Section 1: Intro

- Task: Natural Language Inference (§1.1)
- From Human Label Variation (§1.2) to Within-label Variation (§1.3)

Section 2: Related Work

- NLI dataset landscape (§2.1)
- Explaining NLI labels (§2.2)
- Background in HLV (§2.3)
- Taxonomies of variation in NLI (§2.4)
- Explanation generation with LLM (§2.5)

RQ1: Can a linguistically grounded taxonomy capture and structure the diversity of human reasoning underlying the same NLI label?

RQ2: Can the taxonomy improve the interpretability and alignment of explanations generated by large language models?

RQ3: Is the taxonomy applicable across diverse NLI datasets with different explanation styles and annotation protocols?

Section 3: LiTEx

Taxonomy Design (LiTEx)

- Linguistically grounded reasoning categories (§3.1)
- **Two main categories:** Text-Based Reasoning (§3.1.1) and World-Knowledge Reasoning (§3.1.2)
- Human annotation with LiTEx (§3.2)

Validation + Analysis

- IAA validation: Classification IAA (§3.3.1) and Highlights IAA (§3.3.2)
- Model classification experiments (§3.4)
- Annotation results analysis (§3.5)

Section 4: Generation

LLM Explanation Generation

- Baseline (§4.1), Highlight-guided (§4.2), Taxonomy-guided (§4.3) prompting
- Similarity assessment against human-written explanations (§4.4)
- Semantic coverage via t-SNE visualization (§4.5)
- Human validation (§4.6)

Section 5: Generalizability

Cross-Dataset Evaluation

- Two additional datasets: LIVENLI and VariErr NLI (§5.1)
- LiTEx on LIVENLI (§5.2) and VariErr NLI (§5.3)
- Similarity across categoryies and labels (§5.4)

Section 6: Discussion

- **LiTEx captures diverse reasoning strategies**, revealing meaningful within-label variation in NLI.
- **Human-interpretable and model-learnable**, enabling reproducible annotation and effective classification.
- **Enhances explanation generation**, producing more human-like and semantically rich outputs.
- **Generalizes across datasets**, supporting robust analysis and future research.