ProMPT - Professional Modeling Prompting Tool

Filip Nykvist
Politecnico Of Turin
Sweden
s347839@studenti.polito.it

Sofia Lopes Suesdek Rocha
Politecnico Of Turin
Brazil
s334681@studenti.polito.it

Leonardo Merelli
Politecnico Of Turin
Italy
s334062@studenti.polito.it

Mohamad Parichehrehteroujeni
Politecnico of Turin
Iran
s328173@studenti.polito.it

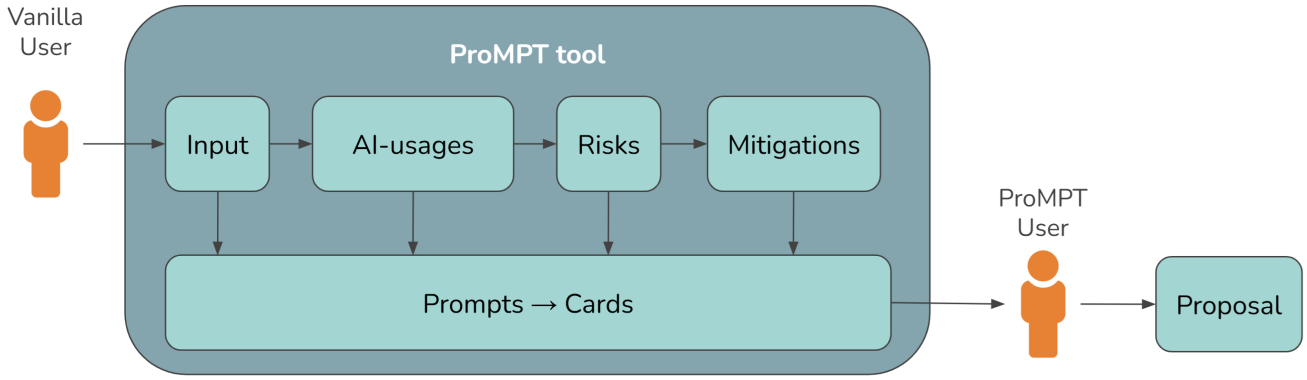

Figure 1: *ProMPT* Tool Flow: The user — an AI developer, policymaker, or civilian seeking better-informed decision-making process — inputs a post-war scenario. *ProMPT* identifies the region's main issues by retrieving and verifying conflict-related data. With the help of the tool it is then possible to map the issues for a potential AI solutions, using both a database and large language models (LLMs) for additional insight. The tool also identifies risks tied to the proposed AI uses and suggests mitigations. The tool outputs tailored prompts to insert into an LLM of choice. The results are compiled structured cards, which support the user in developing, evaluating, and refining AI-based solutions for post-war recovery.

Abstract

Post-conflict recovery is critically important for restoring stability, rebuilding institutions, and preventing renewed violence in war-torn regions. Artificial Intelligence (AI) offers significant potential to support this process through applications in humanitarian assistance, infrastructure restoration, and civil governance.

However, while AI has been widely used to enhance combat operations, its role in post-conflict contexts remains underexplored. Research has largely overlooked how AI can support peacebuilding, justice, and reconstruction, resulting in a lack of guidance.

To address this gap, we make three contributions. First, we propose a taxonomy of post-conflict problems, describing key characteristics of war and their impacts. Second, we develop a tool that, given post-conflict problems as input, generates prompts for large language models to explore AI-based solutions—outlining benefits, risks, and mitigations—and compiles them into impact assessment cards. Finally, we evaluate whether the tool can support decision-making for AI post-war solutions through a two-step study: in the first step, participants choose and critique AI solutions independently; in the second, they do so with the support of the impact assessment cards. We compare their responses to assess the tool's effect on structured decision-making. The ProMPT project repository is accessible at: https://github.com/Sofia-ME/ProMPT.git

Keywords

Artificial Intelligence, LLM, Post-war, Impact Assessment Cards

1 Methods

The conceptualization and development of this project were based on two key considerations that remained central throughout the process.

First, while Artificial Intelligence (AI) is increasingly playing a pivotal role in modern society and has been extensively applied to enhance combat effectiveness, its post-conflict potential remains underexplored. Existing research and applications have largely neglected the unique challenges of post-conflict environments, and there is a lack of specialized tools designed to guide the responsible use of AI in such contexts [2, 3, 12].

