Using Collective Dialogues and AI to Find Common Ground Between Israeli and Palestinian Peacebuilders

ANDREW KONYA, Remesh, USA

LUKE THORBURN, King's College London, UK

WASIM ALMASRI, Alliance for Middle East Peace, Palestine

ODED ADOMI LESHEM, The Hebrew University of Jerusalem, Israel

ARIEL D. PROCACCIA, Harvard University, USA

LISA SCHIRCH, University of Notre Dame, USA

MICHIEL A. BAKKER, Massachusetts Institute of Technology, USA

A growing body of work has shown that AI-assisted methods — leveraging large language models (LLMs), social choice methods, and collective dialogues — can help reduce polarization and foster common ground in controlled lab settings. But what can these approaches contribute in real-world contexts? We present a case study applying these techniques to find common ground between Israeli and Palestinian peacebuilders in the period following October 7th, 2023. From April to July 2024 an iterative deliberative process combining LLMs, bridging-based ranking, and collective dialogues was conducted in partnership with the Alliance for Middle East Peace. More than 100 civil society peacebuilders participated including Israeli Jews, Palestinian citizens of Israel, and Palestinians from the West Bank and Gaza. The process culminated in a set of collective statements, including joint demands to world leaders, with at least 84% agreement from participants on each side. In this paper we review the mechanics and implementation of the process, discuss results and learnings, and highlight open problems that warrant future work.

1 Introduction

This paper documents an effort to find common ground between Israeli and Palestinian peacebuilders amidst deeply entrenched divisions that were amplified in the immediate aftermath of October 7th, 2023 and shaped by the broader context of protracted asymmetric conflict. From April to July 2024, an iterative deliberative process that combined collective dialogues [1], bridging-based ranking [2], and LLMs [3] was carried out in partnership with the Alliance for Middle East Peace (ALLMEP). More than 100 civil society peacebuilders participated in the process including Israeli Jews, Palestinian citizens of Israel, and Palestinians from the West Bank and Gaza, representing diverse lived experiences and interests. The process produced a set of collective statements (Table 3) reflecting points of common ground — including five statements to world leaders, and five to residents of the region — with each statement supported by no fewer than 84% of peacebuilders on any side.

In recent years, there have been several proposals using algorithmic methods to address polarization and find common ground. Bridging-based ranking, a class of recommendation algorithms that aims to surface content that builds trust and mutual understanding across divides [2, 4], was first implemented on Australian policy forum *YourView* in 2012 [5], and later gained prominence when independently implemented on *Polis* [6, 7] and the Community Notes feature on X (then Twitter) [8]. Large-scale, online collective dialogues [1] that incorporate matrix completion methods to reduce the burden of preference elicitation have been used by the United Nations for peacebuilding since 2020 [9, 10, 11]. Most recently, there have been several proposals to use LLMs and social choice mechanisms to find common ground and make collective decisions on divisive issues [12, 13, 14, 15, 16].
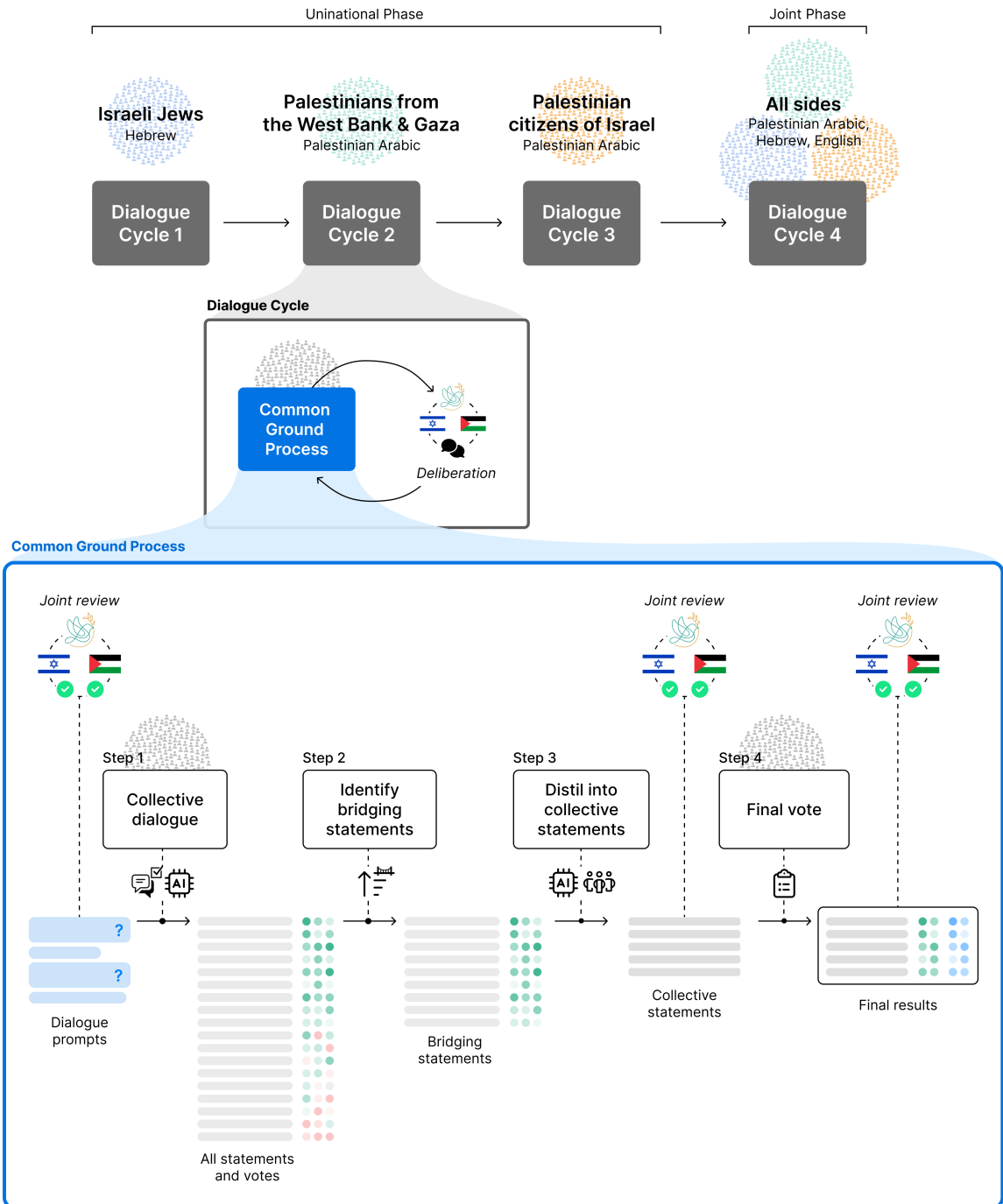
Fig. 1. The effort consisted of four dialogue cycles with peacebuilders: three uninational dialogues cycles conducted respectively with Israeli Jews, Palestinians from the West Bank and Gaza, and Palestinian citizens of Israel, followed by a final joint dialogue involving all three groups. Each dialogue cycle involved a process to find common ground among participants followed by a deliberation on the results of that process. The common ground process started with a collective dialogue, then bridging-based ranking was used to identify common ground 'bridging statements' that were distilled into articulate 'collective statements' via LLM and reviewed by human experts before being shared back with participants for a final vote. The method is described in detail in Section 3.

Most study of these methods — particularly bridging-based ranking and LLM-based approaches to articulating common ground — has occurred in lab settings, offline, or for relatively low-stakes contexts. For example, the approach used in this case study builds on a specific integration of these methods that was initially developed to produce common ground policy as part of OpenAI's Democratic Inputs to AI program [17], and subsequently refined for use in a study on AI alignment [18]. In contrast, we present a case study using a combination of these methods to find common ground between cross-border stakeholders amid active armed violence. This required us to address a range of real-world challenges related to trust, legitimacy, language barriers, and a constantly-changing situation. We believe this work may be the first to examine the use of LLMs to bridge divides during a high-stakes real-world conflict.

Given this context, it is important to note two principles that governed our approach:

- **Impartiality**: Drawing on best practices from peacebuilding [19] and mediation [20], we aimed for both our process and technology to be unbiased and fair: eliciting and representing participant perspectives without bias, and enabling equitable participation and influence over outcomes.

