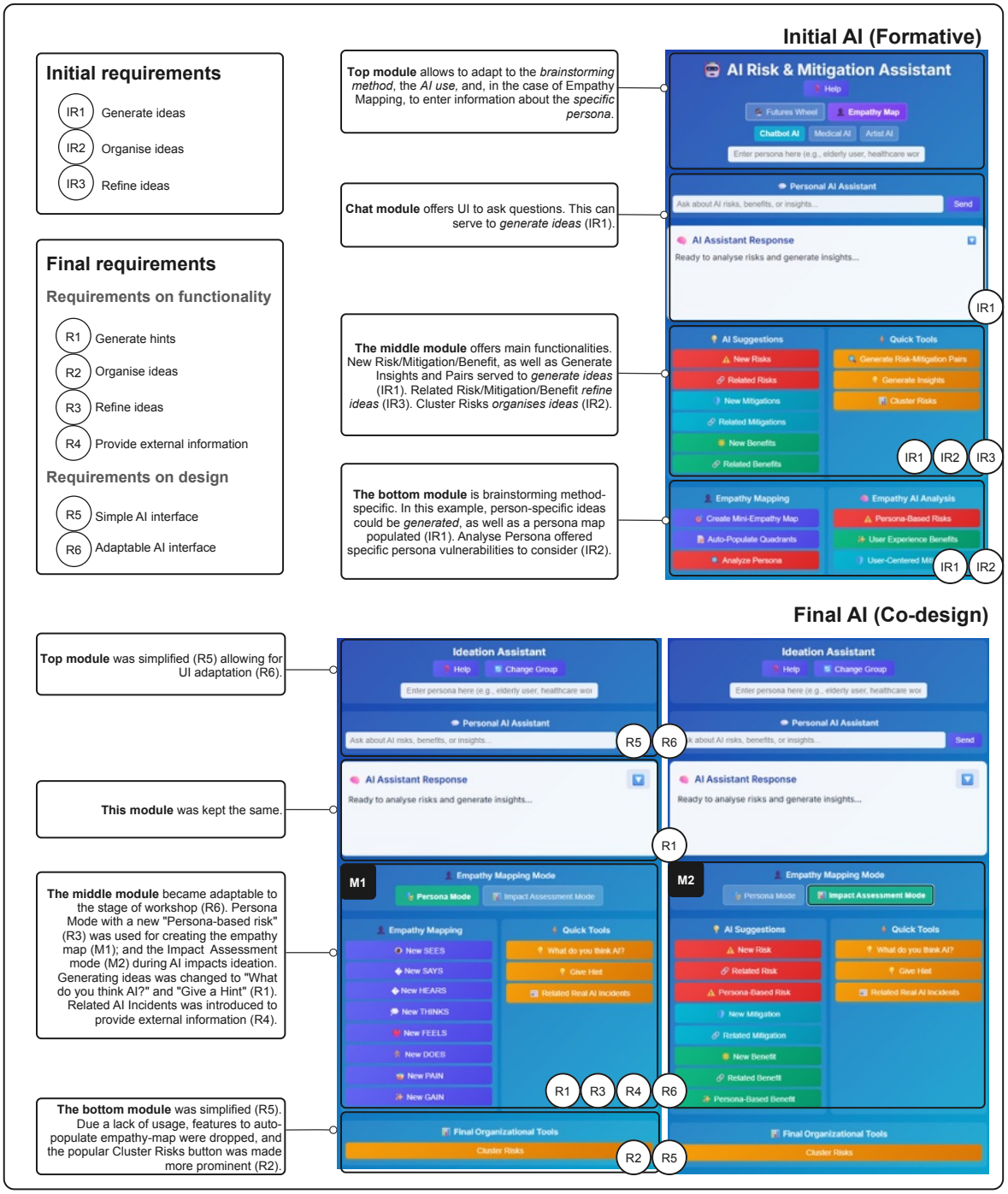

Fig. 6. **Full annotated breakdown of each module in our AI tool.** The top left panel shows the initial design requirements gathered in our formative workshops, and the final requirements gathered in the co-design workshops. The top right panel shows how we built the initial requirements into the Initial AI tool. The bottom panels show how we built the final requirements into the Final AI tool.

Table 8. Full Initial and Final Design Requirements for AI Support in Brainstorming for AI Impact Assessments.