Second, public understanding of AI tends to be superficial and fragmented, and societal acceptance of AI technologies continues to be divided and inconsistent. This gap in awareness and

engagement further reinforces the need for tools that promote thoughtful, inclusive, and well-informed decision-making around AI deployment in sensitive environments.

To address these gaps, the methodology of this project is structured around three core steps framework:

(1) **Conducting a thorough investigation into war and post-conflict dynamics**, with the goal of building a specialized understanding of the typical problems, conditions, and recovery needs that arise in the aftermath of armed conflict.
(2) **Developing a tool** that supports both the responsible design of AI-based solutions and their informed implementation by key stakeholders, such as government agencies, international organizations, and civic groups involved in post-conflict recovery and policy development.
(3) **Evaluating and validating the tool**, named *ProMPT* — Professional Modeling Prompting Tool, through user-centered testing to assess its effectiveness in guiding structured, informed, and risk-aware decision-making.

The primary objective of *ProMPT* is to support users in making well-informed decisions regarding the implementation of AI solutions in post-conflict regions. The tool aims to help users:

- Identify relevant impacts of AI technologies more clearly;
- Think critically about associated risks and unintended consequences;
- Make more confident, conscious, and safer decisions;
- Propose meaningful adjustments that enhance the value or reduce the potential harm of selected AI interventions.

ProMPT achieves this by generating a sequence of prompts for a large language model (LLM), tailored to the user's selection of relevant post-conflict issues. These issues are drawn from a predefined and extensible dataset comprising a wide array of challenges commonly observed in post-war environments.

Once the user identifies the relevant problems, the tool constructs a customized prompt sequence, which is input into an LLM. The LLM then generates what this report refers to as Impact Assessment Cards (IACs)—structured, multi-dimensional summaries that assist the user in understanding benefits, risks, and appropriate mitigation strategies for potential AI applications.

Recognizing that the definition of a "problem" can vary significantly depending on cultural, political, and personal contexts, the dataset is designed to be flexible and expandable. This adaptability ensures that the tool remains relevant across diverse post-conflict settings and user groups.

The remainder of this section will detail the technical structure of the dataset, the logic behind prompt generation, and how these elements collectively align with the three-step methodology outlined above.

1.1 Step 1 - The Uppsala Conflict Data Program

To generate relevant and grounded prompts, the tool relies on a structured and scientifically validated database of armed conflicts. The *Uppsala Conflict Data Program* (UCDP) [11] provides detailed information on organized violence and active conflicts worldwide. By using UCDP's downloadable datasets in Excel format, it was possible to filter and extract relevant conflict data for integration into the tool. This allows the system to ground its outputs in empirically validated information and thus ensure that the generated IACs are context-aware and aligned with real-world future post-war situations. UCDP's taxonomy for actors, conflict types, and timelines further enables structured identification and prioritization of issues within the tools pipeline.

1.2 Step 1 - Taxonomy of Post-War Problems and Conflict Types

Building on the types of actors, conflict-types, and timeline classifications provided by the *Uppsala Conflict Data Program* (see section 1.1), we developed a taxonomy tailored to a post-conflict setting. This taxonomy can thus be used for problem-scraping and information–extraction that categorizes real-world challenges in a way that aligns with UCDP's framework. The taxonomy is:

- **Security**: landmines, militia violence, arms trafficking, civil unrest, violence.
- **Governance**: corruption, lack of elections, political instability, power vacuum.
- **Economy**: unemployment, collapsed economy, poverty, inflation.
- **Health**: trauma, disease outbreaks, mental health, lack of hospitals, malnutrition.
- **Displacement**: refugees, internally displaced persons, repatriation, migration.
- **Infrastructure**: destroyed infrastructure, lack of electricity, damaged roads, communication breakdown.

Similarly, to mirror UCDP's conflict-type classifications and thus ensure consistency in the context-setting, we defined the following list of conflict types:

- Civil war.
- Ethnic conflict.
- Religious conflict.
- Invasion.
- Border war.
- Insurgency.
- Terrorism.
- Revolution.
- Proxy war.
- Occupation.