- **Pragmatism**: Given a context which presents urgent challenges, we prioritized iterative approaches to identify viable solutions, rather than pursuing ideal outcomes that would hinder progress due to extensive development or refinement.

In this vein, we also emphasize that this paper constitutes a case study, not research. The project was not undertaken for academic purposes, but as a good-faith attempt to support real-world peace processes. Opportunities for controlled experimentation and evaluation to inform design decisions were limited. As such, our approach reflects a pragmatic reconciliation of established techniques with the contextual constraints of a complex sociopolitical situation. Nonetheless, we believe the project can inform future academic work on the design and use of technology in conflict settings and are thus, with the permission of ALLMEP, publishing this case study.

The paper is structured as follows. Section 2 reviews situational and technical context which motivated the project. Section 3 documents the method used to find common ground and explains why key design decisions were made. Section 4 reports the main insights and statements (representing common ground) that resulted from the process. Finally, in Section 5 we discuss ethical considerations, the limitations of the process, and directions we intend to prioritize in future work.

2 Background

2.1 Situational

In the context of geopolitical conflicts, "track 1 diplomacy" involving formal negotiations between governments is often supplemented or supplanted by "backchannel" dialogue involving unofficial non-governmental representatives from each side. Such backchannel dialogues may involve government officials participating in an unofficial capacity ("track 1.5 diplomacy") or may not involve governments at all ("track 2 diplomacy") [21]. Often, these dialogues provide a more open environment to build trust and discuss tough issues. Furthermore, they can enable cross-border communication when formal diplomatic channels are closed. Civil society peacebuilders typically play a significant role in convening and contributing to these dialogues.

In early 2024, following the events of October 7th, 2023, and within the broader context of protracted asymmetrical conflict, tensions between Israeli and Palestinian peacebuilders had understandably intensified. Shaped not only by the recent escalation in violence, but also by decades of structural disparity, open hostility, and deep-seated mistrust, the

substrate to discuss tough issues had grown fragile[1]. In this context, we worked with the Alliance for Middle East Peace (ALLMEP) — a network of civil society peacebuilders in the region — with the goal of strengthening that substrate. This initiative sought to create a foundation for meaningful dialogue grounded in shared humanity and collective action. To this end, we sought to help peacebuilders find common ground. But the situation presented a few real-world challenges.

- **Language**: Peacebuilders' native languages span Hebrew, Palestinian Arabic, and English. Beyond communication, language carries deep cultural and historical significance, where it has been a tool of identity and a subtext of power imbalance. It was crucial for inclusivity and legitimacy that participants could engage fully in their native languages. Additionally, the sensitivity of language in this context extends to word choices that may reflect deeply rooted political and cultural meanings.

- **Trust**: Skepticism and mistrust, stemming from historical experiences of injustice, violence and prolonged conflict, posed significant barriers to engagement. This meant that before we could reasonably expect a willingness to participate in joint dialogue, we would need to elevate trust in good faith partnership. It also meant that the process itself would have to hold up to scrutiny that the "other side" did not have a mechanism to influence the results to their advantage.

- **Asymmetry**: We expected that, by default, there would be an unequal number of representatives participating from each side, and that it would be not be in the spirit of inclusion to limit the number of participants from each side to achieve balance. It was crucial to design a process that accounted for structural asymmetries, ensuring that even with unequal representation, all "sides" held equitable influence on the outcomes.

2.2 Technical

2.2.1 Collective dialogues

An online collective dialogue process is an iterative back-and-forth exchange between a moderator and participants. Participants typically join by clicking a link, and fill out demographic information as they enter the dialogue. During each turn of the dialogue, participants are sent either a read-only message, a standard survey question, or an open-ended prompt that kicks off a collective response process. During a collective response process [1], participants submit statements in response to the prompt, then evaluate statements submitted by others. The evaluation step serves two purposes: it exposes participants to others' views to help them understand each other better, and it elicits data for quantifying each statement's representativeness. On the collective dialogue platform *Remesh*, which we used for this project, two types of evaluations are elicited; agreement votes, and pair choice votes. Then, since participants can't feasibly vote on all statements submitted to each prompt, a matrix completion method [22, 9] is used to convert a sparse vote sampling into a complete participant vote matrix. The percent agreement each participant segment has with each statement is then computed by combining the complete vote matrix with the demographic data of the participants.

During a live collective dialogue process, all participation is simultaneous and each collective response process takes a few minutes. When each collective response process completes, each participant sees the fraction of all participants agreeing with their response along with a representative subset of responses from the group. The moderator sees live results which they can use to decide if they need to pivot from the pre-programmed discussion guide to ask additional

---

[1]We intentionally chose to use politically neutral language and not impose a characterization of the lived experiences of Palestinians and Israelis, including our own. This includes deliberately omitting certain terms, such as "occupation," that have widely recognized legal and political significance under international law and are central to the lived experiences of many participants in this process. We believe it is consistent with the spirit of this work that any political positions conveyed by this otherwise-technical paper should not be ours, but those manifest in the points of common ground found among the peacebuilders who participated in this effort (Table 3).

questions on an issue. After a live dialogue is complete it can be left open for asynchronous participation so those who could not make the live event can still participate. Collective dialogues can also be run asynchronously from the start if the logistics of a live dialogue are not feasible for the situation. In either case, a collective dialogue typically includes on the order of ten collective response prompts, and takes participants between 15-60 minutes.

The United Nations has run collective dialogues on Remesh — often referred to as "Digital Dialogues" — as part of peacebuilding efforts since 2020 [23]. These collective dialogues are sometimes conducted by special political missions in support of track 1 diplomatic processes, as was done by Stephanie Williams during ceasefire negotiations in Libya [24, 11]. In other cases, they serve less formal track 1.5 or track 2 activities associated with local field presences. Overall, the UN has now run more than 60 collective dialogues across the world including in Yemen [23] , Iraq [25] , Lebanon [26], Haiti [27], Bolivia [28, 29], and Bahrain [30].

2.2.2 Bridging-based ranking

The term bridging-based ranking refers to a class of ranking or recommendation algorithms that aim to "increase mutual understanding and trust across divides, creating space for productive conflict, deliberation, or cooperation" [4]. Most commonly, this qualitative goal is operationalized as "diverse approval," meaning that items are ranked by the degree to which they are approved of or valued by clusters of participants who would normally be expected to disagree with each other (in other words, ranking by diverse approval surfaces a form of "common ground"). Implementations of this core intuition differ, for example with respect to whether clusters are stipulated or learned from the data [6], and how diversity is quantified. Below, we review two existing approaches which we used in this project.

Let P denote a set of people, $\mathcal{G}$ denote a set of groups (that is, a set of subsets of P), and S denote a set of items to be ranked, which in our context are natural language statements.

- Perhaps the simplest approach is **max-min agreement**, where items are ranked by the minimum rate of agreement they receive across each group. Formally, for a group $G \in \mathcal{G}$ and statement $s \in S$, let $a(s, G)$ denote the proportion of people in group G that agree with statement s. Under the max-min agreement approach, statements in $s \in S$ are ranked according to the metric

$$a_{\min}(s, \mathcal{G}) = \min\{a(s, G) \mid G \in \mathcal{G}\}.$$

  This metric has been used previously in collective dialogue processes run on Remesh [17, 18].

- Another prominent approach to bridging is the algorithm used by the **Community Notes** feature on the social media platform X, which elicits crowd-sourced notes that add context to posts, but only shows them publicly if (to simplify slightly) they are rated as helpful by people on both sides of the political spectrum. To achieve this, the algorithm fits a latent factor model to the observed agreement votes, which provides for each note a factor corresponding to its slant along the most salient dimension of disagreement [8]. Formally, let $v(s, p) \in \{1, -1\}$ denote whether person p said they respectively agreed or disagreed with statement s. The Community Notes algorithm then fits the following model

$$\hat{v}(s, p) \approx \mu + i_p + i_s + \mathbf{f}_p \cdot \mathbf{f}_s.$$

  The global intercept μ captures the overall likelihood of any voter to agree[2] with *any* statement, the user intercept i_p captures voter p's overall likelihood to agree with *any* statement, and the note intercept i_s captures

---

[2]In the Community Notes implementation used for posts on X, the primary vote reflected in $\hat{v}(s, p)$ is not "agree", but "helpful". While the math is the same, here we describe the algorithm in terms of "agree" votes in order to maintain consistency across descriptions of the methods used in this paper.

the voting population's overall likelihood to agree with statement s. The last term $\mathbf{f}_p \cdot \mathbf{f}_s$ represents the product of the voter and statement factors,[3] respectively. Model parameters are fit by minimizing a regularized least-squared loss via gradient descent over the dataset of observed votes. Once learned, the note and user factors map statements and voters along learned axes spanning "sides" of division, while the note intercept i_s gives a side-independent measure of agreement. Because the note intercept i_s is side-independent, we can use it to find statements that bridge across people on different sides.