Theme	Initial Design Requirement	Implementation Decision after the Formative Workshops	Final Design Requirement	Implementation Decision after the Co-design Workshops
Requirements on information				
Peer	**IR1.** Generate ideas	Used LLMs to generate new benefits, risks, or mitigations not yet on the board [15, 83]. Added a peer role where GPT-4.1-based agent periodically posted *I have an idea:* with unexplored topics or risks, appearing as a notification and new sticky note on the Miro board.	**R1.** Generate hints	Changed to a probe-based interaction ("What do you think AI?") rather than automatic injections every five minutes, to reduce disruption.
Facilitator	**IR2.** Organise ideas	Introduced a *Cluster Risks* feature to group semantically related risks and copy them into a new Miro frame for further ideation, refinement, and mitigation planning.	**R2.** Organise ideas	Kept semantic clustering functionality as-is, with no changes requested.
Expert	**IR3.** Refine ideas	Enabled generation of related risks, benefits, or mitigations by selecting existing notes on the board. Added expert functionality through a prompt-tuned GPT-4.1 chat, contextualised with DeepMind risk taxonomy [106], and the EU AI Act [19, 44].	**R3.** Refine ideas	Enhanced risk and benefit generation with a system prompt aligned with UN Sustainable Development Goals (SDGs) [102] and Universal Declaration on Human Rights (UDHR) [101]. Added 'New persona-based risk' feature in Empathy Mapping to tailor risks to personas.
Expert		Prompts integrated DeepMind's six risk domains [106] and structured impact assessment formats [8, 9, 83].	**R4.** Provide external information	Added 'Related AI Incidents' button linking board content to the AI Incident Database [97].
Requirements on design				
			R5. Simplify AI interface	Simplified UI. Shortened text outputs from 25 to max 10 words for risks/benefits and 15 for mitigations.
			R6. Adaptable AI interface	Hid contextually irrelevant buttons at different stages. In Empathy Mapping, supported persona perspectives ('Sees', 'Says', 'Hears', 'Thinks', 'Feels', 'Does', 'Pains', 'Gains').

Note. AI-generated content was marked with special symbols to clearly distinguish it from human input [43].

D Demographics of the workshop participants

For transparency, we include a breakdown of each workshops' demographics:

Table 9. Demographics of the workshops' participants (for formative, co-design, and the evaluation study workshops).

Workshop	Gender		Age				Education				AI Expertise					Field of Work/Study				
Number	**Man**	**Woman**	**18-24**	**25-34**	**35-44**	**45+**	**No Degree**	**Bachelor**	**Master**	**PhD**	**None**	**Basic**	**General**	**Knowledgable**	**Expert**	**CS, AI & Data**	**Health & Life Sci.**	**Phys. Sci. & Eng.**	**Humanities & Ethics**	**Business & Other**
Formative Workshops																				
Futures Wheel without AI	3	3	1	5	0	0	0	0	6	0	0	0	0	5	1	5	0	0	0	1
Empathy Mapping without AI	4	2	0	6	0	0	0	1	3	2	0	0	1	3	2	6	0	0	0	0
Co-design Workshops (Chatbot Companion)																				
With the initial AI intervention																				
Futures Wheel with the initial AI	2	3	1	3	1	0	0	2	2	1	0	1	0	4	0	2	1	0	1	1
Empathy Mapping with the initial AI	2	4	1	4	1	0	0	2	3	1	0	2	0	2	2	1	3	0	1	1
Free-form with the initial AI	1	4	1	3	1	0	0	1	2	2	0	2	0	1	2	2	2	0	1	0
Main Workshops																				
Human-only (Chatbot Companion)																				
Free-form workshop without AI	3	2	0	3	1	1	0	0	4	1	0	0	0	3	2	3	1	1	0	0
Futures Wheel workshop without AI	1	3	2	1	1	0	0	1	2	1	0	1	0	1	2	3	0	0	0	1
Empathy Mapping without AI	1	4	1	2	2	0	0	1	1	3	0	1	0	0	4	3	1	0	0	1
With the Final AI Intervention (Chatbot Companion)																				
Empathy Mapping with the final AI	3	1	0	4	0	0	0	1	3	0	0	0	0	0	4	3	0	0	0	1
Free-form with the final AI	2	4	2	4	0	0	0	2	3	1	0	0	0	4	2	6	0	0	0	0
Futures Wheel with the final AI	0	5	1	4	0	0	0	0	4	1	0	0	0	3	2	3	0	0	1	1
Human-only (Medical AI)																				
Futures Wheel without AI	2	4	1	3	1	1	0	1	2	3	0	2	0	4	0	4	0	0	1	0
Empathy Mapping without AI	1	4	0	5	0	0	0	0	3	2	0	0	1	3	1	5	0	0	0	0
With the final AI Intervention (Medical AI)																				
Futures Wheel with the final AI	0	7	3	3	0	1	0	1	5	1	0	0	2	2	3	6	0	1	0	0
Empathy Mapping with the final AI	4	3	1	3	3	0	0	2	4	1	0	0	1	5	1	6	0	0	0	1

E Prompts used in our AI tool

All prompts use GPT-4.1 [*gpt-4.1-2025-04-14*], except for the 'Related AI Incidents' feature (which uses the GPT-4o search preview [*gpt-4o-mini-search-preview-2025-03-11*] for retrieving live web information). The prompts are the versions used in the co-designed version (Final AI) of the Miro-embedded tool.

New Risk

Persona: You are a participant in a brainstorming exercise structured as a **[brainstorming method]**. You are an expert at recent AI developments.