Although the taxonomy was adapted to focus specifically on post-conflict challenges, it draws on UCDP's classification principles to maintain consistency with the conflict data used elsewhere in the system.

1.3 Step 1 - Finding post-war related problems

In order to maintain the relevance and completeness of our prompting-tool, we implemented a pipeline that can extract structured information from various sources such as news articles or Wikipedia pages and turn this into well-defined problems related to war situations. This way it is possible to transform user-defined web content into usable entries for downstream prompt generation using our tool. It further enables the system to

grow over time by incorporating new events and developments related to post-conflict environments.

The primary goal of the pipeline is to convert natural language articles into a structured format that captures the "who", "where", "what", and "when" of post-conflict situations. The found problems are then stored in CSV- or JSON-format for further use by the prompt-generation module.

The full process for updating the dataset can be divided into five main stages:

(1) **Article retrieval** is handled using the *newspaper3k* python library [4]. This library makes it possible to download and parse content from a given article URL. This tool also removes HTML noise, providing clean, readable text that can be processed further.
(2) **Analyzing article linguistically** using the "*en_core_web_trf*" model from the SpaCy python-library [7]. This is a transformer-based natural language processing model which is capable of high-quality tokenization, sentence segmentation, and named entity recognition (NER) [8]. Thus in this step, the article is broken down into individual sentences and annotated with entities such as people, organizations, locations, and dates.
(3) **Extracting the key contextual elements** using a combination of rule-based logic and fuzzy matching techniques with the python-library *rapidfuzz*. Fuzzy matching is basically a technique used to compare strings and determine their similarity, allowing approximate matches even when there are minor differences such as typos or formatting variations. This is exploited in order to, for each article, extract the following entities:
    - **Geographic regions** are identified by filtering named entities labeled as geopolitical entities (GPE) or locations (LOC), and are then ranked based on frequency of appearance in the text.
    - **Actors**, including individuals, groups, or organizations—are extracted using NER categories such as ORG (organization), PERSON, and NORP (nationalities, religious, or political groups), and are also ranked to ensure relevance.
    - **Post-war problems** are identified by comparing each sentence in the article against a taxonomy (see section 1.2) of issues (ex. "militia violence", "refugees", "mental health"). If a sentence semantically fuzzy-matches one of the keywords in the taxonomy, it is marked as describing a relevant problem.
    - **Timeline information** is extracted from date entities that include a four-digit year and appear near problem-related sentences, increasing contextual relevance.
    - **Conflict type** (ex. "civil war", "invasion", "proxy war") is inferred by sliding a window across the article and comparing text chunks against a taxonomy (see section 1.2) over conflict categories using fuzzy string similarity.
(4) **System constructs a structured record**, from the extracted data, for each sentence that contains a detected post-war problem. Each record includes the region(s), actor(s), problem keyword, descriptive sentence, relevant date (timeline), conflict type, and the original article's URL. The design also allows the system to extract multiple problems from a single article, ensuring a high degree of coverage. A typical entry for the dataset looks like:

    Region: South Sudan
    Conflict Type: Civil War
    Timeline: 2013
    Problem: Inflation
    Description: In October 2017, the IMF reported that real income had halved since 2013 and inflation was more than 300% per annum.
    Actors Involved: Nuer, SPLM, SPLA, Machar, Kiir, Dinka, Shilluk, UN
    Source: *wikipedia.org/wiki/South_Sudanese_Civil_War*

(5) **The output are saved** in both CSV- and JSON-formats. The filenames also include a timestamp and a user-defined case label (ex. "*SouthSudan_problems*_20250526.*csv*") for versioning and easy organization.

Currently there's no Graphical User Interface so the problem-scraping program requires mild programming skills in order to use. The entire data-extraction process is however encapsulated into a single function call: $analyze\_article(url, name, project\_path)$, where the variable *url* is a string containing the link to an article, *name* is a custom label for the particular run and $project\_path$ is a string that specifies where to locally store the programs output. The program is suited for both manual, single-article processing and automated batch updates.

1.4 Step 2 - AI uses, risks and mitigations

To generate accurate and context-aware prompts, we enrich them with real-world AI use cases, associated risks, and suggested mitigation strategies for each risk. These elements are sourced and structured using *ExploreGen* [6] and *RiskRAG* [10] inspired datasets, compiled in a CSV format.