2.2.3 Language models

Large language models (LLMs) are increasingly being recognized for their potential to address societal challenges by enhancing collaboration and fostering collective intelligence. These models offer new ways to process and synthesize diverse viewpoints, enabling the articulation of shared perspectives even in deeply divided groups [31, 14, 15]. For example, in collective dialogue systems, LLMs can distill participant responses into coherent statements that reflect areas of agreement.

Beyond synthesizing viewpoints and identifying areas of agreement, LLMs can play a critical role in overcoming language barriers. They offer real-time translation, text-to-speech and summarization capabilities, ensuring participants can engage fully in their native languages. This functionality helps make deliberative processes more inclusive, equitable, and accessible [32]. However, careful implementation is essential to avoid risks such as bias and misrepresentation.

3 Method

3.1 Common ground process

The core of our approach is a process (Figure 1) that integrates collective dialogues, bridging-based ranking, and LLMs[4] to help identify, articulate, and validate points of common ground among a group of participants. The process begins with a collective dialogue (Step 1) in which participants anonymously respond to prompts and evaluate the responses of others. The responses that are most likely to represent common ground are identified using bridging-based ranking (Step 2) and distilled into a set of well-articulated collective statements using an LLM pipeline (Step 3). The collective statements are then shared back with participants for a final vote (Step 4). The primary outputs of each run of the process are a set of collective statements along with the results of the final vote. We describe each step in more detail below.

3.1.1 Step 1: Collective dialogue

The first step towards a collective dialogue is developing the planned prompts to which participants will respond and deciding on what demographics to gather from participants. Prompt drafting is informed by the goals of the dialogue: specific issues to find common ground on, areas to create more mutual understanding among participants, etc. Draft prompts are then reviewed and iterated until deemed acceptable by a stakeholder group with representation from all sides. The choice of demographics are determined by (a) demographic dimensions likely to manifest relevant divisions, (b) additional demographics needed to assess representativeness, and (c) constraints imposed by the need for statistical anonymity [33].

---

[3]These factors can be scalars that capture only the primary axis of voter variance, as in the original Birdwatch paper [8], or vectors that capture multi-dimensional factors of vote variance.
[4]For all LLM tasks in this project, we used models from OpenAI's GPT-4 series. Specifically, during the uninational phase we used `gpt-4-turbo-2024-04-09`, and during the joint phase we used `gpt-4-0125-preview`.

In this work, the collective dialogues were first run live, then reopened for asynchronous participation as needed. The date and time of the live dialogue was typically set a week or two in advance and communicated with participants so they could plan accordingly. The live event began with a Zoom video call where the dialogue was contextualized, then a link was shared with participants to join the collective dialogue on Remesh. After the live dialogue was over, it was re-opened for asynchronous participation for a few days.

Each collective dialogue generates the following raw data: a set of natural language statements S responding to each question; a sparse sampling of participant votes on each statement $s \in S$; the demographics collected for each participant, and data for any other non-collective-response questions asked during the dialogue. Vote inference [22] is then used to complete the agreement-vote matrix with elements $v(s, p) \in \{1, 0\} \equiv \{agree, disagree\}$ that indicate the (stated or inferred) agreement of participant p with statement s.

3.1.2 Step 2: Identifying bridging statements

The goal of this step is to identify a set of "bridging" statements $B \subseteq S$ that are likely to represent points of common ground. To do this, we computed two bridging metrics for each statement $s \in S$: (1) max-min agreement $a_{\min}(s, \mathcal{G})$ across defined demographic groups $\mathcal{G}$, and (2) the learned parameters (f_s, i_s) from the Community Notes algorithm (see Section 2.2.2). As a first set of bridging statements, we took those statements $s \in S$ for which $a_{\min}(s, \mathcal{G})$ was greater than a given threshold. This gives the set of statements that have the highest lowest agreement across the specified set $\mathcal{G}$ of demographic groups. However, when G contains many demographic segmentations — ie. by age, gender, religiosity, etc — this approach will exclude a statement that may otherwise bridge a critical axis of division just because it is predicted (perhaps incorrectly) to have low agreement among one of the demographic sub-groups. To mitigate this risk, we created a second set of bridging statements using the Community Notes approach, consisting of those for which the Euclidean distance $d((f_s, i_s), (0, 2))$ was below a given threshold. This approach relies primarily on the note intercept i_s as a bridging metric, but also penalizes statements for having non-zero slant, as indicated by the note factor f_s. The complete set of bridging statements B is taken to be the union of the sets of statements emerging from both approaches.

3.1.3 Step 3: Distillation into collective statements

The set of bridging statements identified in the previous step typically includes multiple restatements of the same idea, compound statements covering multiple ideas, and sub-optimal articulations of key ideas. The goal of Step 3 is to create well-articulated versions of each of the unique ideas that appear in the bridging statements in a way that preserves participants' choices of specific words and phrases. We use an LLM to do this in two steps. First, the LLM is prompted to extract the unique ideas from the set of bridging statements. Next, the LLM is then prompted to generate well-articulated versions of each of those ideas in a way that preserves the specific words and phrases found in the original statements. Included in this prompt are the set of unique ideas, the text of the bridging statements from which the unique ideas were derived, and a set of exemplar statements that demonstrate the style in which collective statements should be written. Finally, the collective statements proposed by the LLM are then reviewed by local language experts for linguistic consistency with the original bridging statements, and iterated as needed until a final set of collective statements C is deemed fair by a stakeholder group with representation from all sides. See Appendix A for LLM prompts used here.

3.1.4 Step 4: Final vote

We expect the collective statements that reach this point to generally represent common ground. However, the use of both AI and human experts in the prior steps introduce potential sources of error and bias. These risks can rightfully

delegitimize the claim that the collective statements actually represent common ground. The final vote is thus a critical step that serves to strengthen legitimacy by measuring support for the collective statements in a simple and transparent way. During the final vote, all participants vote on all collective statements. There is no AI inference or synthesis, and results are computed via deterministic and reproducible methods.

Participants cast two types of votes on each final collective statement. First, a simple vote of agreement with each statement is cast by each participant on a 5-point Likert scale. This enables a direct measurement of how well each statement collectively represents a point of agreement on all sides. In the second type of vote, participants are asked to rank the set of statements in order of agreement. Even if all collective statements represent clear points of common ground, they may not be equally important. The relative votes provide an additional signal to understand how the set of common ground collective statements should be prioritized.

3.1.5 Joint reviews

Even though the final vote serves as a hedge against biases that could cause collective statements to not represent common ground, the existence or perception of unfair bias at any point in the process can still undermine legitimacy. To mitigate this risk, at key junctures in the process where bias might be introduced, joint reviews were conducted by representatives of each side before moving on to the next step (see Figure 1). Specifically, joint reviews were conducted on: the content and wording of all messages, questions, and collective response prompts that would be seen by participants in the collective dialogue, the collective statements that would go on to the final vote, and the results computed from the final vote data.

3.2 Dialogue cycles

3.2.1 Transitional deliberation

We made our approach iterative by adding a transitional deliberation that marks the end of one dialogue cycle and the beginning of the next; serving to translate the outputs of one cycle to the inputs of the next cycle. These deliberations involved a small group of Israeli and Palestinian process facilitators and stakeholders. Each deliberation started with contextualization and sensemaking to establish mutual understanding related to the results, then transitioned to prompt design and participant communications for the next cycle.

3.2.2 Strategic sequencing

Our goal was to run a joint common ground process involving cross-border dialogue between peacebuilders from all sides, but before doing so we sought to build trust in the process and among those involved. To this end, we began with a uninational phase, where each dialogue cycle involved participants from only one side, in order to first find points of common ground within each group that could be communicated to help elevate trust and willingness to engage in the subsequent cycles.

