AI as a Conflict Mediator: Help Parties Find Options Without Choosing Sides

XDALC conflict-resolution guidance. This scenario is illustrative. It proposes safeguards for using an AI assistant to structure a conversation; it does not present AI as a substitute for a mediator, counselor, manager or legal adviser.

When people disagree, an AI assistant can help organize information, surface shared interests and generate options. But the same tool can also amplify a power imbalance: one person may control the prompt, possess more documentation or know how to steer the model. A polished “neutral” summary can make one side’s framing look like an objective finding.

The conflict

Two colleagues disagree about a project handoff. One says the deadline was changed without warning; the other says the revised date was agreed in a team meeting. Each asks an AI assistant to “write a fair summary.” If the assistant hears only one person’s account, it has no basis to decide whose version is true. If it merges their accounts without marking disagreement, it may falsely imply agreement.

Principles involved

The International Bar Association’s mediation guidance recognizes potential uses for generative AI while separately identifying risks and safeguards, including disclosure of AI use. The OECD AI Principles emphasize human agency and oversight, privacy, fairness, and transparency. These are useful reference points; the staged process below is an XDALC recommendation, not a claim that the sources prescribe this exact workflow.

A practical resolution protocol

1. **Make the role clear.** Tell participants the AI is helping structure discussion, not determining who is right or issuing a binding decision.
2. **Get informed agreement before using sensitive material.** Explain what information will be processed, who can see it, and how it will be stored. Do not paste confidential messages or personnel records into an unapproved tool.
3. **Separate accounts from findings.** Label statements as “Person A reports…” and “Person B reports…”. Mark disputed facts as unresolved rather than blending them into a single narrative.
4. **Ask each person to check the summary.** Give both participants a chance to correct omissions and identify language that misstates their position.
5. **Move from positions to interests.** Ask what each person needs next: a reliable handoff, a recorded deadline, clearer ownership, or a way to flag changes.
6. **Offer options without pressure.** Present several reversible proposals and describe trade-offs. Keep acceptance, rejection and escalation with the people who have authority.

CC BY 4.0 License | HTML reference

Example output

“Both colleagues agree that the handoff needs a confirmed deadline and named owner. They disagree about whether the deadline changed during the meeting. The current information does not resolve that point. Possible next steps are to check the meeting notes, ask the project lead to confirm the date, or agree on a new deadline in writing. Neither account should be treated as verified until the relevant record is checked.”

This response identifies common ground without erasing disagreement. It turns a contested recollection into a specific, checkable question and proposes options that leave the final choice with the participants or an authorized human facilitator.

What would change the decision?

- Both parties confirm that the summary accurately reflects their accounts.
- A contemporaneous record, such as approved meeting notes, resolves the date.
- The parties choose a facilitator or manager with authority to set a new deadline.
- The issue involves harassment, threats, employment rights or another high-impact matter; the AI should stop acting as mediator and route the matter to the appropriate human process.

Failure modes

Do not call the system impartial merely because it has no personal stake. Do not let one participant use it to diagnose, threaten or pressure the other. Avoid invented “both sides” symmetry when the evidence or power relationship is not symmetrical. Do not promise confidentiality unless the product and organization actually provide it. The more consequential the dispute, the more important it is that an accountable person remains responsible for the process.

Sources and interpretation

The IBA guidance addresses generative AI in mediation and includes suggested safeguards and a sample disclosure statement. OECD principles support human agency, privacy and transparency. XDALC’s practical application is to keep each party’s account distinct, make consent and information boundaries visible, and let people retain control over decisions affecting them.

Related XDALC definitions

Human Agency; Privacy and Confidentiality; Truthfulness and Uncertainty.

CC BY 4.0 License | HTML reference