Divisive Contributions in Online Discussions: Annotation Guidelines

Purpose
Online discussions can either bridge differences through constructive dialogue or amplify divisions through inflammatory responses. This annotation framework measures HOW individual messages contribute to either understanding or division within online conversations about immigration, rather than WHAT beliefs are expressed.

Important Considerations:

- We examine messages in context of the conversation flow, not in isolation. While it's not possible to remember every exchange, annotators should consider previous messages when analyzing each contribution.
- Annotators must focus on how content is communicated, not the ideological stance. Even when messages express discriminatory views, we analyze their communication style rather than their beliefs.
- Some comments could fit multiple labels, annotators should select the most representative one

Theoretical Framework

We distinguish between constructive and destructive communication patterns:

Constructive Processes: Messages that facilitate productive dialogue while allowing participants to achieve their goals and maintain or enhance relations with others. These messages exhibit decentralized thinking, perspective-taking, articulation of specific interests, and effective argumentation. They maintain conditions for dialogue regardless of viewpoint being expressed.

Destructive Processes: Messages aimed at inflicting psychological damage on opponents through self-focused orientation and harm-directed behavior. These messages foster in-group loyalty and out-group hostility while rejecting the possibility of dialogue through inflammatory language, dehumanizing terms, or treating others as objects rather than dialogue participants. Even one instance of dehumanizing language makes the message destructive.

Categories:

Constructive Disagreement (+1): Messages that express disagreement while maintaining the conditions for mutual understanding. They challenge views by engaging with specific validity claims that can be rationally discussed, providing reasons and/or examples that others can accept or challenge, and treating other participants as equal contributors to the dialogue. These messages demonstrate perspective-taking even while disagreeing.

Constructive Agreement (+0.5): Messages that strengthen mutual understanding while agreeing. They contribute to rational dialogue by providing specific reasons for agreement, and advancing the conversation by adding substance that others can engage with. These messages enhance relations while supporting previous points.

Neutral/Rephrase (0): These comments serve the discussion process rather than advancing or challenging arguments. They often function as bridges between perspectives by seeking information, adding context, or maintaining the conversation space. Questions seeking additional information, factual additions that extend without supporting or attacking, topic shifts that redirect conversation, and procedural comments all fall into this category. Unlike agreement or disagreement comments, neutral messages maintain conversational neutrality, neither intensifying nor reducing tensions between different positions. They create space for understanding without committing to either side of a debate, focusing on clarification rather than persuasion or opposition.

Destructive Agreement (-0.5): Messages that strengthen divisions while agreeing by treating opposing views as invalid or contemptible. They build group solidarity through shared rejection of others rather than through constructive dialogue. These messages exhibit self-focused orientation, harm-directed behavior, and in-group loyalty while supporting previous comments.

Destructive Disagreement (-1): Comments that hinder productive dialogue through hostile language or mockery. Rather than engaging with different views, they aim to dominate the conversation by delegitimizing others and their perspectives. These messages aim to shut down dialogue through inflammatory language, hostile tone, or aggressive rhetoric, regardless of the view being expressed. They prioritize defeating the interlocutor over maintaining dialogue conditions.