

Modeling Human Label Variation in Natural Language Inference: From Human to LLM and Chain-of-Thought Explanations



UNIVERSITY OF
CAMBRIDGE

Beiduo Chen *ELLIS supervisor* Barbara Plank \blacktriangle *co-supervisor* Anna Korhonen \blacklozenge

\blacktriangle MaiNLP, CIS, LMU Munich & MCML, Germany

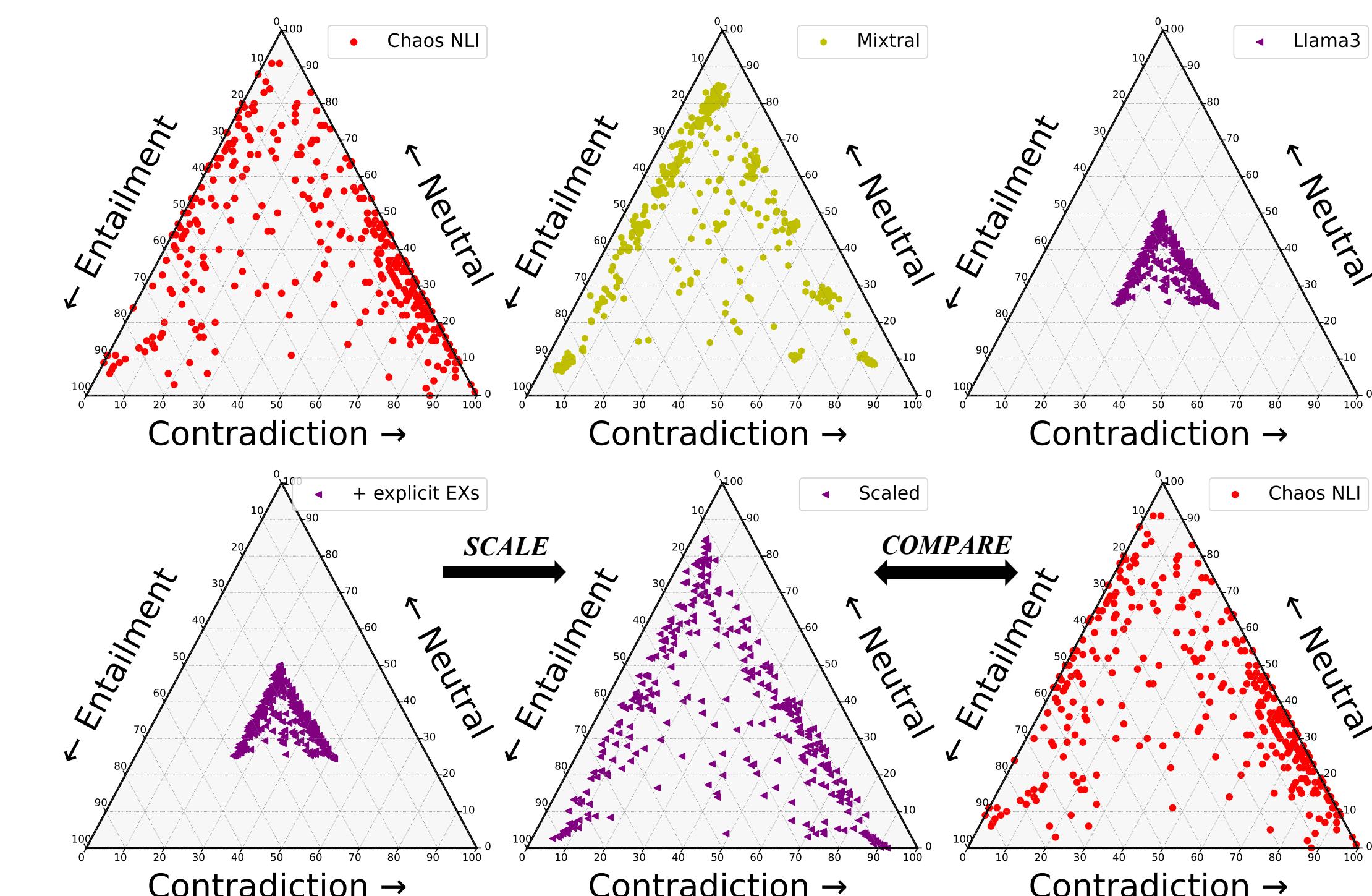
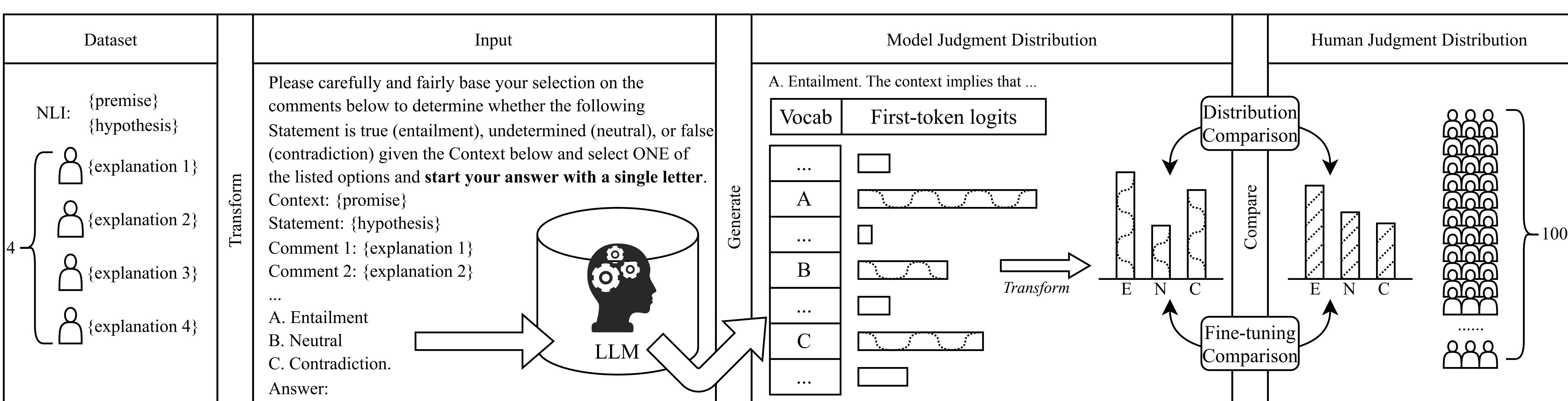
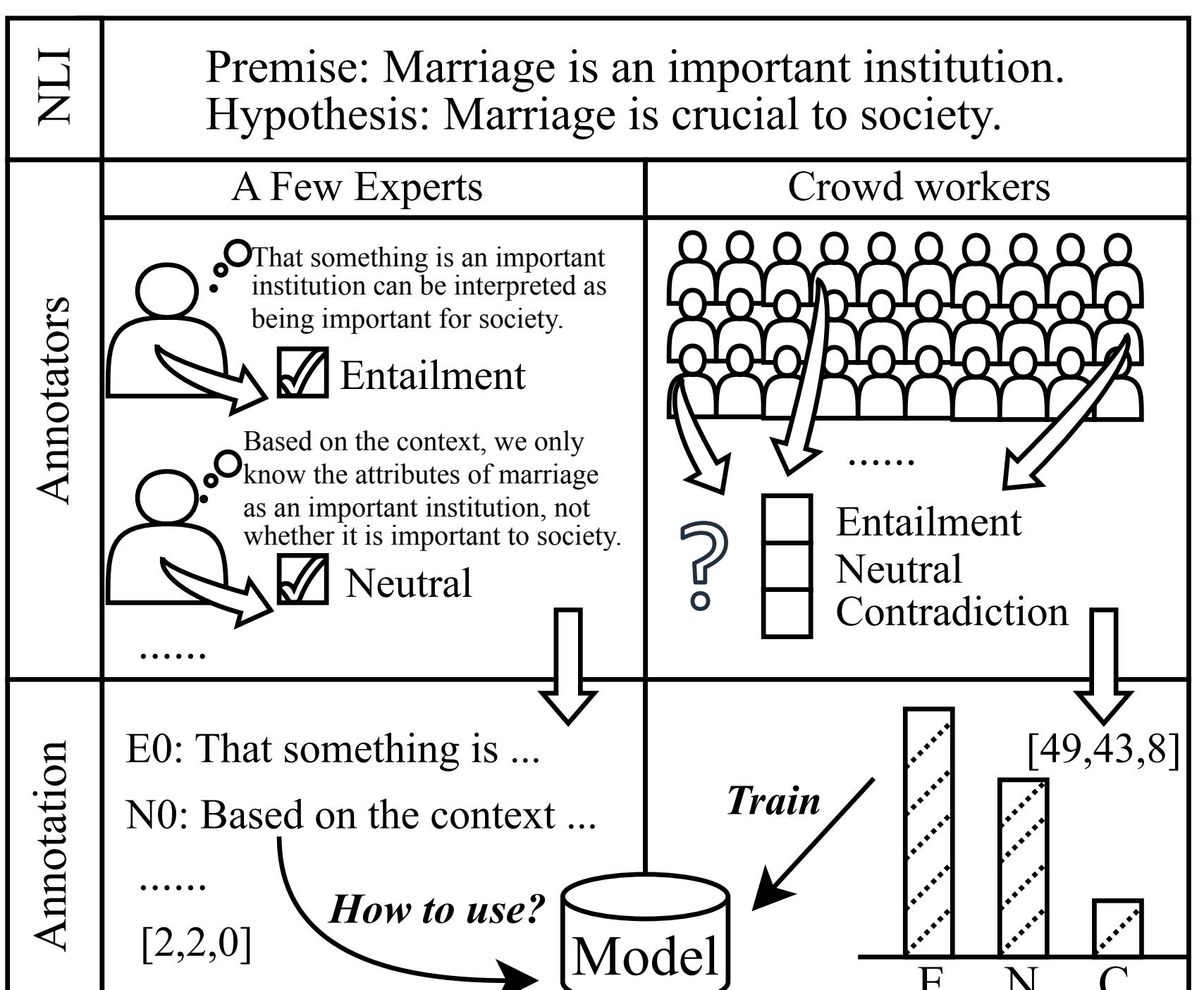
\blacklozenge Language Technology Lab, University of Cambridge, United Kingdom



