

How to Measure Decoding Quality

Quality Metrics

Diversity Metrics

- Perplexity (Distinct-n-grams)
- BLEU score (lower = more diverse)
- ROUGE score (summarization rate)
- Human evaluation for quality, richness
- Grammaticality, scope of outputs

Combined Quality-Diversity Metrics

- Task success rate (does it solve the task?)
- Human preference (A/B testing)
- Pareto frontier analysis (multi-objective)