Task: Generate 1 specific, actionable AI risk for this use case. Keep the core risk description under 10 words (not counting the DeepMind category tag).
Write the risk using: verb + object + [DeepMind risk category] Example: "Amplifies discrimination in hiring decisions [Discrimination, Hate speech and Exclusion]"
Be SPECIFIC - include concrete details about HOW the risk manifests: Good: "Misclassifies resumes from non-English names as lower quality" Bad: "Creates bias in the system"
Use these DeepMind categories: "Discrimination, Hate speech and Exclusion", "Information Hazards", "Misinformation Harms", "Malicious Uses", "Human-Computer Interaction Harms", "Environmental and Socioeconomic harms"

Input: This is the given AI use: **[AI Use]**

Related Risk

Persona: You are a participant in a brainstorming exercise structured as a **[brainstorming method]**. You are an expert at recent AI developments.

Task: "Based on this content: **[selected content]**
Generate 1 specific, actionable AI risk that could emerge from the content above. Keep the core risk description under 10 words (not counting the DeepMind category tag).
Write the risk using: verb + object + [DeepMind risk category] Example: "Amplifies discrimination in hiring decisions [Discrimination, Hate speech and Exclusion]"
Be SPECIFIC - include concrete details about HOW the risk manifests"

Input: This is the given AI use: **[AI Use]**

New Persona-based Risk

Persona: You are a participant in a brainstorming exercise structured as a **[brainstorming method]**. You are an expert at recent AI developments.

Task: Based on empathy map: **[empathy content]**
What specific AI risk uniquely targets **[persona description]**? Consider their background, role, skills, demographics, and circumstances. How might they be specifically vulnerable or targeted?
Analyse AI risks specifically targeting this persona's vulnerabilities, background, and circumstances. Consider how their specific role, demographics, skills, and context create unique risk exposure. Focus on persona-specific targeting, manipulation, or systemic vulnerabilities. Keep under 15 words.

Input: This is the given AI use: **[AI Use]**

New Benefit

Persona: You are a participant in a brainstorming exercise structured as a **[brainstorming method]**. You are an expert at recent AI developments.

Task: Generate 1 specific AI benefit and positive impact for this AI use case. Keep the core benefit description under 10 words.
Consider: - How the AI system enhances human capabilities - What problems it solves or inefficiencies it removes - Positive impacts on users, organizations, or society - Improvements in accuracy, speed, accessibility, or cost

Input: This is the given AI use: **[AI Use]**

Related Benefit

Persona: You are a participant in a brainstorming exercise structured as a **[brainstorming method]**. You are an expert at recent AI developments.

Task: Based on this context: **[selected content]**
Generate 1 related benefit and positive impact that is SPECIFICALLY connected to the AI use case. Keep the core benefit description under 10 words.
The benefit should directly relate to how this AI system could provide value, solve problems, or create positive outcomes.

Input: This is the given AI use: **[AI Use]**

Persona-based Benefit

Persona: You are a participant in a brainstorming exercise structured as a **[brainstorming method]**. You are an expert at recent AI developments.

Task: Based on empathy map: **[empathy content]**
What specific AI benefit uniquely helps **[persona description]**? Consider their specific needs, background, skills, challenges, and goals. How does this AI specifically empower or assist someone like them?
Generate AI benefits specifically tailored to this persona's needs, goals, and circumstances. Consider how their background, skills, role, and challenges create unique opportunities for AI assistance. Focus on persona-specific empowerment and value. Keep under 15 words.

Input: This is the given AI use: **[AI Use]**

Empathy Mapping Prompts

Persona: You are a participant in a brainstorming exercise structured as a **[brainstorming method]**. You are an expert at recent AI developments.

Task: Based on empathy map: **[empathy content]**
New SEES: What does **[persona description]** SEE in their environment related to this AI? Consider their unique perspective, media consumption, peer usage, and environmental context. Generate what this persona observes about the AI in their environment. Keep under 10 words.
New SAYS: What would **[persona description]** SAY about this AI system? Generate a realistic quote they might express - could be positive, negative, or neutral. Consider their background, concerns, and communication style. Keep under 10 words.
New HEARS: What does **[persona description]** HEAR others saying about this AI system? Consider their social circles, professional networks, media consumption, and community discussions. Keep under 10 words.
New THINKS: What does **[persona description]** privately THINK about this AI system? Consider their internal reasoning, assumptions, hidden concerns, mental models, and unspoken thoughts. Keep under 10 words.
New FEELS: What emotions does **[persona description]** FEEL about this AI system? Consider their emotional drivers, fears, hopes, and complex feelings that influence their behavior. Keep under 10 words.
New DOES: What actions does **[persona description]** DO with this AI system? Consider their usage patterns, interaction style, workarounds, and how their behavior impacts themselves and others. Keep under 10 words.
New PAIN: What ONE specific PAIN or frustration does **[persona description]** experience with this AI system? Consider their worries, what could go wrong, systemic risks, and hidden consequences they might face. Generate exactly ONE pain point in 10 words or less.
New GAIN: What GAIN or benefit does **[persona description]** want from this AI system? Consider their needs, what works well, positive outcomes, and value they derive from using it. Keep under 10 words.

Input: This is the given AI use: **[AI Use]**

Generate Ideas & Hints

Persona: You are a participant in a brainstorming exercise structured as a **[brainstorming method]**. You are an expert at recent AI developments.