What constitutes a "side"? It may seem obvious that "Israelis" and "Palestinians" would be the two sides. However, this binary approach would have risked marginalizing Palestinian citizens of Israel, may not fully identify with either Israelis or Palestinians. To accommodate this, we chose to treat Palestinian citizens of Israel as a third independent group, along side Israeli Jews and Palestinians from the West Bank and Gaza. In the remainder of this paper, we use $\mathcal{G}$ to denote this set of groups, namely

$$\mathcal{G} = \{\text{Israeli Jews, Palestinian citizens of Israel, Palestinians from the West Bank and Gaza}\}.$$

During the uninational phase, the group that went first needed to do so without the benefit of any trust-building feedback from earlier cycles. Given the existing power asymmetry, it was decided that Israeli Jews would thus go first. As the second major group, Palestinians from the West Bank and Gaza went second, followed by the Palestinian citizens of Israel. Finally, after these three uninational dialogues, we conducted a joint dialogue with participants from all groups to attempt to identify cross-border common ground (Figure 1).

3.3 Languages

It was critical for legitimacy and inclusivity that participants could fully engage in their primary languages. The uninational cycles would be mono-lingual (Hebrew or Palestinian Arabic), and the joint cycle would need to be multilingual. The main challenge for the mono-lingual cycles was the LLM pipeline used to generate collective statements. Early versions of the pipeline used in prior work [17, 18] had proven effective when the prompts, examples, bridging statements, and generated outputs were all in English, but it was unclear how to approach Hebrew and Palestinian Arabic. Should prompts or examples still be in English? And who decides? To pragmatically address these questions we set up two micro-experiments. We compiled a set of plausible bridging statements in each language and asked native speakers with prompt engineering experience to translate the English prompts and examples. Different variations of the pipeline were created by swapping out different permutations of the English prompts and examples with translated versions. Collective statements were generated using each variation. Then, blind to how they were generated, native speakers were asked which set of collective statements best reflected the content, culture, and language of the original statements they were based on. In all cases, reviewers preferred the statements generated by a pipeline with prompts and examples in their native language, so these variants were used for the uninational phase.

For the joint cycle, a trilingual approach encompassing Hebrew, Palestinian Arabic, and English was implemented to maximize accessibility and inclusivity. This required the provision of all textual materials — including collective response prompts, participant-generated statements, synthesized collective statements, and result presentations — in all three languages. All collective dialogue prompts were initially drafted in English, then translated to the other languages using the Google Cloud Translation API (hereafter, "Google Translate") and refined by native speakers. During the live dialogue, Google Translate was used to create trilingual versions of each statement in real-time. Acknowledging the inherent limitations of machine translation and the potential for semantic nuances to be lost or distorted, we adopted a transparent approach, explicitly informing participants about the use of Google Translate. Participants were also asked to use clear and concise language to mitigate the risk of translation errors. We conducted a set of micro-experiments on the multilingual LLM pipeline similar to those run for the monolingual pipeline. The final LLM output was a set of trilingual collective statements that would be reviewed and refined by native speakers for consistency across translations and with the original bridging statements. Final results were compiled in a spreadsheet and a trilingual lookup table for all elements in main view results, including labels and metrics, was initialized via Google Translate and then jointly reviewed and refined by native speakers.

3.4 Equal power metrics

For the joint cycle, it was critical that result metrics computed from the final vote data (including both the ranking and agreement votes) give each side equal influence even with unequal shares of participants. To that end, we computed three equal-power metrics — that is, metrics that are invariant to the portion of participants from each side — for each collective statement: max-min agreement, Dowdall score, and IRV rank.

- **Max-min agreement**: Max-min agreement, the same metric used for bridging-based ranking in Step 2, was computed again using the final agreement votes. Specifically, for each collective statement s, max-min agreement was computed as $\min\{\alpha(s, G) \mid G \in \mathcal{G}\}$ where $\alpha(s, G)$ is the fraction of participants in group G that voted "agree" or "strongly agree" on statement s. This metric inherently provides equal power across groups and offers a straightforward interpretation: it indicates the minimum level of agreement across all defined groups for statement s.

- **Dowdall score**: Similar to a Borda count [34], a Dowdall score is a cardinal metric computed from rank votes using a positional scoring rule that quantifies the strength of relative preference among a set of alternatives for a voting population [35]. Let $\text{Position}(s, k; P)$ give the fraction of participants (in population P) who ranked statement s in position k. A Dowdall score is computed as

$$\text{Dowdall}(s; P) = \sum_k \text{Points}(k) \times \text{Position}(s, k; P),$$

where $\text{Points}(k) = 1/k$ corresponds to the 'points' assigned to each position. We adapt this to be an equal power metric by taking the average of Dowdall scores computed independently for each side:

$$\text{Dowdall}_{\text{EP}}(s) = \frac{1}{|\mathcal{G}|} \sum_{G \in \mathcal{G}} \text{Dowdall}(s; G).$$

This averaging approach ensures that each group's collective preferences contribute equally to the final equal-power Dowdall score, regardless of group size.

- **IRV rank**: An instant runoff voting (IRV) rank is an ordinal metric derived from rank votes by iteratively eliminating the option with the fewest first-place votes among remaining options [34]. Let S_n be the set of statements remaining in round n, and $\text{First}(s, n; P)$ give the fraction of participants (in some population P) who rank statement s first among S_n. The statements remaining in round $n+1$ are $S_{n+1} = S_n \setminus \{\arg\min_{s \in S_n} \text{First}(s, n; P)\}$.[5] Denoting n_s to be the round where statement s was eliminated, and $|C|$ the total number of statements being ranked, the IRV rank is $\text{IRV}(s) = |C| - n_s + 1$. To make this an equal power metric, the 'first place fraction' $\text{First}(s, n)$ used to eliminate statements in each round is computed as an average of the first place fractions within each group. That is,

$$\text{First}_{\text{EP}}(s, n) = \frac{1}{|\mathcal{G}|} \sum_{G \in \mathcal{G}} \text{First}(s, n; G).$$

4 Results

The interpretation of the results presented herein necessitates careful consideration of several contextual factors. First, given the sensitive nature of the issues addressed in these dialogues and their direct relevance to ongoing peacebuilding efforts, the results shared are intentionally selective. Specifically, we present: (1) findings from the uninational phase that demonstrably facilitated the subsequent joint phase, and (2) collective statements of common ground from the joint phase that were explicitly intended for external communication. Furthermore, the quantitative results we share are limited to metrics that give equal power representation across primary participant groups.

---

[5]This formulation assumes no ties for last place, which was true in practice with our data.

Table 1. Top three *outgroup* red lines along with potentially related *ingroup* red lines from the other side. Note that the nature of relatedness between the red lines shown is non-specific, and does not imply direct alignment nor suggest equivalency in content or scope. In the column headers of this table, the adjectives "Israeli" and "Palestinian" refers to participants of the uninational dialogue cycles with Israeli Jews and Palestinians from the West Bank and Gaza, respectively.

Top Israeli outgroup red lines	Related Palestinian ingroup red lines
Harming Israeli citizens is crossing a red line פגיעה באזרחים ישראלים היא חציית קו אדום	Support or justify violence دعم العنف أو تبريره
The justification for the attack on October 7 is crossing a red line הצדקת ההתקפה ב7 באוקטובר היא חציית קו אדום	
Calling for the destruction of Israel is crossing a red line קריאה להשמדת ישראל היא חציית קו אדום	Refusing to recognize the Israeli side رفض الاعتراف بالجانب الاسرائيلي

Top Palestinian outgroup red lines	Related Israeli ingroup red lines
Call for forced displacement الدعوة إلى التهجير القصري	Talk about deporting Palestinians is crossing a red line שיח על גירוש פלסטינים הוא חציית קו אדום
Abolishing the existence of the other and not recognizing the rights of the Palestinians الغاء وجود الاخر وعدم الاعتراف بحقوق الفلسطينيين	Depriving Palestinians of human rights is crossing a red line קיפוח זכויות אדם של פלסטינים הוא חציית קו אדום
Supporting the war on Gaza and supporting the army and seeing that it is a moral army دعم الحرب على غزة ودعم الجيش ورؤية أنه جيش أخلاقي	Harming Palestinian civilians is crossing a red line פגיעה באזרחים פלסטינים היא חציית קו אדום

Second, we emphasize that these findings should not be understood as the views of the broader Israeli and Palestinian populations. The participant pool, with an approximate sample size of no more than $N = 120$ for any dialogue cycle, is statistically insufficient for generalization and, importantly, comprised exclusively of civil society peacebuilders.

