post society

Grassroots Depolarisation

Preliminary Design Requirements and Rationale

Author Max Kortlander

Initial development of this project received funding from Provincie Zuid Holland, in coordination with Kennis Zuid Holland and Leiden University's research track on depolarisation.

Content

1. Introduction

Current online platforms amplify division and hostility. They reward engagement over understanding, reaction over civility, and virality over reflection.

Efforts to fix these problems typically rely on after-the-fact interventions like content moderation, AI filtering, or deplatforming. While important, these tools are reactive – they don't address the underlying structure of how online conversations are designed.

To what extent can civility and depolarisation be embedded into the structure of digital conversations before moderation, AI, or algorithmic sorting are needed? Our research hypothesizes that healthier dialogue is possible by optimising the basic rules of interaction for civility and depolarisation:

- How conversations are structured
- How feedback is offered
- How groups are composed

As a part of the Grassroots Depolarisation research project this document presents general values-based assumptions (Section 2.1) and proposes technical requirements for the first prototyping iteration (Section 2.2). To arrive at this list of technical requirements, we:

- Reviewed established models of civil/depolarising conversation (Annex 1);
- Considered how functions in these existing models could be translated into technical requirements for an online space (Annex 1);
- Developed a longlist overview of potential technical requirements that contribute to civility and depolarisation (Annex 2);

Note that the analysis and subsequent design proposals in this paper are preliminary, exploratory, and non-final. In general, they are overly simplified for brevity – the goal is to arrive at an implementable design proposal for prototyping. Early prototyping roots our research in makership and drives an iterative and co-creative design process. We anticipate a return to the subjects of this document (existing conversational models, mechanisms, and design requirements) again when equipped with further technical progress and community insight.

2. Design Requirements

2.1 Platform Assumptions

General Assumptions

We assume the following regarding the digital civility mechanisms developed in the Grassroots Depolarisation project:

- Compatible with Mastodon, PubHubs and/or other ActivityPub protocols.
- Intended for public-interest communication and adoption of mechanisms by other platforms; rather than user growth and engagement optimization.
- Are not dependent on profiling or personal data.
- Source code is open and auditable.
- Infrastructure avoids big tech dependencies.

2.2 Consolidation: Design Requirements Shortlist (to implement for minimum viable product)

To begin building this prototype, our first iteration should start lightly. An initial prototype of the Grassroots Depolarisation platform may implement one or more of the following elements on a Mastodon instance or PubHubs Hub and observe the impact of their implementation:

- **Grounding template** in OP field (soft gate)
    - *Purpose:* Encourages discourse to begin from shared values or common concern.
    - *Mechanism:* Input prompt fields like “I think we agree on...”; “I think our common ground is…”; “I think we all care about…”.
    - *Relevance:* Centers the conversation around topics of shared interest (as opposed to conflicts).
- **Reflection suggested for first reply** (soft gate)
    - *Purpose:* Encourages participants to acknowledge or rephrase the previous post before responding critically.
    - *Mechanism:* Input prompt fields like ; ‘I heard you say…’; ‘If I understand correctly…’; ‘In summary…’
    - *Relevance:* Promotes active listening, slows reactive posting.
- **Civility feedback** on posts and comments (viewable only privately to author)
    - What it does: Allow personal reflection without social scoring (Encourages reflection, avoids performativity).
    - Mechanism: Optional, **non-public** feedback buttons or sliders per post, e.g.:
        i. “This taught me something.”
        ii. “This challenged my perspective.”
        iii. “Thank you for sharing this.”

Annex 1. Established (offline) models of civil, peaceful, depolarising conversation

There are existing models for structured conversation designed to optimise for values like depolarization, civility, and conflict resolution. This annex explores a number of those existing 'offline' models from various fields; considers their core principles; and presents a brief reflection on how a digital version of each model might work.

The analyses in this section are NOT design requirements themselves, but rather serve a creative and exploratory function to help us understand how existing mechanisms for civil conversation might be translated into design requirements. From here, as presented in section 2, we selected which design requirements to prototype.

Future research in this project (pending further funding) plans to return to the topic of existing (offline) models via a participatory research process involving experts from related fields.

Shuttle Diplomacy

Overview

- Used when two or more parties refuse to meet directly.
- Structure: A neutral mediator communicates back and forth between groups, delivering messages, proposals, and counter proposals.
- Principles: Indirect, controlled communication; cooling-off periods; private commitments before public statements.