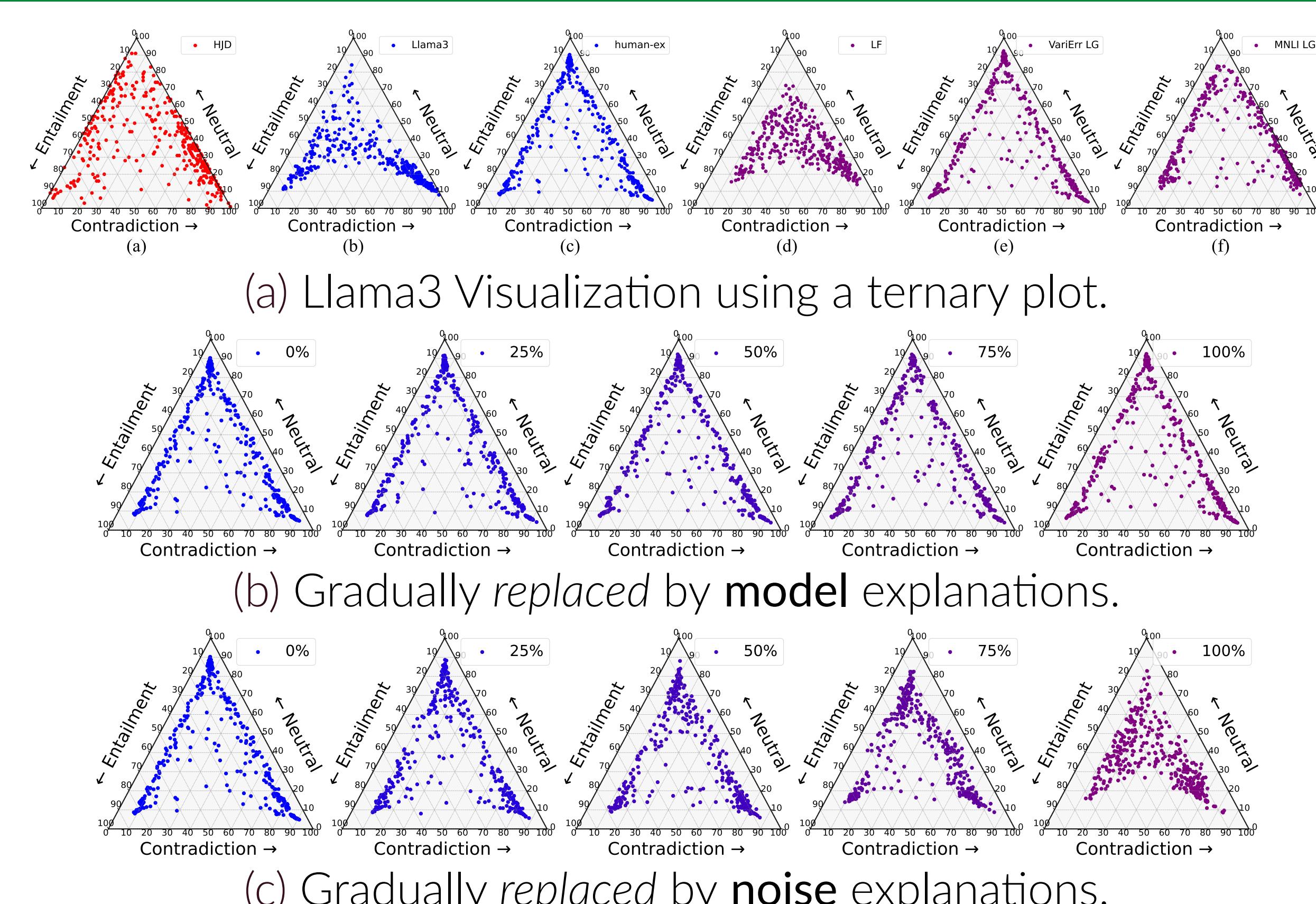