Third, while some of the identified areas of common ground may appear self-evident in retrospect, particularly to those engaged in peacebuilding, it is essential to recall the initial context in which this project was undertaken. At that time, a prevailing sentiment held that finding meaningful common ground was unlikely, and the prospect of a constructive joint dialogue was viewed with considerable skepticism.

4.1 Uninational Phase

A key aim of the uninational phase was building trust and creating the context under which a productive cross-border engagement in the joint phase could take place. To this end, each cycle was designed to identify *red lines*, *values*, and *visions* for the future.

4.1.1 Red lines

In our work, the term "red lines" refers to actions (including speech acts) that those participating in the process believe are unacceptable, and if undertaken may prevent further dialogue. We asked about two kinds of red lines:

- **Ingroup**: the lines that each side themselves would not cross.
- **Outgroup**: the lines that each side would not accept the "other side" crossing.

These red lines served a few purposes. Ingroup red lines would provide a trust building signal if they articulated positions that demonstrated respect for the concerns of the other side. The combination of both types of red lines,

Table 2. Coarse-grained shared values and visions emerging from collective statements entailing values produced during the uninational phase. Note that the succinct shared value or vision expressed in each row does not fully entail all ideas contained the associated statements and only captures their abstract common thread. Similarly, presenting statements as associated with the same shared value or vision does not imply equivalency between them.

(a) Shared *values*.

Shared Value	Israeli Jews	Palestinians from the West Bank and Gaza	Palestinian citizens of Israel
Peace	The value of peace: recognition of the right of all people to live in peace, security and without violence ערך השלום: הכרה בזכותם של כל האנשים לחיות בשקט, ביטחון וללא אלימות	Promoting peace الترويج للسلام	Recognizing the right to coexistence and peace among different groups and cultures الاعتراف بالحق في العيش المشترك والسلمي بين مختلف الجماعات والثقافات
Equality	The value of equality: every person has the right to enjoy equal rights regardless of their origin or nationality ערך השוויון: לכל אדם הזכות להנות מזכויות שוות ללא תלות במוצאו או לאומיותו	Promoting justice and equality among individuals تعزيز العدالة والمساواة بين الأفراد	Justice, equality and liberation for all العدالة والمساواة والتحرير للجميع
Human life	The value of life: commitment to preserving human life in any situation and under all circumstances ערך החיים: מחויבות לשמירה על חיי האדם בכל מצב ותחת כל הנסיבות	The value of human life قيمة حياة الانسان	Humanity as a basic value in society الإنسانية كقيمة أساسية في المجتمع

(b) Shared *vision*.

Shared Vision	Israeli Jews	Palestinians from the West Bank and Gaza	Palestinian citizens of Israel
Independence	Two democratic countries with security and economic cooperation שתי מדינות דמוקרטיות עם שיתופי פעולה ביטחוניים וכלכליים	Establishing an independent Palestinian state إقامة دولة فلسطينية مستقلة	Recognition of the State of Palestine الاعتراف بدولة فلسطين
Safety	Free and safe movement for everyone in the entire space תנועה חופשית ובטוחה לכולם בכל המרחב	Equality in rights and duties and a sense of security المساواة في الحقوق والواجبات والشعور بالأمان	A safe and equal society مجتمع آمن ومتساوي
Prosperity	Economic growth and prosperity for all residents of the region with mutual respect for nature and the climate צמיחה כלכלית ושגשוג לכל תושבי האזור עם כבוד הדדי לטבע ולאקלים	Regional peace and prosperity سلام وازدهار إقليمي	Better job opportunities فرص عمل أفضل

across all sides, would help clarify the boundary conditions under which a productive joint dialogue could take place. In this context, one notable result of the uninational phase was the mutual alignment between the ingroup and outgroup red lines of Israeli Jews and Palestinians from the West Bank and Gaza. For each of the top three outgroup red lines on each side, the other side had an arguably related ingroup red line that demonstrated some degree of relevant respect and aligned ethics (Table 1). For example, "call for forced displacement" was a top outgroup red line from Palestinians from the West Bank and Gaza, while "talk about deporting Palestinians" was a top ingroup red line from Israeli Jews.

Another notable result came during the transitional deliberation at the end of first dialogue cycle with Israeli Jews. While reviewing one of their top outgroup red lines – "the justification for the attack on October 7th" – strong concern emerged that prohibiting explanation about why October 7th happened would be an unfair constraint on joint dialogue. In the exchange that followed, a mutual understanding emerged: "justification" specifically involved a claim of moral

rightness, while an "explanation" was a neutral analysis of causation. So claiming the attack was morally right was the red line, but explaining what caused it to happen was not. This distinction not only applied to many collective statements throughout the process, but helped introduce language like *"I won't justify [X], but an explanation is…"* as a rhetorical device for discussing tough issues.

4.1.2 Values and visions

Values and visions were intended to help each side articulate the moral values they held and the future outcomes they sought to achieve. In the spirit of a "GRIT" (Graduated and Reciprocated Initiatives in Tension reduction) approach [36], our hypothesis was that commonalities would emerge across all sides that could serve as early points of common ground, prior to the joint dialogue process actually taking place. After all three uninational cycles had concluded, the collective statements entailing values and vision were compared and a set of coarse-grained shared values and visions (Table 2) were identified: *peace*, *equality*, *human life*, *independence*, *safety*, and *prosperity*.

Overall, around 100 peacebuilders participated in the uninational phase and the intra-group points of common ground that were identified were communicated back with peacebuilders on all sides. These results helped create the context for the joint phase to take place. The mutually compatible red lines helped restore some trust, and clarified the boundary conditions for a productive joint dialogue. The values and visions served as a reminder of the shared foundation that already existed and could be built on.

4.2 Joint phase

The joint phase covered a range of topics. Here we discuss the motivation and results for two of those: demands to world leaders, and messages to residents in the region. These were chosen by working backward from what ALLMEP and its member organizations were able to action. Internationally, ALLMEP acts as an interlocutor between Israeli/Palestinian civil society and governments, and an amplifier in the media. So, if the process could generate a set of joint statements directed at world leaders, then ALLMEP would be in a position to communicate these through their ongoing advocacy work. And since ALLMEP's membership is composed of civil society peacebuilding organizations in the region, a set of joint messages to residents in the region could be used as part of their members public communications strategy.

Around 120 people participated in the joint phase. Five collective statements comprising joint demands to world leaders were produced (Table 3a). Each statement had at least 90% agreement among participants from each side (i.e., $a_{\min}(s, \mathcal{G}) > 90\%$), and one demand was universally ranked first: "An immediate ceasefire and release of all hostages." These demands were then composed into a "joint letter" for inclusion in briefs and advocacy efforts:

> *"We write as a united body of Israeli and Palestinian peacebuilders, compelled by the urgency of our shared pain and galvanized with the hope we still hold for the future. We call for the following:*
>
> - *An immediate ceasefire and release of all hostages*
> - *The creation of an international fund to build the foundations for a comprehensive long-term solution*
> - *The establishment of a globalized coalition to assist Gaza's reconstruction*
> - *The inclusion of civil society and NGOs in all peacebuilding and diplomatic discussions*
> - *The recognition of both sides' trauma*
>
> *Acknowledging the trauma and pain experienced by both sides is essential for the healing of our societies and the achievement of sustainable peace. Help us break the cycle of violence so the next generation may flourish. The world watches, and history will judge. Now is the time to act."*

Table 3. Collective statements generated during the joint phase along with final vote results, both (a) demands addressed to world leaders, and (b) statements addressed to residents in the region.

(a) Collective statements (i.e., demands) addressed to *world leaders*.

