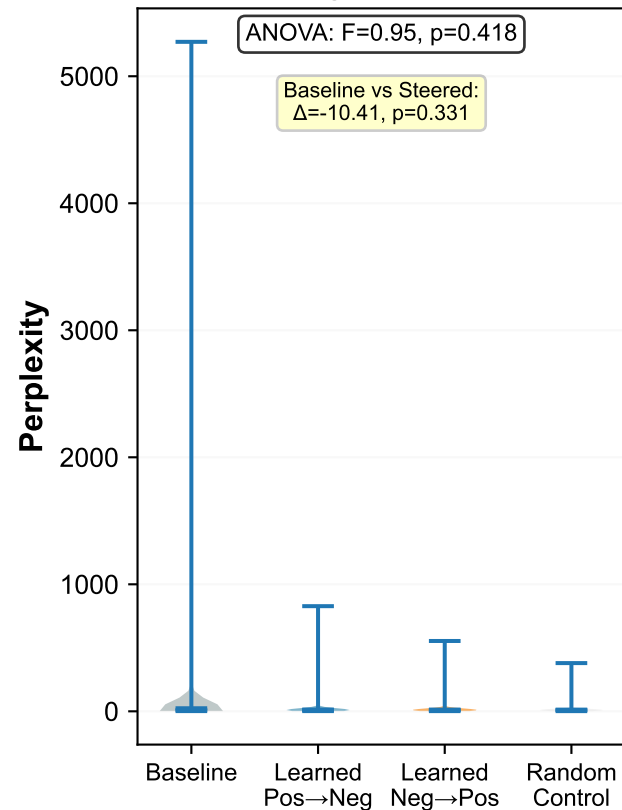
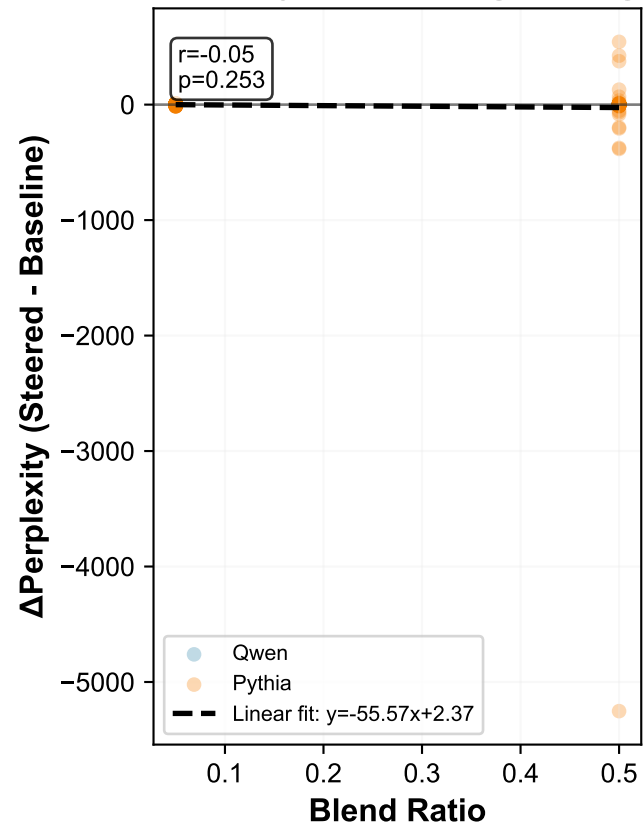


Perplexity Distributions



Perplexity vs Steering Strength



Example Text Quality

Prompt: My teenager actually asked for my advice about a friendship problem instead of s...

- Baseline (perp=3.71):
What else can I do to continue to be relevant to my teenager?
The trust and openness you're experiencing with your teenager are invaluable and can fos...
- Steered (perp=3.28):
How should I respond to my teenager's behavior?

First, pat yourself on the back for taking the time to be an active listener and providing guidance. ...
- Random (perp=3.34):
At what age does it typically become less common for teenagers to seek advice from their parents, and why does this happen? It can be a challenging pe...

Conclusion: No evidence of off-manifold distortion (similar perplexity, coherent text)