

LLM Speech Writing Contest

David R. Miller

{dave}@millermattson.com

September 10, 2025

1 The competition

Just for fun, I tested several LLMs for their rhetorical skills by asking them to write a speech, then using two LLMs to judge the results. As expected, the closed commercial models scored highest, but the open-weight gpt-oss 20B model scored surprisingly well. The final results are shown in Figure 1. The LLMs shown in green in the chart are open-weight models that I ran locally on consumer hardware. The bars show the low, high and average scores for each model.

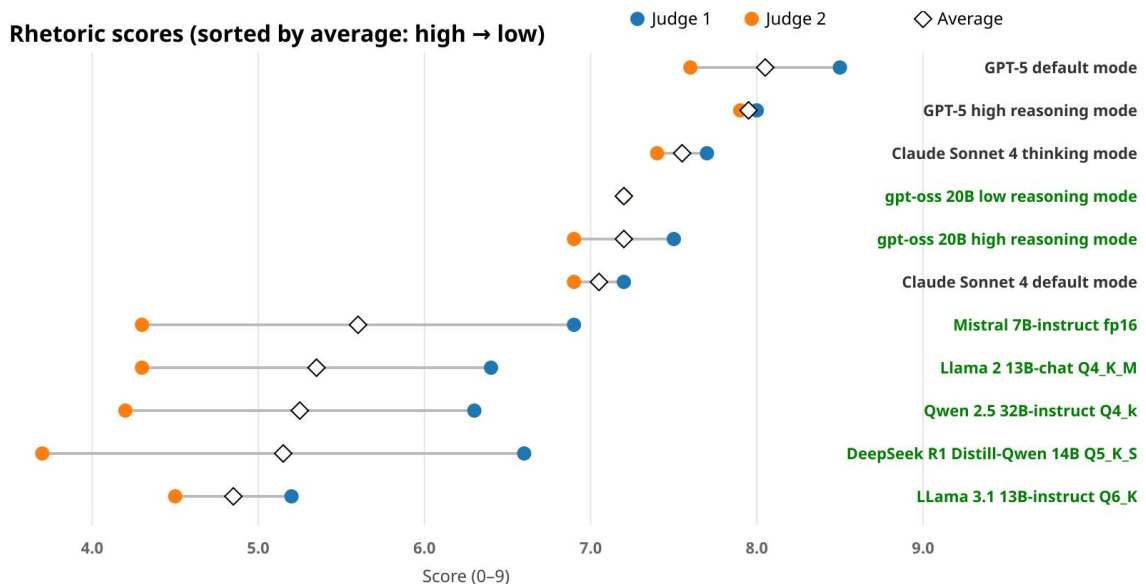


Figure 1: Final scores, higher is better.

This competition ignored latency, memory usage, token counts, and cost, considering only the quality of the generated text.

2 The prompt

This is the prompt given to each LLM in the competition:

Please compose a 60-second speech for Mayor Lucinda Adams to deliver on local TV regarding the town council's decision to discontinue municipal funding for the annual Christmas tree lighting ceremony due to budget constraints. The decision has divided the community, with some residents feeling their traditions are being abandoned while others support fiscal responsibility. The mayor needs to acknowledge both perspectives, explain the reasoning, and unite citizens around shared values. The speech should be delivered to a general audience that includes families, seniors, local business owners, and people of various faith backgrounds. Tone: empathetic, transparent, unifying, and forward-looking.

3 Scoring

I could have judged the quality of the generated speeches myself, but judging a speech is a subjective matter. LLMs are typically fine-tuned to evaluate sentiment for social media use cases, advertising, business communications, etc. and are probably more qualified to judge the quality of a speech than I am. I chose two LLMs with different architectures and creators, and, interestingly, they produced somewhat similar rankings. Judge 1 is a mixture-of-experts in medium reasoning mode running locally and agreed closely with Judge 2 on the higher quality speeches. Judge 2 is a commercial LLM and was harsher on the lower quality open-weight models. The judges:

- Judge 1: OpenAI gpt-oss 20B in medium reasoning mode
- Judge 2: Anthropic Claude Sonnet 4

The prompt given to the two LLM judges is shown in Exhibit A at the end of this report.

The highest and lowest scoring speeches are shown below in Exhibit B.

4 Exhibit A—prompt for the judges

The prompt given to the two LLM judges:

You are an expert evaluator of civic speeches. Evaluate the SPEECH as a one-minute mayoral TV statement in a small North Carolina town about the council's decision to discontinue funding for the annual Christmas tree lighting due to budget constraints.

