Appendix

Supplementary Materials for Counterspeech or Disagreement? Comparing Oppositional Replies in China's AI Controversies on Social Media

A1. Framework and Examples of LLM Annotation Instructions

Component	Shared coding guidance	Task-specific adaptation
Task	The instructions defined the coding objective, unit of analysis, and relation between the two comments in each parent–response dyad.	**Stance annotation:** The model assessed the response comment in relation to the position expressed in its parent comment. **Response-type annotation:** The model assessed the primary form of opposition expressed in the response comment.
Input	Each coding instance contained the complete parent comment and its direct response, with the two texts presented as a linked dyad.	
Labels	The available labels and their definitions.	**Stance annotation:** Opposition, Support, and Neutral. **Response-type annotation:** Personal Attack, Affiliation Accusation, Contradiction, Justification, Request, and Counterclaim.
Contextual interpretation	The prompt first guided the model to summarize the parent comment's core claim, evaluation, and target and to identify its stance, before assessing the response against this contextual basis.	
Decision criteria	Coding criteria specified the information relevant to classification and the basis for selecting among candidate labels.	**Stance annotation:** The model considered the direction of the response toward the parent comment's central claim, judgment, or attitude and assigned one stance label. **Response-type annotation:** The model considered the principal target and argumentative operation of the response and assigned one primary type, together with one secondary type when another substantive operation was also present.
Boundary conditions	Additional rules addressed ambiguous cases arising from surface wording or category overlap.	**Stance annotation:** e.g., a response that only corrects a peripheral detail without addressing the parent comment's core stance is coded as neutral. **Response-type annotation:** e.g., A question is coded as Request (C3) when it seeks information, clarification, or evidence that could materially advance the dispute; when the interrogative form serves only to express rejection instead of soliciting an answer, it is coded as Contradiction (C1).

Few-shot guidance	Each example included the parent comment, response comment, assigned label, and classification rationale.	**Stance annotation:** e.g., Parent: “Domestic LLMs are all hype.” Response: “Yeah, yeah—you pounding away at a keyboard counts as real technology.” Label: Opposition. Rationale: The response conveys opposition through exaggerated sarcasm despite appearing to agree. **Response-type annotation:** e.g., Parent: “DeepSeek got such a simple question wrong. It’s not worth using at all.” Response: “Ordinary people use AI to save time and money, not to watch it win a math Olympiad.” Label: Counterclaim. Rationale: The response shifts the evaluative standard from test performance to everyday usefulness rather than providing evidence about the disputed performance.
Output format	The model was required to return a predefined code and a brief justification in a standardized format, facilitating automated extraction and subsequent analysis.	**Stance annotation:** The model returned one categorical stance label. **Response-type annotation:** The model returned one primary response type and, where applicable, one secondary response type.

Note. During pilot coding, boundary conditions and few-shot examples were iteratively adjusted based on comparisons between preliminary model outputs and human coding, with particular attention to typical sources of coding inconsistency.