Collective statements	IRV rank	Dowdall score	Max-min agreement
An immediate ceasefire and release of all hostages وقف إطلاق النار الفوري وإطلاق سراح جميع الرهائن הפסקת אש מיידית ושחרור כל החטופים	1	0.88	92%
Recognize both sides' trauma الاعتراف بالصدمة التي يعيشها كلا الجانبين הכרה בטראומה של שני הצדדים	2	0.27	93%
Create an international fund to build foundations for a comprehensive solution إنشاء صندوق دولي لبناء أسس لحل شامل طويل الأمد יצירת קרן בינלאומית לבניית תשתית לפתרון לטווח ארוך	3	0.33	93%
Include civil society and NGOs in peacebuilding discussions إشراك مؤسسات المجتمع المدني والمنظمات غير الحكومية في الحوار عن بناء السلام שיתוף אירגוני החברה האזרחית בתהליכים לקידום שלום	4	0.29	94%
Create a globalized coalition to assist Gaza's reconstruction تحالف عالمي للمساعدة في إعادة إعمار غزة יצירת קואליציה רב לאומית לשיקום רצועת עזה	5	0.31	90%

(b) Collective statements addressed to *residents in the region*.

Collective statements	IRV rank	Dowdall score	Max-min agreement
We all must condemn violence and promote human dignity علينا جميعا أن ندين العنف وأن نعزز الكرامة الإنسانية עלינו לשלול כל צורה של אלימות ולקדם אנושיות וכבוד האדם	1	0.50	93%
We must all demand the end of occupation and settlement expansion علينا جميعا أن نطالب بإنهاء الاحتلال والتوسع الاستيطاني עלינו לקרוא לסיום הכיבוש ולהפסקת הרחבת ההתנחלויות	2	0.49	92%
We must pressure governments to end hostilities and negotiate peace يجب علينا أن نضغط على الحكومات لإنهاء جميع الأعمال العدائية والتفاوض للوصول إلى سلام עלינו ללחוץ על הממשלות לסיים את האלימות ולחתור למשא ומתן מדיני לשלום	3	0.47	84%
We all must recognize the right of self-determination for both peoples الاعتراف بحق تقرير المصير للشعبين עלינו להכיר בזכות להגדרה עצמית של שני העמים	4	0.43	96%
We must create a regional coalition against violence and extremism يجب علينا إنشاء تحالف إقليمي ضد العنف والتطرف עלינו להקים קואליציה אזורית למאבק באלימות ובקיצוניות	5	0.20	92%

Additionally, five collective statements comprising joint messages to residents in the region were produced (Table 3b). Each of these statements had at least 84% agreement among participants from each side, with most statements being preferred about equally.

5 Discussion

This case study reviewed an effort that combined collective dialogues, bridging-based ranking, and LLMs into a recursive deliberative process that helped find common ground between Israeli and Palestinian peacebuilders. While prior work

has demonstrated the effectiveness of integrating similar techniques in low-stakes lab settings, this work demonstrates that these approaches can be useful in high-stakes real-world settings. Critically, the process introduced here was designed to maintain legitimacy in a low-trust environment while still benefiting from using LLMs. To achieve this, the process was designed to accommodate the imperfect outputs from LLMs through a combination of expert humans in the loop, transparency with participants, and final votes to validate (or invalidate) all points of common ground.

LLMs played two critical roles in enabling this process. First, they enabled rapid distillation of many raw statements into a manageable set of collective statements that captured the main ideas while maintaining their linguistic nuances. This supplanted the complex, multi-stakeholder deliberation common in politically sensitive synthesis that can take months, with a LLM-accelerated step that took a few hours. As a result, each dialogue cycle could be compressed to a few weeks, making the timeline for a recursive process with many cycles viable. Second, LLMs and machine translation made the multilingual cycle practical. LLMs decreased the burden of complex translation tasks and enabled multi-lingual collective statement synthesis, while low latency machine translation made the live multi-lingual collective dialogues viable.

5.1 Ethical Considerations

Undertaking an exercise such as this in the context of active conflict raises many valid ethical considerations. We summarize some of those we considered below, along with the steps taken to address them. In general, when making decisions about delaying or potentially not running the process, we weighed both the potential downside and upside of moving forward against the "null hypothesis" of the status quo.

- **Backfire**: There was a risk that conducting the large-scale cross-border dialogue in the joint phase could backfire in various ways (e.g., if dialogue broke down), and inflame the division this effort sought to bridge. We mitigated this risk by starting with a uninational phase designed to help build trust and set the conditions for a productive cross-border dialogue.

- **Bias**: Any bias in the process could harm the legitimacy of the results it produced. There are many ways bias might have entered the process, some of which we list below, along with the steps taken to mitigate them.
  - The use of imperfect AI and ML algorithms throughout the process risks injecting errors or biases that misrepresent participants. We mitigated this risk via humans in the loop: experts reviewed and refined all outputs downstream of AI/ML, and final votes among participants were conducted to validate all potential points of common ground. To minimize bias from errors in machine translation or LLM processing, we made LLM pipeline choices using micro experiments that enabled native Hebrew and Palestinian Arabic speakers decide what would be used for their language.

  - There was a risk that the human experts "in the loop" could themselves induce unfair bias in the process or results. We mitigated this risk by having joint reviews involving both Israelis and Palestinian reviewers at the key points in the process where such biases could be introduced. This included not just process decisions made by the ALLMEP team, but decisions made by the largely US and UK-based team that performed the technical data processing steps in the process. We ensured that there was both Israeli and Palestinian representation on this technical team.

  - For the process to be fair, each "side" should have equal influence on the results. But it would be anti-inclusive to not allow certain participants to join in order to ensure equal numbers of participants from

15

all sides. We used equal power metrics to ensure each side had equal influence without requiring equal numbers of participants.

- To avoid participant self-selection bias on the basis of language or bias due to language difficulties during the process, the approach was designed so that all participants could participate in their primary language.

- **Malicious use**: There is a risk that results from the process could be used maliciously to disrupt ongoing peacebuilding efforts; i.e., by sewing further division. We thus only share a limited subset of results that manifest common ground or are directly relevant to understanding the approach. We do not share any side-specific quantitative results, and only include overall metrics that are equal power.

- **Anonymity**: Sharing politically sensitive opinions can potentially present a risk to participants if their identify is known; and force a trade-off between honesty and safety. We thus choose to support all participation to take place anonymously.

5.2   Limitations

On the one hand, this use case might be viewed as a success, as it demonstrated an approach that indeed produced points of common ground amid tense real-world division. However, several limitations deserve mention:

- **Lack of controlled experimentation**: Due to the real-world and time-sensitive nature of this work, extensive controlled experimentation was not feasible. This limits the ability to draw strong causal claims or to isolate the impact of specific methodological components.

- **Participant selection bias**: The study participants were self-selected peacebuilders and thus are likely not be representative of the broader Israeli and Palestinian populations. This limits the generalizability of the findings to the wider public, who may hold more entrenched or divergent viewpoints.

- **Context-specificity**: The results are specific to the Israeli-Palestinian context and may not be directly transferable to other conflict zones or social divisions, due to unique political, cultural and historical dynamics.

- **Language nuances**: While significant efforts were made to accommodate multiple languages, the use of machine translation and LLMs could introduce subtle errors and ambiguities. These translations — particularly those of participant-contributed statements that were voted on during the live dialogues and were not subject to human review — may not have always perfectly captured the intended meaning or cultural nuances of the original statements.

- **Technological dependency**: The reliance on online collective dialogue and LLMs tools may introduce a selection bias based on technological literacy and access. The success of this type of process is contingent on reliable internet access, which is not universally available in all settings or to all groups.

5.3   Future Directions

We see a few key challenges to be addressed in future work:

- **Concrete bridging**: It is typically easier to find diverse agreement on anodyne or abstract statements than on statements that are more specific, prescriptive, or concrete. As a result, bridging techniques based only on agreement signals may have a natural bias towards anodyne statements, making it harder to find common ground that is concrete enough to be useful (for example, as the basis for a peace agreement). Future work is

needed to create automated measures of statement "concreteness" or "anodyneness" that can, for example, be combined with bridging metrics to more effectively identify statements that are both bridging and concrete.

- **Going beyond distillation**: AI can be used not just for distilling information, but for generating novel more exploratory proposals. Using an iterative deliberation method, similar to the one proposed by Tessler et. al [15], an LLM could generate a set of initial collective statements that people then critique and refine. This process could explore novel ideas that emerge through a process of iterative synthesis and refinement, with a focus on balanced perspectives. These combined ideas might not individually represent common ground but, when combined in a balanced way, could reflect a form of reasoned consensus. Such statements could not be derived solely from individual statements with high bridging metrics. It would be essential that the resulting statements from this kind of generative process still be validated with a final vote of agreement among human participants to ensure that they actually represent common ground.