ExploreGen is a LLM framework that creates realistic and diverse AI use case scenarios. It also classifies the associated risk levels based on the criteria defined by the EU AI Act [5].

RiskRAG, on the other hand, is a tool designed to enhance AI risk reporting. It addresses the shortcomings of traditional model cards, which often overlook risks or reuse generic, non-actionable content. *RiskRAG* identifies a wide range of model-specific risks, prioritizes them, proposes concrete mitigation strategies, and contextualizes them for real-world applications.

1.5 Step 2 - *ProMPT* - Professional Model Prompting Tool

ProMPT is a system designed to support AI-assisted reasoning in post-war recovery. Given a specific post-conflict scenario, it generates a sequence of prompts for LLMs to produce IAC that describe a problem, propose an AI solution, and outline risks and mitigations.

ProMPT addresses the challenge of navigating fragmented post-war data by grounding prompts in real-world conflict information, using the UCDP database and issue-specific web scraping. It simulates a comprehensive, collaborative, and policy-oriented workflow in which each prompt builds on the previous one to produce transparent, traceable reasoning and structured analysis.

The goal is to streamline decision-making, policy development, and strategic planning by providing users with structured, actionable insights. Evaluation will compare unaided writing with tool-assisted generation through a controlled user study.

ProMPT showcases how targeted AI prompting can enable transparent and scalable post-conflict reconstruction.

1.5.1 Database.

ProMPT works with two main databases:

- The first database catalogs post-conflict problems and factual war-related events, structured according to a predefined taxonomy. For the South Sudan 2013 civil-war case study, this was implemented by formatting a JSON file of conflict-related problems into structured strings.
- The second database consists of AI technologies, use cases, associated risks, and mitigation strategies. These entries are sourced and structured using datasets inspired by *ExploreGen* and *RiskRAG*, compiled in a CSV format. The risks and mitigations are further refined into keyword lists and sets that support accurate prompt generation and grounded policy evaluation.

Based on the user's selected conflict scenario, *ProMPT* dynamically accesses both databases to inform the content of each prompt. This ensures that AI suggestions and the corresponding IAC are both contextually relevant and grounded in empirical evidence.

1.5.2 Prompts.

ProMPT employs a sequential pipeline of seven prompts, each designed to simulate the roles of different research and policy experts. These prompts can be deployed using any preferred LLM; however, in the current case study, OpenAI's ChatGPT was used. The prompts were crafted following OpenAI's prompt design guidelines, emphasizing task decomposition, logical progression, and context preservation [9].

(1) **Prompt 1**: simulates a research assistant summarizing the post-conflict consequences based solely on verifiable data. The output is a set of factual bullet points describing humanitarian, infrastructural, political, and cultural damages.
(2) **Prompt 2**: represents a senior analyst reviewing the previous summary, filtering out speculative or unverifiable claims to ensure a clean, verified set of bullet points.
(3) **Prompt 3**: extracts the core post-conflict problems from the verified summary, producing concise statements that represent the ground-level issues needing intervention.
(4) **Prompt 4**: introduces AI technology by matching each identified problem with relevant AI tools and use cases from the database. Each match is justified in 1–2 sentences.
(5) **Prompt 5**: analyzes the realistic risks of applying each AI tool in the post-conflict environment, drawing on a curated keyword list from the CSV dataset.
(6) **Prompt 6**: suggests mitigation strategies that could realistically reduce or prevent the risks identified, tailored to the conflict context.
(7) **Prompt 7**: synthesizes all the previous outputs into complete IAC for each AI tool, summarizing the proposed use, associated risks, mitigation strategies, and key benefits in a post-conflict setting.

Each prompt in the pipeline enhances the output's quality, creating a traceable chain of reasoning that supports trustworthy and scalable recommendations.

1.5.3 Impact Assessment Cards (IAC).

The IAC are the final output of the *ProMPT* pipeline and were designed to address the need for high-quality, context-sensitive evaluations of AI applications in post-conflict settings, as highlighted in the literature [1, 6, 10].

Each card represents a distinct AI use case, matched to a specific post-war problem identified through the *ProMPT* analysis pipeline. Figure 2 shows one of the IAC generated by *ProMPT* and later used for the case study presented in the evaluation.