Core Rules

- No direct public dialogue between opposing participants.
- All communication goes through a trusted intermediary.
- Messages can be summarized, reframed, or anonymized.

Example Online Implementation

In shuttle diplomacy, mediators move between two (or more) parties who aren't speaking directly. In a digital setting, this would mean that users can send posts **not to a public timeline**, but to a **private mediation space** where a facilitator sees and manages the conversation.

A prototype **"Shuttle"** instance could prevent direct public replies and introduce a mediator thread to coordinate. Consider three separate rooms: team red, team green, and team neutral. Two conversations occur first, one within team red and one within team green. Each has listening members of team neutral in the conversation. Then, team neutral meets in a third room to share messages, trade proposals, and perhaps even try to hash out a plan themselves.

Example features include:

- **Phase-gated conversation structure**
    - Various phases of the conversation – the private mediation space, the neutral meeting space, and any further phases of the conversation – are separate from and dependent on one another

- **Mediated Message Threads**:
    - Create a "mediation room" where posts by Party A are only visible to the mediator, not to Party B, and vice versa.

- **Mediator Dashboard**:
    - Intermediaries can read all inputs, summarize them, and post *synthesized summaries* publicly or semi-publicly.

World Café

Overview

- Utilises rotating small group conversations around specific questions. Insights are synthesised and shared after each round.
- Used in community visioning, municipal policy, social innovation.
-
- Principles: Informality, inclusion, idea cross-pollination, collective sensemaking.

Example Online Implementation

Core functions of an online world café:

- **Slowly grow the size of the discussion groups**. First, a discussion has many small groups, and a rotating member in each group synthesises the main ideas at the end of their discussion. These smaller groups join together to make increasingly fewer, larger groups. Ideas are synthesized at the end of each round.
    - Expand group size after each round as the discussion progresses (e.g. a group of 64 people - 16 groups of 4, then 8 groups of 8, then 4 groups of 16, then 2 groups of 32, then one group of 64).
- **Synthesis stewardship and writing** – Each new group includes at least two “synthesis stewards” who bring forward key insights from prior rounds; and at least two “synthesis writers” who lead the writing synthesis at the end of the session.
- **Timebox** discussion and synthesis in each discussion round.

Example technical tools:

- Pinned synthesis posts: allow 2+ stewards to pin collaborative summaries.
- Simple text input templates: prompt for opening posts and synthesis posts.
- Thread lifecycle management: auto-delete or anonymize threads after round completion.

Nonviolent Communication (NVC)

Overview
Nonviolent communication is based around four components:

- **Observation** – What actually happened? (No judgment)
- **Feelings** – How do I feel about it?
- **Needs** – What universal human need is behind this feeling?
- **Request** – What specific, do-able request can I make?

Example Online Implementation

- **Structured Online Post (OP) Composer** – OP composer has 4 fields:
  - Observation: free text, or prompted by "I observe that…"
  - Feeling: drop-down (sad, angry, hopeful, curious, etc.)
  - Need: drop-down (safety, understanding, inclusion, etc.)
  - Request: free text, or prompted by "I request that…"

- **OFNR Tags for Replies**:
  - First replies must be *empathetic reflections* using NVC core rules (especially paraphrasing feelings/needs).
    - When a user clicks "Reply" on a post tagged with `#NVCThread`, they are presented with a **pre-filled template** in the reply input:

      ```
      [Observation] What I notice is...

      [Feeling] I feel…

      [Need] Because I need…

      [Request] Would you be willing to...?
      ```

    - These tag brackets can either be:
      - Simple labels above the text box (non-enforced), or
      - Four required input fields before you can post (enforced).

    - Posts with this structure are visually formatted with icons or markers for each component, making it easy for other users to recognize that the post is intended as a good-faith, empathic response.

- **Reflection Gate**:
  - Thread design where you must acknowledge and reflect before you can critique.
  - For threads marked as `#ReflectFirst`, **replies must begin with a short paraphrase** of the original post.

    - A lightweight system-enforced prompt could be used like:
      "In your own words, what is the poster saying?"
      (50 character minimum before the 'Continue' button is enabled)
    - Once that reflection field is filled, the second field appears:
      "Now share your response, idea, or perspective."
    - The reflection part is posted first in a different font or indentation style,making the user's acknowledgement visually clear.

