Perplexity vs Quantity of Gold Standard Fine-tuning Data

GPT-2 5M English and Synthetic Training Examples

Legend:
- Cebuano pretrained on english5m
- Frisian pretrained on english5m
- Frisian pretrained on synthetic frisian
- Occitan pretrained on english5m
- Occitan pretrained on synthetic occitan

Y-axis: Perplexity

X-axis: Fine-tune Data Quantity (1000s of sentences)