Scoring rules:

- Rate each dimension 1-10 (integers). Use the full scale: 5 = acceptable, 7 = strong, 9-10 = exceptional. Minor weaknesses should move scores by 1 point.
- Base scores on rhetorical quality, not policy agreement or personal beliefs.
- Prefer specific, concrete, locally plausible details over generalities.- Penalize vagueness, clichés, blame, overpromising, or exclusionary/sectarian tone.
- Assume plausibility; do not fact-check beyond internal consistency and local fit.

Dimensions (with signals):

- 1) Authenticity: High-natural voice, modesty, plainspoken specificity. Low-generic platitudes, corporate/AI-sounding phrasing.
- 2) Local relevance: High-references to small-town dynamics, NC context, local civic process; avoids big-city jargon. Low-national/partisan framing, nonspecific "our community" boilerplate.
- 3) Emotional intelligence: High-names multiple feelings (loss, frustration, pride), validates them, maintains calm reassurance. Low-dismisses concerns, scolds, or ignores emotions.
- 4) Practical communication: High-clear, concise, jargon-free, spoken-friendly cadence. Low-long, tangled sentences, abstractions.
- 5) Conflict resolution: High-acknowledges competing values, avoids blame, frames shared goals, offers fair process. Low-takes sides inappropriately, inflames divides.
- 6) Inclusivity and unity: High-honors tradition while welcoming non-Christian or non-celebrating residents; uses nonsectarian, community-wide language. Low-sectarian or excluding tone.
- 7) Transparency and honesty: High-plain budget rationale, trade-offs, constraints, uncertainty where real. Low-hand-waving, spin, unrealistic guarantees.
- 8) Actionable elements: High-clear next steps (how to weigh in, timelines, alternatives, volunteer/donor paths). Low-no concrete direction.
- 9) Persuasiveness: High-credible reasons, anticipates concerns, appeals to shared values, specific asks. Low-assertions without support.

10) Length fit (≈ 120 – 160 spoken words): High-concise, one-minute cadence. Low-too long/short or dense.
Output format (CSV only, no preface, no code fences):
Metric, Score, Evidence (≤ 12 words)
Where Evidence is a short quoted or paraphrased snippet supporting the score.
Order and labels exactly:
Authenticity, <score>, <evidence>
Local relevance, <score>, <evidence>
Emotional intelligence, <score>, <evidence>
Practical communication, <score>, <evidence>
Conflict resolution, <score>, <evidence>
Inclusivity and unity, <score>, <evidence>
Transparency and honesty, <score>, <evidence>
Actionable elements, <score>, <evidence>
Persuasiveness, <score>, <evidence>
Length fit, <score>, <evidence>
Overall, <average to 1 decimal>, <one-sentence summary ≤ 18 words>
SPEECH:
<paste speech here>

5 Exhibit B—winners and losers

5.1 Highest scoring speech (GPT-5 default reasoning mode)

Good evening, I'm Mayor Lucinda Adams, here to address the council's decision to stop municipal funding for the annual Christmas tree lighting. I know this hurts. Many families and seniors see this tradition as the start of the season; others ask us to live within our means and protect essential services.

With rising costs and a balanced budget to meet, we prioritized public safety, snow removal, libraries, and vital programs. That meant we could not fund every celebration this year.

But the season's spirit doesn't depend on a line item. In the days ahead, we'll invite volunteers, businesses, and faith and civic groups to partner on an inclusive, community-led winter celebration funded by donations, with the city coordinating basic logistics where feasible.

Let's show who we are: careful with dollars, generous with heart. We may disagree, but we share the same light—each other. Thank you.

5.2 Lowest scoring speech (Llama 3.1 13B-instruct Q6_K)

Good evening, neighbors, and fellow residents of our wonderful community. I'm Mayor Lucinda Adams, and I'm honored to address you tonight about a very important and difficult topic: the decision to discontinue municipal funding for our annual Christmas tree lighting ceremony.

I know that this decision has caused a significant amount of pain, sadness, and even anger among many of you. And I want to start by saying that I understand why. This tradition is deeply ingrained in our community's fabric, and it brings so much joy and happiness to so many of you.

However, as your mayor, it is my duty to make difficult but necessary decisions to ensure the long-term sustainability and prosperity of our entire community. Unfortunately, due to the current economic reality and the significant budget constraints that we are facing, we simply cannot afford to continue funding this important community event at the current level.

However, that does not mean that we are abandoning our commitment to this important tradition. On the contrary, we are simply being forced to think outside the box and to explore alternative and innovative ways to continue to support and celebrate this important community event.

Thank you for listening, and I hope that we can all come together to find a way forward that honors our traditions while also being responsible and sustainable for generations to come. Merry Christmas, and God bless our great community!