1 - Autonomous Vehicles

Use
Unmanned ground or aerial vehicles deliver aid where roads are unsafe or inaccessible.

Benefits
- Ensures consistent aid delivery despite road insecurity or destroyed infrastructure.
- Reduces risk to human drivers in hostile environments.

Risks	*Mitigations*
Electronic Interference: GPS jamming/spoofing may misdirect vehicles	Anti-jamming measures and safe-mode fallback protocols
Route Failures: vehicles may be stranded due to damaged infrastructure	Up-to-date post-war maps and redundant sensors (LiDAR, radar).
Military Exploitation: platforms could be hacked or repurposed for surveillance or attacks	International norms separating civilian vs. military use

Figure 2: Example of impact assessment card generated by *ProMPT*.

For every AI use, the card provides:

- A concise description of how the AI technology can be applied in the given post-conflict context, including the uses and benefits.
- An outline of risks associated with its deployment.
- Suggested mitigation strategies tailored to the conflict-specific scenario, aimed at reducing or managing those risks effectively.

The cards were developed with accessibility and clarity in mind. They are intended to be used by individuals from a range of backgrounds—including policymakers, AI developers, and members of the general public with decision-making influence, such as through referendums or civic participation. By presenting structured, scenario-specific information in an understandable format, the cards aim to facilitate more informed, transparent, and

robust decision-making about the role of AI in post-conflict reconstruction and recovery.

1.6 Evaluation Study

To explore public perceptions, concerns, and the perceived effectiveness of various proposed AI-based solutions, we conducted a controlled evaluation study. The study was designed to simulate a democratic policy-making process in which citizens deliberate and vote on publicly funded proposals. While the study focused on the context of the South Sudan conflict, this choice was not due to any unique suitability of the tool for that specific context.

The study involved 30 participants. Seventeen were recruited via the online platform Prolific, commonly used in behavioral and social science research. An additional thirteen participants were invited through academic networks, they all specializes in the engineering field or similar. All participants were over 18 years of age and provided informed consent before participating.

1.6.1 Procedure.

The study was structured into four sequential sections:

- **Section 1 – Demographics and Contextual Understanding:** Participants provided demographic information (age, gender, and country of origin), self-assessed their familiarity with AI and the post-war context in South Sudan, and read a brief background summary of the conflict. A short comprehension check ensured that they understood key contextual details.
- **Section 2 – Initial Solution Voting (Without IACs):** Participants were presented with four pairs of AI-based policy proposals. For each pair, they selected the solution they believed would have the greatest positive impact with the fewest negative consequences. After each choice, they rated the selected solution on three dimensions: perceived relevance, confidence in its effectiveness, and perceived risk, using a 1–5 Likert scale. Participants could also provide suggestions for improvement in an open-text response.
- **Section 3 – Voting with Impact Assessment Cards (IACs):** Participants revisited the same four solution pairs, now accompanied by Impact Assessment Cards (IACs) for each option. These cards provided structured information about each solution's use case, benefits, risks, and potential mitigations. Participants again made their selections and completed the same rating and feedback process.
- **Section 4 – General Feedback on the IACs:** In the final section, participants evaluated the IACs on six criteria: clarity, usefulness, ability to support deeper reflection, ability to surface new risks, influence on their decisions, and whether they would recommend using such tools in public policy settings. Each criterion was rated on a 1–5 Likert scale.

To assess the IACs, we analyzed average ratings across the six evaluation dimensions and conducted a thematic analysis of open-ended responses to identify perceived strengths and areas for improvement.

2 Results

This section summarizes the findings from the evaluation study, see section 1.6. The study involved 30 participants who used IAC generated with the *ProMPT* system to make decisions about AI-related technologies for post-war applications. The study provided both quantitative and qualitative feedback.

2.1 Participant Demographics

The study sample was relatively balanced and diverse. After cleaning the data, the gender distribution was 20 male participants and 10 female participants. The age distribution showed that 43.33% of participants were aged 18–24, followed by 26.67% aged 25–34 and 23.33% aged 35–44. Older age brackets (45–54 and 55+) were minimally represented, each at 3.33%. This suggests a sampling bias toward younger individuals.