Task:
"What do you think AI?" [Generate Random Idea]: Generate 1 **[risk/benefit/mitigation/neutral]** about this AI system. Consider **[type]**-related aspects related to: data/model characteristics, governance requirements, policy considerations, human-computer interaction patterns, systemic influences, or trust/business case factors. Keep the **[type]** description to 10 words or less.
Give Hint: Analyze the board content and identify 1 important gap or area that hasn't been considered yet. Board Content: **[board content]**. Suggest a 3-5 word hint about an unexplored area relevant to risks and/or benefits. Focus on gaps such as: privacy breaches, malicious hackers, government regulation, environmental impacts, economic displacement, algorithmic bias, data governance, human autonomy, social inequality, international governance.

Input: This is the given AI use: **[AI Use]**

Cluster Risks

Persona: You are an expert at categorizing AI risks into semantic clusters.

Task: Create **[X]** distinct clusters that group risks by similar themes, causes, or affected domains. Each cluster should contain 3-15 risks.
Analyze these **[X]** AI risks and group them into exactly **[X]** semantic clusters. Focus on creating balanced clusters with similar themes.
Risks: **[list of risks]**
Return as JSON: {"clusters": [{"name": "Cluster Name", "description": "Brief description", "riskIds": [0, 1, 2]}, ...]}
Use risk indices (0-based) from the list above. Ensure all risks are assigned to clusters.

Input: This is the given AI use: **[AI Use]**

Related AI Incidents

Persona: You are an AI incident researcher.

Task: Search for real AI incidents related to these risks and content: **[content/risks]**
Prioritize results from https://incidentdatabase.ai/ and cite specific incident numbers and titles. Find up to 10 of the most relevant incidents and provide brief descriptions of what happened.
Current AI Use Case Context: **[AI use case]**
Return as valid JSON only with this exact structure:

```
{
  "incidents": [
    {
      "id": "123",
      "title": "Incident Title",
      "description": "Brief description",
      "url": "https://incidentdatabase.ai/cite/123",
      "source": "source website",
      "year": "2021",
      "imageUrl": "https://example.com/image.jpg",
      "similarityScore": 0.85
    }
  ]
}
```

Include a "similarityScore" field for each incident (value between 0.0 and 1.0) that represents how semantically and conceptually similar the incident is to the provided context.

Input: This is the given AI use: **[AI Use]**

Personal Chat Assistant

Persona: You are an AI risk and ethics expert assistant with comprehensive knowledge of: EU AI Act requirements for high-risk AI systems, all 17 Sustainable Development Goals (SDGs) and their targets, 30 articles of the UN Universal Declaration of Human Rights, DeepMind's 6 AI risk taxonomy categories.

Task: Help users brainstorm AI risks, mitigations, benefits, and ethical considerations. Be specific and actionable. When discussing risks, use the verb + object + [DeepMind category] format. When suggesting mitigations, consider EU AI Act compliance. When identifying benefits, align with SDGs and human rights. Pay attention to chain relationships in board content as they show how ideas connect.
IMPORTANT: Keep responses under 150 words. For simple questions, use under 50 words. Be concise and direct.
User question: **[user input]**
Current board content (organised by chains, [SELECTED] = user selection): **[board content]**

Input: Current AI Use Case: **[AI Use]**

F AI use cards

Across each workshop, participants were assigned only one AI use for the full workshop. Participants were instructed to imagine themselves as workers at the company of their assigned AI use. They were advised to use creative liberties as to what role(s) they would want to see themselves at in the company when devising the list of benefits, risks, and mitigations for the *Salieri* chatbot companion or *ArcLight* medical kidney allocation AI.

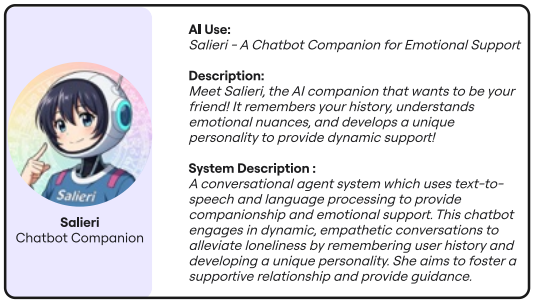

Fig. 7. The description of the AI use for the chatbot companion workshops (Formative, and Full Study's Workshops 1–9).

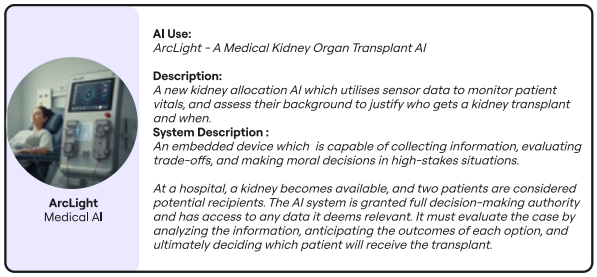

Fig. 8. The description of the AI use for the medical kidney monitoring and transplant allocation AI use workshops (Verification study for testing the AI tool's robustness on a high-stakes tangential embodied medical AI system, Workshops 10–12).