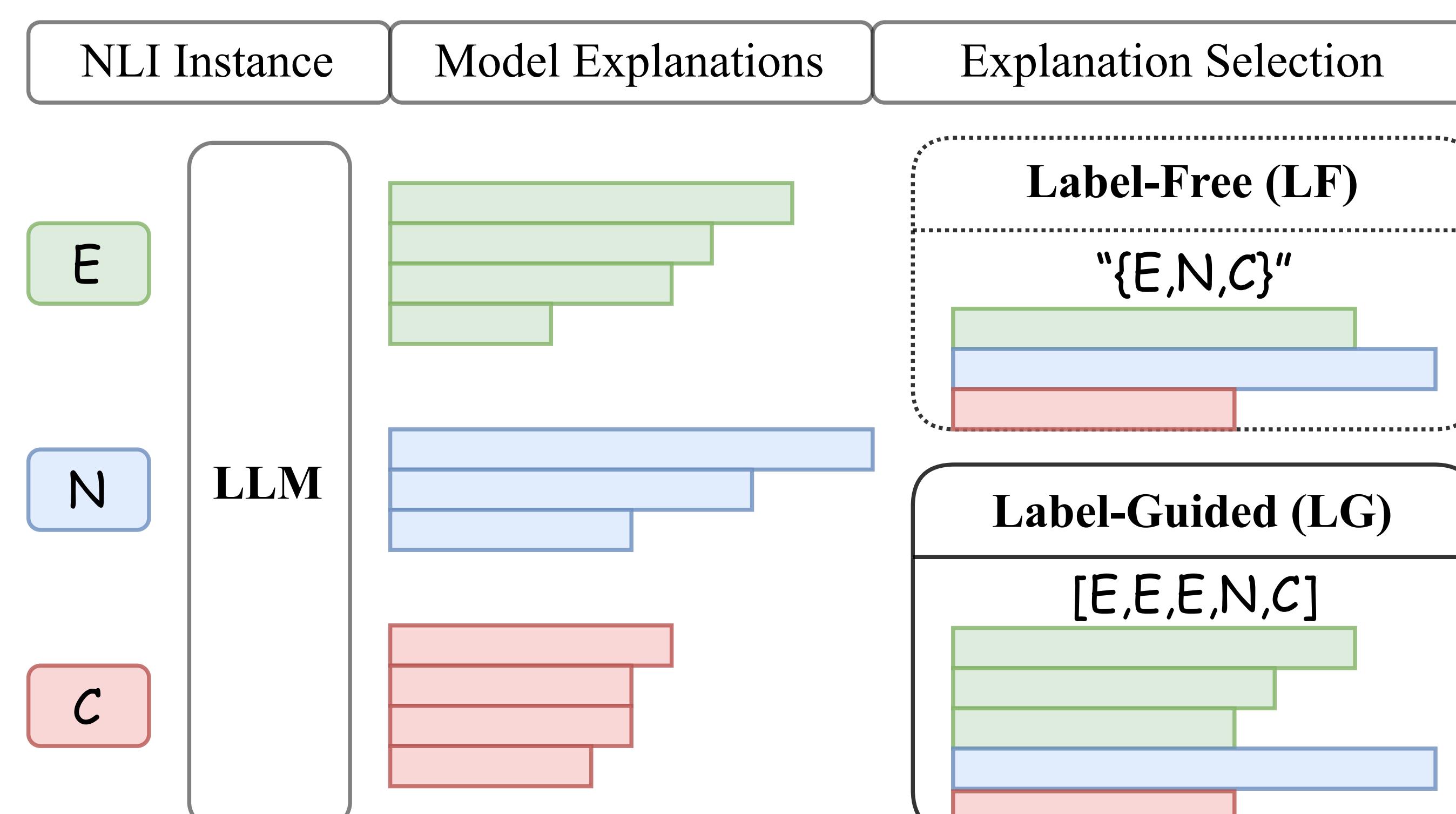
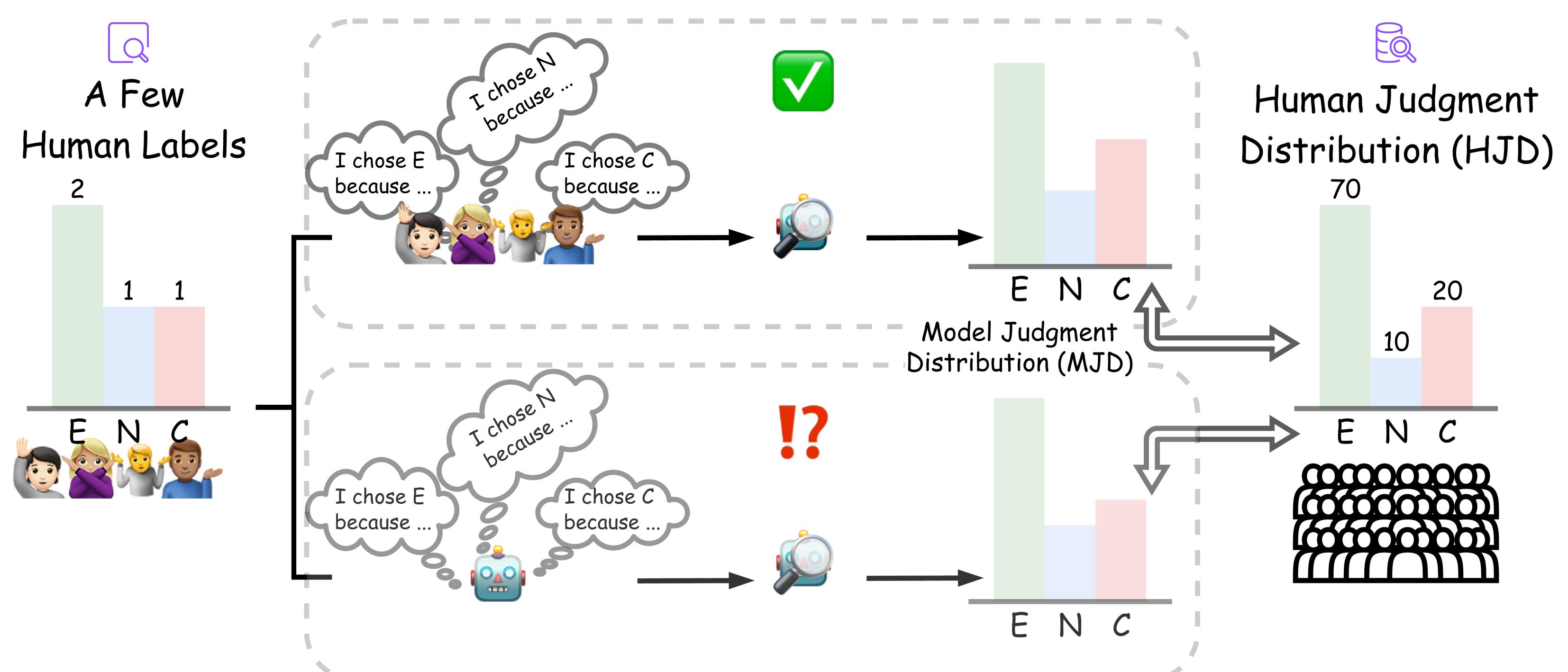
MCML
Munich Center for Machine Learning