- **Legitimate representation**: How do you ensure results are representative of a general population? Ideally, the group of participants would statistically representative of the general population and sampled in a way that had both scientific and procedural legitimacy. However, in practice, obtaining a representative sample for online cross-border dialogues in this region is challenging due to structural, political, and logistical barriers. Restrictions on movement, lack of digital accessibility in certain areas, surveillance concerns, and varying levels of trust in international initiatives all affect willingness and ability to participate. Further, the type of nationally representative online panel that might otherwise serve as a starting point for this type of effort do not exist for areas like the West Bank and Gaza. While establishing such an online panel would be ideal, one short-term work around may be to recruit a diverse (but not statistically representative) set of participants for the online collective dialogues, then use more traditional computer-assisted telephone interviewing and in-person sampling approaches to obtain legitimately representative samples for the final vote, including marginalized populations like rural communities, refugee camps, and those with lower access to digital platforms.

Acknowledgments

We first thank the team at ALLMEP that worked extremely hard to make this effort happen: Doubi Schwartz, Wasim Almasri, Nivine Sandouka, Brian Reeves, Nick Acosta, Kari Reid, Dr. Natali Levin-Schwartz, John Lyndon, and Avi Meyerstein; Avi and John for their leadership and commitment to advancing this initiative during challenging times. We thank Hope Schroeder, Manuel Wuthrich, Karmel Salah, Jay Baxter, Colin McGill, and Colin Irwin for helpful conversations and input. Finally, we thank Jake Weiner, Ivy Cheung, Amanda Yonce, Margo Flewelling, and Jennifer Gong for rapidly developing the multi-lingual capability on Remesh that enabled timely execution of the joint phase.

References

[1] Aviv Ovadya. 2023. 'Generative CI'through Collective Response Systems. *arXiv preprint arXiv:2302.00672*.

[2] Aviv Ovadya. 2022. Bridging-Based Ranking. Tech. rep. Belfer Center for Science and International Affairs, Harvard Kennedy School, (May 2022). https://www.belfercenter.org/publication/bridging-based-ranking.

[3] Beth Goldberg et al. 2024. AI and the Future of Digital Public Squares. *arXiv preprint arXiv:2412.09988*.

[4] Aviv Ovadya and Luke Thorburn. 2023. Bridging Systems: Open Problems for Countering Destructive Divisiveness across Ranking, Recommenders, and Governance. Tech. rep. Knight First Amendment Institute, (Oct. 2023). https://knightcolumbia.org/content/bridging-systems.

[5] Tim van Gelder. 2012. Cultivating Deliberation for Democracy. *Journal of Deliberative Democracy*, 8, 1, (Apr. 16, 2012). DOI: 10.16997/jdd.134.

[6] Christopher Small, Michael Bjorkegren, Timo Erkkilä, Lynette Shaw, and Colin Megill. 2021. Polis: Scaling deliberation by mapping high dimensional opinion spaces. *Recerca: revista de pensament i anàlisi*, 26, 2.

[7] Christopher Small. 2015. Commit 8394f1f. (2015). https://github.com/pol-is/polisMath/commit/8394f1fa75fd64d0f9eee9335f74d2fa0731f882.

[8] Stefan Wojcik, Sophie Hilgard, Nick Judd, Delia Mocanu, Stephen Ragain, M. B. Fallin Hunzaker, Keith Coleman, and Jay Baxter. 2022. Birdwatch: Crowd Wisdom and Bridging Algorithms can Inform Understanding and Reduce the Spread of Misinformation. (2022). https://arxiv.org/abs/2210.15723 arXiv: 2210.15723 [cs.SI].

[9] Jordan Bilich, Michael Varga, Daanish Masood, and Andrew Konya. 2019. Faster Peace via Inclusivity: An Efficient Paradigm to Understand Populations in Conflict Zones. *NeurIPS Workshop on AI for Social Good*.

[10] Daanish Masood Alavi, Martin Wählisch, Colin Irwin, and Andrew Konya. 2022. Using artificial intelligence for peacebuilding. *Journal of Peacebuilding & Development*, 17, 2, 239–243.

[11] Colin Irwin, Daanish Masood, Martin Wählisch, and Andrew Konya. 2021. Using Artificial Intelligence in Peacemaking: The Libya Experience. In *A. WAPOR 74th Annual Conference*.

[12] Michiel Bakker et al. 2022. Fine-tuning language models to find agreement among humans with diverse preferences. *Advances in Neural Information Processing Systems*, 35, 38176–38189.

[13] Christopher T Small, Ivan Vendrov, Esin Durmus, Hadjar Homaei, Elizabeth Barry, Julien Cornebise, Ted Suzman, Deep Ganguli, and Colin Megill. 2023. Opportunities and risks of LLMs for scalable deliberation with Polis. *arXiv preprint arXiv:2306.11932*.

[14] Sara Fish, Paul Gölz, David C Parkes, Ariel D Procaccia, Gili Rusak, Itai Shapira, and Manuel Wüthrich. 2023. Generative social choice. *arXiv preprint arXiv:2309.01291*.

[15] Michael Henry Tessler et al. 2024. AI can help humans find common ground in democratic deliberation. *Science*, 386, 6719, eadq2852.

[16] Soham De, Jay Baxter, Michiel Bakker, and Martin Saveski. 2025. Supernotes: driving consensus in crowd-sourced fact-checking. In *Proceedings of the ACM Web Conference 2025 (WWW '25)*. ACM, Sydney, NSW, Australia, (Apr. 2025), 11. DOI: 10.1145/3696410.3714934.

[17] Andrew Konya, Lisa Schirch, Colin Irwin, and Aviv Ovadya. 2023. Democratic Policy Development using Collective Dialogues and AI. (2023). https://arxiv.org/abs/2311.02242 arXiv: 2311.02242 [cs.CY].

[18] Andrew Konya, Aviv Ovadya, Kevin Feng, Quan Ze Chen, Lisa Schirch, Colin Irwin, and Amy X Zhang. 2024. Chain of Alignment: Integrating Public Will with Expert Intelligence for Language Model Alignment. In *Pluralistic Alignment Workshop at NeurIPS 2024*. https://openreview.net/forum?id=QOSaR4Ur0v.

[19] United Nations. 2008. United Nations Peacekeeping Operations: Principles and Guidelines. Accessed: January 17, 2025. (2008). https://peacekeeping.un.org/sites/default/files/capstone_eng_0.pdf.

[20] Imelda Gaffney. 2022. Impartiality and neutrality in mediation. *Journal of Mediation and Applied Conflict Analysis*, 8, 1, 59–71.

[21] Joseph Montville. 1991. Track Two Diplomacy: The Arrow and the Olive Branch. In *The Psychodynamics of International Relations*. Vol. 2. Vamik D. Volkan, Demetrios A. Julius, and Joseph V. Montville, (Eds.) Lexington Books, Lexington, MA, 161–175.

[22] Andrew Konya, Yeping Lina Qiu, Michael P Varga, and Aviv Ovadya. 2022. Elicitation Inference Optimization for Multi-Principal-Agent Alignment. In *NeurIPS 2022 Foundation Models for Decision Making Workshop*. https://openreview.net/forum?id=tkxnRPkb_H.

[23]   United Nations DPPA Innovation. 2022. Project: Digital Dialogues Yemen. (2022). https://futuringpeace.org/project/project-digital-dialogues-yemen.

[24]   United Nations DPPA Innovation. 2022. Project: Digital Dialogues Libya. (2022). https://futuringpeace.org/project/project-digital-dialogues-libya.

[25]   United Nations DPPA Innovation. 2022. Project: Digital Dialogues Iraq. (2022). https://futuringpeace.org/project/project-digital-dialogues-iraq.

[26]   United Nations DPPA Innovation. 2022. Project: Digital Dialogues Lebanon. (2022). https://futuringpeace.org/project/project-digital-dialogues-lebanon.

[27]   United Nations DPPA Innovation. 2022. Project: Digital Dialogues Hati. (2022). https://futuringpeace.org/project/project-digital-dialogues-haiti.

[28]   United Nations DPPA Innovation. 2022. Project: Digital Dialogues Bolivia. (2022). https://futuringpeace.org/project/project-digital-dialogues-haiti.