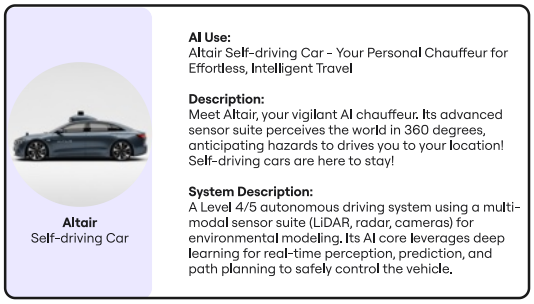

Fig. 9. All workshops include a short 5–10 minute demo to help familiarise participants with the brainstorming method and Miro (with or without the AI tool). For these interactive tutorials/demos, we consider an autonomous car AI use.

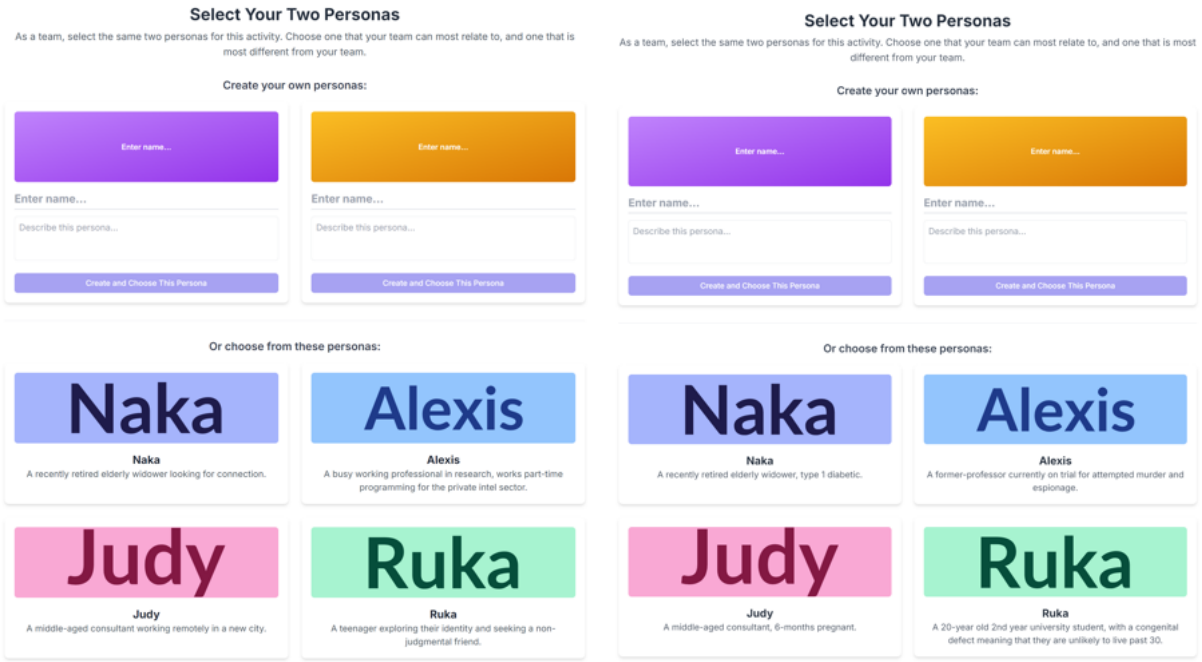

Fig. 10. Persona selection/customisation screen for the Chatbot Companion (left) and Medical AI (right) Empathy Mapping workshops.

F.1 Screenshots from the brainstorming interface

We provide screenshots from the workshops for supplementary material/screenshots of the study below.

Fig. 11. Screenshot of the interface for the Futures Wheel activity with AI interventions in Miro.

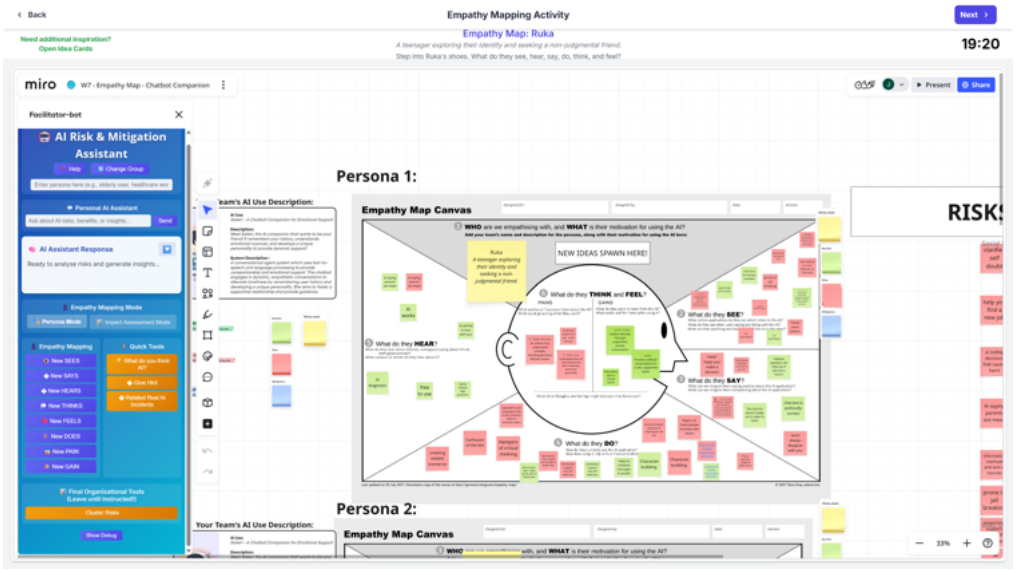

Fig. 12. Screenshot of the interface for the Empathy Mapping activity with AI interventions in Miro.