Language
Technology
Lab

"Seeing the Big through the Small": Can LLMs Approximate Human Judgment Distributions on NLI from a Few Explanations? (EMNLP 2024)



A Rose by Any Other Name: LLM-Generated Explanations Are Good Proxies for Human Explanations to Collect Label Distributions on NLI (ACL 2025)



Threading the Needle: Reweaving Chain-of-Thought Reasoning to Explain Human Label Variation (EMNLP 2025)

Input: Please read the following social scenario and the accompanying question, choose the most appropriate answer from the options provided and start your answer with a single letter. Scenario: Ash redeemed themselves after retaking the test they failed. Question: How will Ash feel as a result?
A. relieved B. accomplished C. proud Answer:

CoTs: Okay, let's see. The scenario is that Ash redeemed themselves by retaking a test they failed. The question is asking how Ash will feel as a result. The options are A. relieved, B. accomplished, and C. proud.

Reasoning Base: First, I need to understand what each of these emotions means in this context. Redeeming oneself after failing a test suggests that Ash probably worked hard to retake it and succeeded this time. So, the feelings here would be related to overcoming a challenge.

Discourse Units: "Relieved"(A) would mean that Ash is feeling a release from the stress or worry about having failed before. Passing the test might take away that anxiety.

EL filter: "Accomplished"(B) implies a sense of achievement, having completed something successfully, especially after effort. Since Ash had to retake the test, putting in the work to pass it now would make them feel like they've accomplished something.

Proud(C): is more about self-esteem or satisfaction in one's own achievements. It's a bit similar to accomplished but more focused on the personal pride aspect.

Now, considering the scenario: Ash didn't just pass a test; they redeemed themselves. That suggests that failing initially might have caused some shame or disappointment, so passing now would alleviate that. But redemption often involves not just relief but also a sense of achievement. So both A and B are possible. But "proud" might be more about the internal sense of pride, whereas "accomplished" is about the feeling of having achieved something through effort. Which one fits better here? If Ash worked hard to retake and pass, then feeling accomplished makes sense. Relieved is possible, but maybe the primary feeling is accomplishment. Proud could also be there, but maybe B is more directly related to the effort and success.

I think the best answer is B. Accomplished.