The study also had a good spread of pre-knowledge about the conflict in South Sudan (which were the conflict focused on within the study, see section 1.6). 16/30 participants answered that hadn't heard about the conflict, or had heard about but didn't know any details. The other 14 participants answered that they had general understanding or where quite familiar with the topic, only one person answered that they knew the topic deeply.

Geographically, the participants in the survey spread across a range of countries. The most represented country was South Africa (8 participants), followed by Sweden (4), Italy (4), USA (3), Brazil (2), Iran (2), the United Kingdom (2) and Kenya (2). This global spread strengthens the survey's findings.

2.2 Overall Effectiveness of the *ProMPT* Cards

Participants responded strongly and positively to the IACs. The responses from each question regarding the overall effectiveness of the cards are compiled in table 1.

Table 1: General feedback on the use of IACs.

Question	Average Rating
What did you think of the cards, were they useful?	4.20
Were the cards useful in helping you reflect more deeply?	4.13
Did the cards make you consider new risks or impacts?	4.13
Did the cards influence your vote?	3.47
Were the cards easy to understand and clear?	4.33
Would you recommend the use of such cards in public decision-making processes?	4.43

As shown in Table 1, participants were generally positive toward the use of IACs, with the highest average score (4.43) indicating a strong willingness to recommend the cards for public decision-making. However, the cards received a lower average score (3.47) regarding their direct influence on participants' final

decisions, but still 100% of the individuals changed at least one of their selections after using the IACs.

While this limited change might initially seem like a drawback, it can also be interpreted as a positive feature. The goal of *ProMPT* is not to direct users toward any specific "correct" choice, but rather to support reflective, informed decision-making. In this light, the IACs appear to fulfill their purpose: encouraging users to consider broader impacts and ethical dimensions without pressuring them into predetermined answers.

It is also important to note that thoughtful deliberation may, in some cases, lead a participant to select a solution that appears less ethical on the surface but which they consider more practical or impactful within the specific context. This nuance highlights ProMPT's strength as a neutral scaffold for ethical reasoning, rather than a persuasive tool — allowing users to arrive at their own, well-considered conclusions.

2.3 Perceived Relevance, Confidence, and Risk

One interesting aspect of the analysis was comparing how participants viewed the relevance, potential harm, and their confidence in different AI solutions across the two rounds of choosing AI solutions. The findings seems to have a certain theme.

- **Relevance** of the chosen solution was on average rated very highly in both rounds, averaging 4.07 in Round 1 (without IAC:s) and 4.03 in Round 2 (with IAC:s).
- **Confidence** in the effectiveness of selected solutions slightly decreased from 4.23 to 3.93 between rounds, possibly reflecting deeper reflection or more cautious optimism after introducing additional perspectives from the IAC:s. This is interesting since the same thing occurred in the RiskRAG paper [10].
- **Perceived Harm** remained low, with averages of 1.93 in Round 1 and 1.90 in Round 2. This could indicate that participants found the proposed AI solutions broadly acceptable and not particularly risky in the first place.

One particularly notable finding from the survey was that question 2 stood out, with nearly one-third (27%) of participants changing their answer between Round 1 and Round 2. This rate of change was the highest among all four questions (see figure 3). The high response changing rate may indicate that the trade-offs presented in this question were especially difficult to assess, even after participants were given access to the IACs.

The AI solutions compared in this question included Speech Recognition and Voice Synthesis, which facilitates real-time translation and communication in multilingual refugee camps, and Natural Language Processing (NLP), which analyzes media to monitor peace agreement violations and community grievances. Both solutions address important, yet highly abstract issues; language access, misinformation, trust, and surveillance, which may make them more ethically and practically ambiguous than, for example, logistical tools like autonomous vehicles.

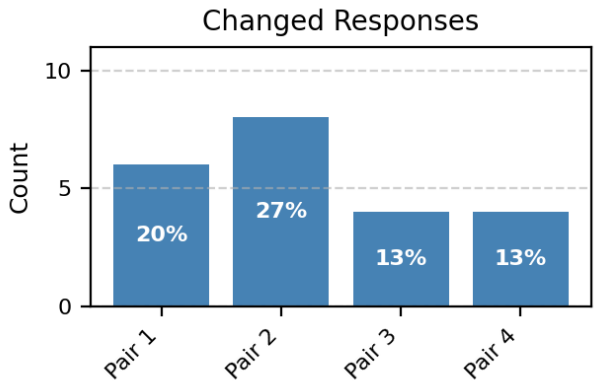

Figure 3: Number of participants who changed their responses between Round 1 and Round 2, with the introduction of IACs, for each preferred AI-related solution. Inside each bar is the percentage of the total number of participants in the study.