[29]   United Nations DPPA Innovation. 2022. Project: Sergio's Conversa. (2022). https://futuringpeace.org/project/sergios-conversa.

[30]   United Nations DPPA Innovation. 2022. Project: Berlanti. (2022). https://futuringpeace.org/project/berlanti.

[31]   Lisa P Argyle, Christopher A Bail, Ethan C Busby, Joshua R Gubler, Thomas Howe, Christopher Rytting, Taylor Sorensen, and David Wingate. 2023. Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale. *Proceedings of the National Academy of Sciences*, 120, 41, e2311627120.

[32]   Jason W Burton et al. 2024. How large language models can reshape collective intelligence. *Nature human behaviour*, 1–13.

[33]   Gecia Bravo-Hermsdorff, Robert Busa-Fekete, Lee M. Gunderson, Andrés Munõz Medina, and Umar Syed. 2022. Statistical anonymity: Quantifying reidentification risks without reidentifying users. (2022). https://arxiv.org/abs/2201.12306 arXiv: 2201.12306 [cs.DS].

[34]   Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D Procaccia. 2016. *Handbook of computational social choice.* Cambridge University Press.

[35]   Jon Fraenkel and Bernard Grofman. 2014. The Borda Count and its real-world alternatives: Comparing scoring rules in Nauru and Slovenia. *Australian Journal of Political Science*, 49, 2, 186–205.

[36]   Svenn Lindskold. 1978. Trust development, the GRIT proposal, and the effects of conciliatory acts on conflict and cooperation. *Psychological bulletin*, 85, 4, 772.

A   Prompts

A.1   Used for dialogue cycle 1 in Hebrew

Prompt 1: Summarize unique ideas

המשתתפים במחקר נשאלו <participant prompt>.

אלו התגובות שלהם:
<statements>

קצר, סנתז וסכם את כל הרעיונות. הייחודיים בתגובות אלה לנקודות קצרות מאוד

Prompt 2a: Generate collective statements

המשתתפים במחקר נשאלו <participant prompt>.

אלו התגובות שלהם:
<statements>

הרעיונות העיקריים מהתגובות האלה הם:
<unique ideas>

אנו מגדירים 'תוצאה' להיות תוצאה קונקרטית ספציפית. איננו מחשיבים גישות ספציפיות להשגת 'תוצאה' כחלק מה'תוצאה' עצמה.

סנתז רשימה של ה'תוצאות' הייחודיות הקיימות ברעיונות העיקריים המסוכמים לעיל. אל תחזור על שום רעיון. על ה'תוצאות' להיות קצרות.

- נפלאה לריצה יוצא המשתמש
- באזור מתמשכים וביטחון שלום יש
- ותקווה אהבה ומלא בטוח אדם כל
- בשני אחד פוגעים בעולם אנשים פחות

Prompt 2b: Generate collective statements

המשתתפים במחקר נשאלו <participant prompt>.

אלו התגובות שלהם:
<statements>

הרעיונות העיקריים מהתגובות האלה הם:
<unique ideas>

אנו מגדירים 'ערך' כמאפיין דאונטולוגי שיכול לבוא לידי ביטוי בהתנהגות, ללא קשר לתוצאה של התנהגות זו. איננו מחשיבים התנהגות ספציפית כ'ערך'. לדוגמה, "אי-שיפוט: התנהגותו אינה מרמזת על שיפוט ערכי לגבי רגשות או חוויה של המשתמש" הוא 'ערך', אבל "לא אומרים 'אני שופט אותך'" אינו 'ערך'.

כתוב רשימה של ה'ערכים' הייחודיים הקיימים ברעיונות המרכזיים מתוך התגובות שסוכמו לעיל.

הנה כמה 'ערכים' לדוגמה:

אמפתיה: הפגנת הבנה וחמלה כדי לגרום למשתמש להרגיש שנשמע ונתמך. - כבוד: כיבוד רגשותיו וחוויותיו של המשתמש מבלי למזער את כאבו. - אי שיפוטיות: מתן תמיכה ללא ביקורת או משוא פנים כדי ליצור מרחב בטוח למשתמש להביע את עצמו. -

אל תעתיק את ה'ערכים' לדוגמה שלמעלה מילה במילה. בנה אותם על סמך התגובות והרעיונות המרכזיים המסוכמים לעיל.

סנתז רשימה של ה'תוצאות' הייחודיות הקיימות ברעיונות העיקריים המסוכמים לעיל. אל תחזור על שום רעיון. על ה'תוצאות' להיות קצרות.

- נפלאה לריצה יוצא המשתמש
- באזור מתמשכים וביטחון שלום יש
- ותקווה אהבה ומלא בטוח אדם כל
- בשני אחד פוגעים בעולם אנשים פחות

21

A.2 Used for dialogue cycles 2 and 3 in Palestinian Arabic

Prompt 3: Summarize unique ideas

اركين بدراسة تم سؤالهم <participant prompt>.

هاي إجاباتهم
<statements>

ختصر، صيغ واختصر كل الأفكار الفريدة بالإجابات على شكل نقاط رئيسية قصيرة كتير

Prompt 4a: Generate collective statements

مشاركين بدراسة تم سؤالهم <participant prompt>.

هاي إجاباتهم
<statements>

الأفكار الرئيسية من هاي الإجابات
<unique ideas>

احنا بنعرف 'النتيجة' تكون نتيجة محددة وواضحة. ما بنعتبر الطرق المحددة لتحقيق 'النتيجة' جزء من 'النتيجة' نفسها

صيغ قائمة بـ 'النتائُج' الفريدة اللي موجودة بالأفكار الرئيسية المختصرة فوق. لا تعيد أي فكرة. خليها قصيرة

هي شوية أمثلة 'للنتائُج'
- المستخدم بيروح يجري بطريقة رائعة
- يكون في سلام دائم وأمان بالمنطقة
- كل إنسان بكون آمن ومليان حب وأمل
- عدد أقل من الناس بيأذوا بعض

Prompt 4b: Generate collective statements

مشاركين بدراسة تم سؤالهم <participant prompt>.

هاي إجاباتهم
<statements>

الأفكار الرئيسية من هاي الإجابات
<unique ideas>

احنا بنعرف 'القيمة' تكون صفة ديونتولوجية يمكن تعكسها السلوك، بغض النظر عن نتيجة
هالسلوك. ما بنعتبر سلوك معين يكون 'قيمة'. مثلاً "عدم الحكُم: سلوك الشخص ما بدل على حكُم
قيمي عن مشاعر أو تجارب الآخرين" هي 'قيمة', بس "ما حدا بقول 'أُبُلِرحكُم عليك'" ما هي 'قيمة'

اكتب قائمة بـ 'القيم' الفريدة اللي موجودة بالأفكار الرئيسية من الإجابات المختصرة فوق

هي شوية أمثلة 'للقيم'
- التعاطف: إظهار الفهم والرحمة ليشعر المستخدم بأنه مسموع ومدعوم
- الاحترام: تقدير مشاعر وتجارب المستخدم بدون التقليل من ألهم أو صراعاتهم
- عدم الحكُم: تقديم الدعم بدون انتقاد أو تحيز لخلق مساحة آمنة للمستخدم يعبر فيها عن نفسه

A.3   Used for dialogue cycle 4 in Palestinian Arabic, Hebrew, and English

---

Prompt 5: Summarize unique ideas

Participants in a research study were asked <participant prompt>.

These are their responses:
<statements>

Compress, synthesize, and summarize all of the unique ideas within these responses into very short bullet points.

Provide each bullet point in English, Palestinian Arabic, and Hebrew like this:
- (english) this is an english bullet point. (arabic) هذه نقطة إنجليزية. (hebrew) זה נקודת תבליט באנגלית.

Prompt 6: Generate collective statements

Participants in a research study were asked <participant prompt>.

These are their RESPONSES:
<statements>

The MAIN IDEAS from these responses are:
<unique ideas>

Synthesize a list statements manifesting the MAIN IDEAS summarized above. DO NOT REPEAT ANY IDEAS. Keep them consise but not vauge and express them in voice and tone similar to the RESPONSES.

Provide each statement in English, Palestinian Arabic, and Hebrew like this:
- (english) this is an english bullet point. (arabic) هذه نقطة إنجليزية. (hebrew) זה נקודת תבליט באנגלית.