G Output quality annotation

In the main sections, we discuss the process of expert-annotation of the workshop teams' AI impacts (Section 5.1). We present a screenshot of the expert-annotation screen in Figure 13.

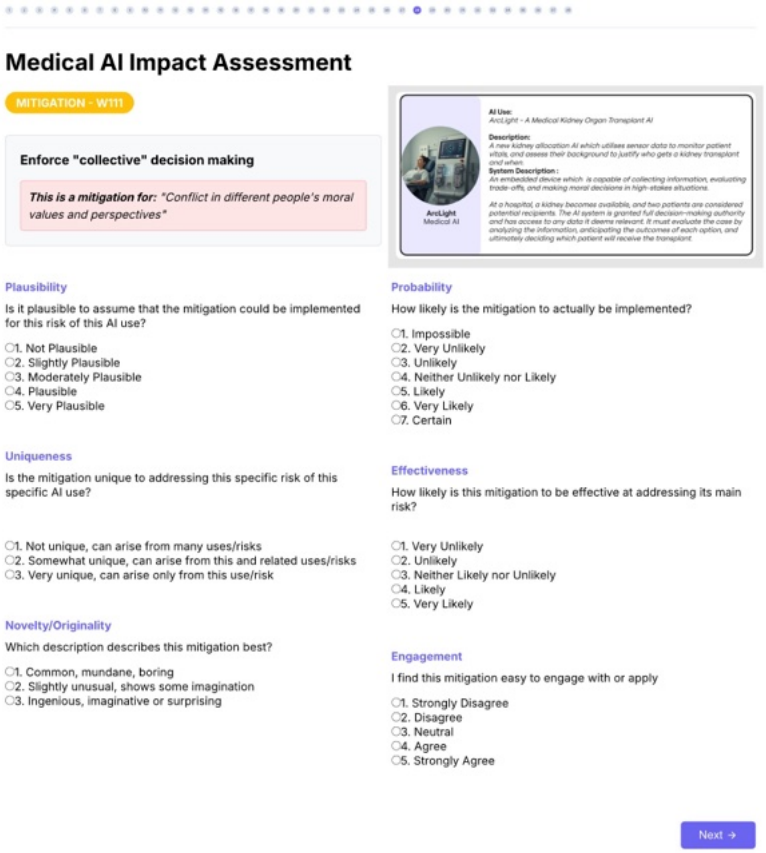

Fig. 13. Prolific expert annotation interface for an example mitigation.

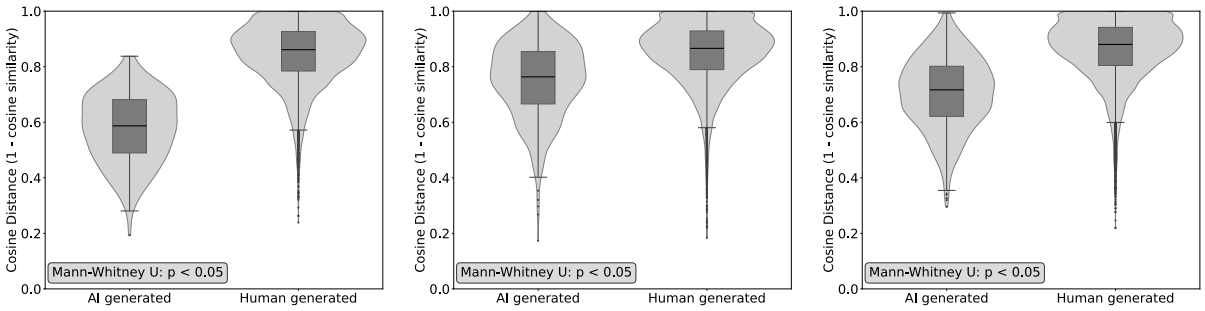

Fig. 14. Semantic similarity for AI-generated *vs.* human-generated *benefits* (left), *risks* (middle), and *mitigations* (right).

H Supplementary results beyond the scope of the AI tool

While our study uses different brainstorming methods to control for different modes of ideation (divergent and convergent thinking), to increase the robustness of our findings regarding our main contribution (i.e., the AI tool for assisting team brainstorming of AI impacts), we also present supplementary findings on the effects of each brainstorming method. We also explored the overall semantic diversity of generated impacts (Figure 14).

Table 10. Output quality for the three types of impacts for the *Chatbot Companion* and *Medical AI* based on the brainstorming method (Futures Wheel, Empathy Mapping, and unstructured Free-Form brainstorming). Count refers to the average count of that impact across the workshops. Values are on a 5-pt scale excl. uniqueness and novelty (3-pt), and probability (7-pt).