Overall, the metrics show a moderate level of consistency, particularly in perceived relevance and risk. This suggests that *ProMPT* helped participants engage more thoughtfully with these complex issues. It's also important to consider the level of ambiguity in the proposed solutions when using or further developing this tool, or similar ones.

2.4 Qualitative Shift in Participant Suggestions for Improvements of the Solutions

A comparison of participants open-ended suggestions for improving the AI-related solutions, before and after the introduction of the IACs, reveals a modest shift in the depth and quality of their answers. Notably, 10 out of 30 participants did not provide any suggestions in Round 1, which may reflect the difficulty many people face when asked to articulate risks or improvements in complex AI scenarios without some kind of structured support.

Of those who gave feedback in Round 1, four participants deepened or significantly elaborated on their answers in Round 2. This suggests that the IACs may have been effective in encouraging more reflective engagement among participants who were already actively thinking about the potential issues. Most of the 10 participants that were in Round 1 also remained silent during round 2 with one exception.

The following 2 examples illustrate how the IACs encouraged participants to reconsider or refine their suggestions:

- **Participant 14** gave *no suggestion* in round 1, but later wrote in round 2: "*This solution needs stronger governance mechanisms to ensure that voice synthesis is not used for surveillance or manipulation.*"
- **Participant 22** initially gave a vague comment in round 1:"*NLP system could help identify early signs of conflict.*" but expanded this in round 2: "*The NLP system could be helpful,*

but it must be transparent in how it analyzes grievances and who controls the system to avoid political misuse."

These examples further demonstrate that while the IACs may not always lead to a change in voting behavior, they can help with more thoughtful, critical, and ethically nuanced reasoning.

These examples further demonstrate that while the IACs may not always lead to a change in voting behavior, they are effective in supporting deeper critical and ethically nuanced reflection. The outcome aligns closely with the goal of the *ProMPT*-tool.

2.5 Summary of Findings and Reflections

The results could indicate that *ProMPT* has the potential to support participants in reflecting on complex and sensitive issues related to AI deployment in post-war settings. The slight decrease in confidence, alongside stable relevance and harm scores, may suggest that the IACs provided the user with valuable insights. This could further imply that our tool could work as a base framework for assessing AI related solutions to implement in the future.

Another interesting aspect is the low and stable perception of harm in both rounds. Despite being prompted to consider negative consequences through the IACs, participants continued to see the proposed AI solutions as generally safe. This may suggest that the IACs introduced risk awareness without generating fear or rejection, aligning well with ProMPT's goal of informed, balanced thinking.

Moreover, while the cards did not significantly alter participant votes (3.47/5 on influence), their effect on reflection (4.13/5) and ethical consideration (4.13/5) highlights the qualitative impact of the tool. Participants may not completely change their decisions, but they appear to reach them through more deliberative reasoning.

The results from the survey suggest that *ProMPT* functions more as a cognitive framework for building structured reasoning, rather than a persuasive device. Based on the study's findings, *ProMPT* seems to have helped the participants to slow down their thought processes, consider broader aspects and more consequences, and also better contemplated the "why" behind their decisions.

This is further supported by both quantitative and qualitative results. Although participants rated the tool's influence on their final decisions moderately (3.47), all of them altered at least one response, with question 2 showing a notable 27% switch rate.

Additionally, several participants elaborated more thoughtfully on their suggestions after being introduced to the IACs, moving from silence or vague comments to specific ethical or governance-oriented recommendations. These findings strengthen the case for ProMPT as a scaffold for structured deliberation, particularly when facing abstract or ethically ambiguous AI solutions.

References

[1] Anonymous. 2025. Impact Assessment Card: Communicating Risks and Benefits of AI Uses. In *Proceedings of the Conference on Computer-Supported Cooperative Work Social Computing (CSCW '25)*. ACM, New York, NY, USA, 1–42. https://researchswinger.org/publications/impact25cscw.pdf To appear.

