Human Evaluation: Best Selection
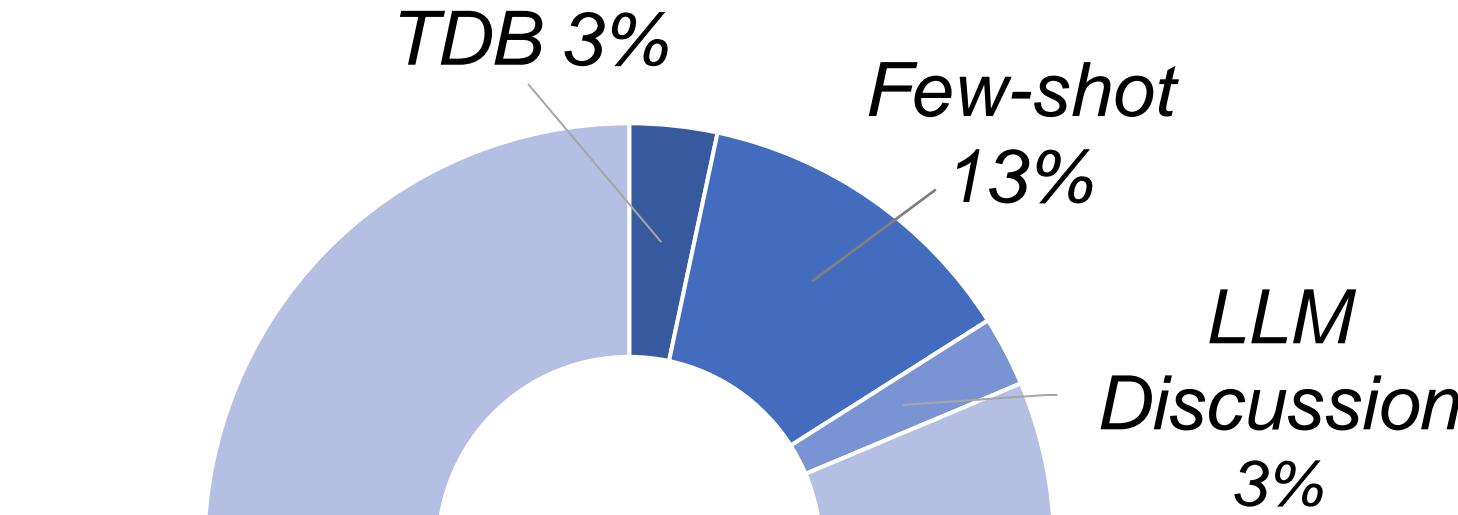

TDB 3%

Few-shot 13%

LLM Discussion 3%

BabyNamer 81%