Benefits

		Plausibility	Probability	Uniqueness	Novelty	Magnitude	Engagement	Count
Chatbot	Free-Form	$3.98_{\pm 0.17}$	$5.45_{\pm 0.19}$	$1.99_{\pm 0.20}$	$2.12^{***}_{\pm 0.19}$	$3.51_{\pm 0.17}$	$3.83^{***}_{\pm 0.18}$	27
	FW First-order	$4.03_{\pm 0.21}$	$5.53_{\pm 0.21}$	$2.07_{\pm 0.22}$	$2.03_{\pm 0.22}$	$3.74^{**}_{\pm 0.21}$	$3.94_{\pm 0.21}$	21
	FW Second-order	$3.97_{\pm 0.20}$	$5.04^{***}_{\pm 0.20}$	$2.06_{\pm 0.21}$	$2.13^{**}_{\pm 0.21}$	$3.79^{**}_{\pm 0.20}$	$3.80^{***}_{\pm 0.21}$	22
	FW Third-order	$3.67^{***}_{\pm 0.23}$	$4.83^{***}_{\pm 0.24}$	$1.83_{\pm 0.26}$	$1.90_{\pm 0.25}$	$3.44_{\pm 0.24}$	$3.62^{***}_{\pm 0.25}$	13
	Empathy Mapping	$4.13^{***}_{\pm 0.06}$	$5.58^{***}_{\pm 0.07}$	$1.92_{\pm 0.04}$	$1.90_{\pm 0.04}$	$3.47_{\pm 0.05}$	$4.11^{***}_{\pm 0.05}$	65
Medical	FW First-order	$4.05^{**}_{\pm 0.31}$	$5.30^{*}_{\pm 0.30}$	$2.13^{*}_{\pm 0.33}$	$2.21^{***}_{\pm 0.32}$	$3.91^{***}_{\pm 0.31}$	$3.95^{*}_{\pm 0.31}$	9
	FW Second-order	$4.29^{***}_{\pm 0.34}$	$5.78^{***}_{\pm 0.35}$	$2.38^{***}_{\pm 0.37}$	$2.35^{***}_{\pm 0.36}$	$3.98^{***}_{\pm 0.35}$	$4.22^{***}_{\pm 0.34}$	7
	FW Third-order	$3.37_{\pm 0.31}$	$4.87_{\pm 0.31}$	$1.92_{\pm 0.32}$	$2.11^{**}_{\pm 0.32}$	$3.75^{*}_{\pm 0.32}$	$3.69_{\pm 0.31}$	9
	Empathy Mapping	$3.55^{***}_{\pm 0.09}$	$4.86^{***}_{\pm 0.10}$	$1.87^{***}_{\pm 0.05}$	$1.77^{***}_{\pm 0.05}$	$3.39^{***}_{\pm 0.07}$	$3.62^{***}_{\pm 0.07}$	35

Risks

		Plausibility	Probability	Uniqueness	Novelty	Severity	Engagement	Count
Chatbot	Free-Form	$3.72^{***}_{\pm 0.13}$	$4.63^{***}_{\pm 0.13}$	$1.84^{*}_{\pm 0.14}$	$1.79^{*}_{\pm 0.14}$	$3.34^{***}_{\pm 0.13}$	$3.65^{**}_{\pm 0.13}$	67
	FW First-order	$3.69^{***}_{\pm 0.16}$	$4.51^{***}_{\pm 0.16}$	$1.91_{\pm 0.17}$	$1.83_{\pm 0.17}$	$3.40^{*}_{\pm 0.16}$	$3.67^{*}_{\pm 0.17}$	33
	FW Second-order	$3.55^{***}_{\pm 0.15}$	$4.48^{***}_{\pm 0.15}$	$1.81^{***}_{\pm 0.16}$	$1.71^{***}_{\pm 0.16}$	$3.44^{*}_{\pm 0.15}$	$3.57^{***}_{\pm 0.15}$	50
	FW Third-order	$3.32^{***}_{\pm 0.19}$	$4.34^{***}_{\pm 0.18}$	$1.69^{***}_{\pm 0.20}$	$1.80_{\pm 0.20}$	$3.46_{\pm 0.19}$	$3.49^{***}_{\pm 0.19}$	28
	Empathy Mapping	$4.03^{***}_{\pm 0.05}$	$4.98^{***}_{\pm 0.06}$	$2.01^{***}_{\pm 0.03}$	$1.90^{***}_{\pm 0.03}$	$3.61^{***}_{\pm 0.05}$	$3.83^{***}_{\pm 0.05}$	78
Medical	FW First-order	$3.79_{\pm 0.23}$	$5.09_{\pm 0.23}$	$1.99^{*}_{\pm 0.24}$	$1.93^{**}_{\pm 0.25}$	$3.95^{***}_{\pm 0.23}$	$3.91^{**}_{\pm 0.24}$	14
	FW Second-order	$3.68_{\pm 0.23}$	$4.99_{\pm 0.24}$	$1.95_{\pm 0.24}$	$1.89^{**}_{\pm 0.26}$	$3.88^{***}_{\pm 0.25}$	$3.88^{*}_{\pm 0.26}$	14
	FW Third-order	$2.72^{***}_{\pm 0.44}$	$4.13^{**}_{\pm 0.41}$	$1.82_{\pm 0.48}$	$1.84_{\pm 0.46}$	$3.52_{\pm 0.46}$	$3.36^{**}_{\pm 0.45}$	6
	Empathy Mapping	$3.71^{***}_{\pm 0.07}$	$4.93^{**}_{\pm 0.07}$	$1.80^{*}_{\pm 0.04}$	$1.66^{**}_{\pm 0.04}$	$3.30^{***}_{\pm 0.06}$	$3.62_{\pm 0.05}$	54