[2] Laurie R. Blank and Daphné Richemond-Barak. 2024. *When Technology Meets Humanity: Shifting Policy Towards Protection in Conflict.* White Paper. The End of War Project. Available at https://www.aclu.org/sites/default/files/field_document/drone_dod_3160_01.pdf.

[3] Ingvild Bode. 2025. *Human-Machine Interaction and Human Agency in the Military Domain.* Policy Brief 193. Centre for International Governance Innovation. https://www.cigionline.org/publications/human-machine-interaction-and-human-agency-in-the-military-domain/ Accessed: 2025-06-01.

[4] Alexander Broussard and contributors. 2025. *newspaper3k.* https://github.com/codelucas/newspaper

[5] European Union. 2024. Regulation (EU) 2024/1244 of the European Parliament and of the Council of 13 March 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Available at: `https://eur-lex.europa.eu/legal-content`. Entered into force on August 1, 2024. Accessed: 2025-05-23.

[6] Viviane Herdel, Sanja Šćepanović, Edyta Paulina Bogucka, and Daniele Quercia. 2024. ExploreGen: Large Language Models for Envisioning the Uses and Risks of AI Technologies. *arXiv* 2407, 12454v1 (2024).

[7] Matthew Honnibal, Ines Montani, Sebastiaan Van Landeghem, and Adriane Boyd. 2025. *spaCy: Industrial-Strength Natural Language Processing in Python.* https://spacy.io

[8] IBM Research Editorial Team. 2020. *What is Named Entity Recognition (NER)?* https://www.ibm.com/think/topics/named-entity-recognition Accessed: 2025-05-26.

[9] OpenAI. 2024. OpenAI Cookbook. https://github.com/openai/openai-cookbook. Accessed: 2025-05-23.

[10] Pooja S. B. Rao, Sanja Šćepanović, Ke Zhou, Edyta Paulina Bogucka, and Daniele Quercia. 2024. ExploreGen: Large Language Models for Envisioning the Uses and Risks of AI Technologies. *University of Lausanne; International Institute of Information Technology Bangalore; Nokia Bell Labs; University of Nottingham; University of Cambridge* CH, 25 (2024).

[11] Uppsala Conflict Data Program. 2025. *UCDP Conflict Encyclopedia.* https://ucdp.uu.se

[12] Baobao Zhang, Allan Dafoe, et al. 2023. The Hall of AI Fears and Hopes: Comparing the Views of AI Influencers and those of Members of the U.S. Public Through an Interactive Platform. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing.* https://arxiv.org/abs/2310.12959

Appendix

Declaration

To construct the ProMPT-tool, the LLM ChatGPT where used to generate parts of the code. ChatGPT have also been used to try out different ideas during testing of the tool. Also in the report ChatGPT have been used for grammar- and spellchecking. Some parts of the text have been also been generated with the help of AI tools.

Division of Labour

The group divided the project into 9 parts, these are:

(1) List ongoing conflicts.
(2) Find problems connected to these conflicts.
(3) Find risks & and associated mitigations.
(4) Coding the *ProMPT* tool and building the prompts.
(5) Constructing the evaluation study.
(6) Launching the evaluation study on Prolific.
(7) Feasibility study
(8) RAI Guidelines Test
(9) Writing the report.

The tasks were then divided in the following way:

- **Filip** → 1, 2 & 6.
- **Sofia** → 4, 5 & 7.
- **Leonardo** → 3, 5 & 8.
- **Mohamad** → 5.
- **Everyone** → 9.

Team Diversity

Our team's diversity significantly contributed to the development and outcome of the ProMPT tool. The team comprises individuals with backgrounds in computer engineering, mechatronics engineering, data visualization, computer science, and artificial intelligence—providing the technical expertise required to design and implement the system. The group consists of three men and one woman and includes a wide range of professional experiences, from participation in student-led initiatives to internships in academic research labs. Culturally, the team represents Europe (Italy and Sweden), Asia (Iran), and South America (Brazil), and reflects a variety of religious affiliations. This diversity offered multiple perspectives on the nature and consequences of war, as well as on the global landscape of AI policies and regulations.

Received 26 May 2025; revised 26 May 2025; accepted 26 May 2025