- **Civility Feedback**:
    - Replace likes with "I felt heard", "This addressed my need", "This helped me understand".

Socratic Dialogue

Overview

- Question-led inquiry, shared curiosity, no rush to judgment.
- Useful for: Exploring complexity rather than resolving disputes.
- Centered on open-ended, clarifying questions rather than assertions.
- Emphasizes shared definition-seeking and cooperative truth exploration.
- Fosters **humility**, **curiosity**, and **reflection**.
- Participants are encouraged to **paraphrase or summarize others' contributions** before responding, to confirm understanding and signal listening.

Example Digital Implementation

Features:

- a **question-first structure**,
- enforced **paraphrasing before contribution**, and
- staged discussion progression toward synthesis.
- **Thread Initiation Requires a Question**:
    - OP field must be in question form (verified via syntax or tagged with e.g. #SocraticStart).
    - Optional field: "What concept are we investigating?"

Technical elements:

- **Phase-Gated Reply System**:
    - **Phase 1 (Inquiry)**:
        - Only follow-up questions allowed. Response composer limited to interrogative syntax or via prompt.
    - **Phase 2 (Clarification/Reflection)**:
        - Responses must begin with a **paraphrase** (pre-filled template: "What I understand you to be saying is…").
        - Post only unlocked once a paraphrasing field is completed.

    - **Phase 3 (Tentative Synthesis)**:
        - Contributions may attempt to build common ground or offer shared definitions.

- **Feedback Tools**:
    - “Refined my thinking”
    - “Deepened the question”
    - “Accurate restatement of ideas”

Annex 2. Design Requirements

We gathered design requirements from the above list (Annex 1), as well as those identified through ideation and research over the course of the project, and developed the following longlist of **civility mechanisms** that might optimise for civility prior to the use of algorithmic, AI, or human intervention. (*Why not algs/AI/moderation? Those things are all necessary to research, too, but this project focuses on the fundamental basics of how online conversation is structured).*

Annex 2.1 Civility Mechanisms – Long List (design input, not to implement)

Response limitations

- Limit number of responses per user (e.g. per thread or per time window)
- Limit response length per user (already done on Mastadon)
- Temporarily limit responses for highly active users/groups to rebalance airtime
- Cap total number of posts about a single view/topic to avoid flooding or crowding
- Introduce waiting periods before re-posting (delayed posting or "cooling-off")

Conversation structure and phasing

- Incorporate structured conversation phases (e.g. Grounding → Reflection → Exploration → Synthesis)
- Start conversations with an agreement or common ground
- Provide pre-built formats for starting discussions, e.g., “I’ve been thinking about...”, “I’m wondering how others experience...”, “A concern I want to understand better is...”
- Incorporate turn-taking mechanics (e.g. encouraging people to respond to a different person each round, or enforcing waiting until others have ‘spoken’)
- Time boxing and pacing (e.g. limited discussion phases; breather period between posting or phases; auto-expiry and closure of threads)
- Incorporate reflection requirements, e.g. through Socratic paraphrasing
- Post reflection prompts e.g. “Did this conversation change your perspective?”; “What would you tell someone who hasn’t read this thread?”; or “What new question do you leave with?”
- Incorporate tagging of threads, response types

Group size

- Require a minimum group size
- Limit maximum group size

Feedback tools

- Reconsider up/down voting (e.g. ‘like’ buttons with buttons or sliders like ‘this is nuanced’ or ‘this adds to my understanding’)

- Remove public reactions, scoring, accumulation and reputation

Timeboxing
- Delayed posting
- Implement necessary waiting time before responding

Humanisation
- Demonstrate ‘humanness’, e.g. through prompts to share creative outputs or anecdotes
- Strike a balance between anonymity, identification, and transparency (e.g. use ABCs to validate someone is from a particular country without having them provide any additional info.)
- Incorporate offline components (e.g. meetups) to complement online conversation

Organisation, visualisation, and presentation
- Group similar responses
- Visually indicate post and/or comment type
- Incorporate a means to synthesise discussion (adopt from existing tech e.g. Polis).

Group composition and dynamics
- Require minimum group size
- Limit maximum group size
- Group based on shared attribute (e.g. city of residence) or opinion.

Note: Portions of this report were developed in conjunction with AI.