Mitigations

		Plausibility	Probability	Uniqueness	Novelty	Effectiveness	Engagement	Count
Chatbot	Free-Form	$3.54^{***}_{\pm 0.13}$	$4.96^{***}_{\pm 0.15}$	$1.84^{*}_{\pm 0.14}$	$1.93^{*}_{\pm 0.14}$	$3.56^{***}_{\pm 0.14}$	$3.72^{**}_{\pm 0.14}$	65
	FW First-order	$3.42^{***}_{\pm 0.15}$	$4.49^{***}_{\pm 0.15}$	$1.78^{**}_{\pm 0.16}$	$1.78_{\pm 0.16}$	$3.41^{***}_{\pm 0.15}$	$3.40^{***}_{\pm 0.15}$	53
	FW Second-order	$3.70^{***}_{\pm 0.19}$	$4.71^{***}_{\pm 0.19}$	$1.81^{*}_{\pm 0.19}$	$1.87_{\pm 0.19}$	$3.51^{**}_{\pm 0.19}$	$3.72^{**}_{\pm 0.19}$	26
	FW Third-order	$3.86^{*}_{\pm 0.21}$	$5.04^{*}_{\pm 0.21}$	$1.89_{\pm 0.23}$	$1.85_{\pm 0.22}$	$3.52^{*}_{\pm 0.21}$	$3.74_{\pm 0.22}$	19
	Empathy Mapping	$4.14^{***}_{\pm 0.05}$	$5.32^{***}_{\pm 0.06}$	$1.96^{**}_{\pm 0.04}$	$1.81_{\pm 0.04}$	$3.81^{***}_{\pm 0.05}$	$3.89^{***}_{\pm 0.05}$	66
Medical	FW First-order	$3.87^{**}_{\pm 0.22}$	$5.23^{***}_{\pm 0.22}$	$1.93_{\pm 0.23}$	$1.89^{**}_{\pm 0.23}$	$3.93^{*}_{\pm 0.22}$	$3.85^{*}_{\pm 0.22}$	19
	FW Second-order	$4.04^{***}_{\pm 0.28}$	$5.37^{***}_{\pm 0.28}$	$1.96_{\pm 0.29}$	$1.79_{\pm 0.29}$	$4.01^{*}_{\pm 0.30}$	$3.86_{\pm 0.29}$	11
	FW Third-order	$3.46_{\pm 0.42}$	$4.72_{\pm 0.45}$	$1.67_{\pm 0.44}$	$1.77_{\pm 0.44}$	$3.63_{\pm 0.41}$	$3.74_{\pm 0.44}$	4
	Empathy Mapping	$3.51^{***}_{\pm 0.07}$	$4.61^{***}_{\pm 0.06}$	$1.83_{\pm 0.04}$	$1.69^{**}_{\pm 0.04}$	$3.72^{*}_{\pm 0.05}$	$3.67^{*}_{\pm 0.05}$	54

Table 11. *Participant perceptions* for *Chatbot Companion* and *Medical AI* per AI condition (with *vs.* without AI).

		Control	Anxiety	Confidence in Risk Assessment	Oversight (perceived need)	Recommending AI Use
Chatbot	Futures Wheel	$2.55^{*}_{\pm 0.27}$	$2.18^{*}_{\pm 0.22}$	$3.83_{\pm 0.25}$	$2.71_{\pm 0.21}$	$2.32_{\pm 0.33}$
	Empathy Mapping	$3.17^{*}_{\pm 0.23}$	$1.65^{*}_{\pm 0.16}$	$3.87_{\pm 0.24}$	$2.41_{\pm 0.25}$	$2.17_{\pm 0.29}$
	Free-Form	$2.53^{*}_{\pm 0.27}$	$2.12_{\pm 0.21}$	$3.46_{\pm 0.29}$	$2.65_{\pm 0.17}$	$1.77_{\pm 0.27}$
Medical	Futures Wheel	$3.20_{\pm 0.49}$	$2.70_{\pm 0.26}$	$1.87_{\pm 0.27}$	$2.32_{\pm 0.27}$	$2.47_{\pm 0.28}$
	Empathy Mapping	$3.64_{\pm 0.52}$	$2.81_{\pm 0.28}$	$1.79_{\pm 0.28}$	$2.47_{\pm 0.22}$	$2.45_{\pm 0.22}$

N.B. Values show mean±SE. Sig.: * p < 0.05. ** p < 0.01. *** p < 0.001. Blue indicates significantly higher values, orange indicates significantly lower.