Evaluating LLM Workshop Projects

Evaluation Criteria

- Explainability
- Creativity
- Correctness
- Efficiency

Explainability

Why it matters:

- In high-stakes or collaborative settings, users need to **understand how and why** a model makes decisions.

Do:

- Use visualizations or plain-language summaries
- Show how the model arrived at outputs
- Mention any tools used (e.g., SHAP, LIME, prompt transparency)

Don’t:

- Say “it just works” without explanation
- Rely on black-box behaviours without interpretability considerations

Creativity

Why it matters:

- LLMs open doors for new ways to **solve problems, present ideas, or interact** with data.

Do:

- Think outside the box—use LLMs for novel or unexpected tasks
- Combine techniques or modalities creatively
- Tailor prompts or agents to new contexts

Don’t:

- Copy existing solutions without adaptation
- Stick to a trivial “Hello world” demo unless it offers something new

Correctness

Why it matters:

- LLMs are powerful—but they're also prone to **hallucinations**. A good application must ground its output in truth.

Do:

- Validate model outputs with references or logic
- Handle edge cases and show robustness
- Clarify what the model **can't** reliably do

Don't:

- Present unverifiable or false outputs
- Ignore failure modes or limitations

Efficiency (Runtime / Energy Use)

Why it matters:

- Sustainable, scalable AI is **computationally aware**. Runtime and energy use are practical constraints.

Do:

- Optimize for smaller models when appropriate
- Use batching, caching, or smart prompt engineering
- Consider cost/benefit tradeoffs of your setup

Don't:

- Burn resources unnecessarily
- Rely on overkill models (e.g., GPT-4 for simple tasks) when a smaller one suffices

How Much Energy Does ChatGPT Use Per Prompt?

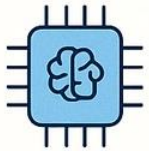
A Look at Its Hidden Environmental Costs

ELECTRICITY CONSUMPTION

Each ChatGPT prompt consumes approximately 0.001–0.01 kilowatt-hours (kWh) of electricity.

WATER USAGE

Each ChatGPT prompt can require 2 to 5 liters of water for server cooling.



COMPUTATIONAL RESOURCES

Using ChatGPT in a web browser typically uses 100–300 MB of RAM per prompt

ENVIRONMENTAL IMPACT

With millions of users, ChatGPTs daily energy consumption can reach 621.4 megawatt-hours (MWh).

@nilanth

Final Thoughts

We’re not just looking for the most advanced project—we’re looking for the most **thoughtful** one.

The best presentations will show:

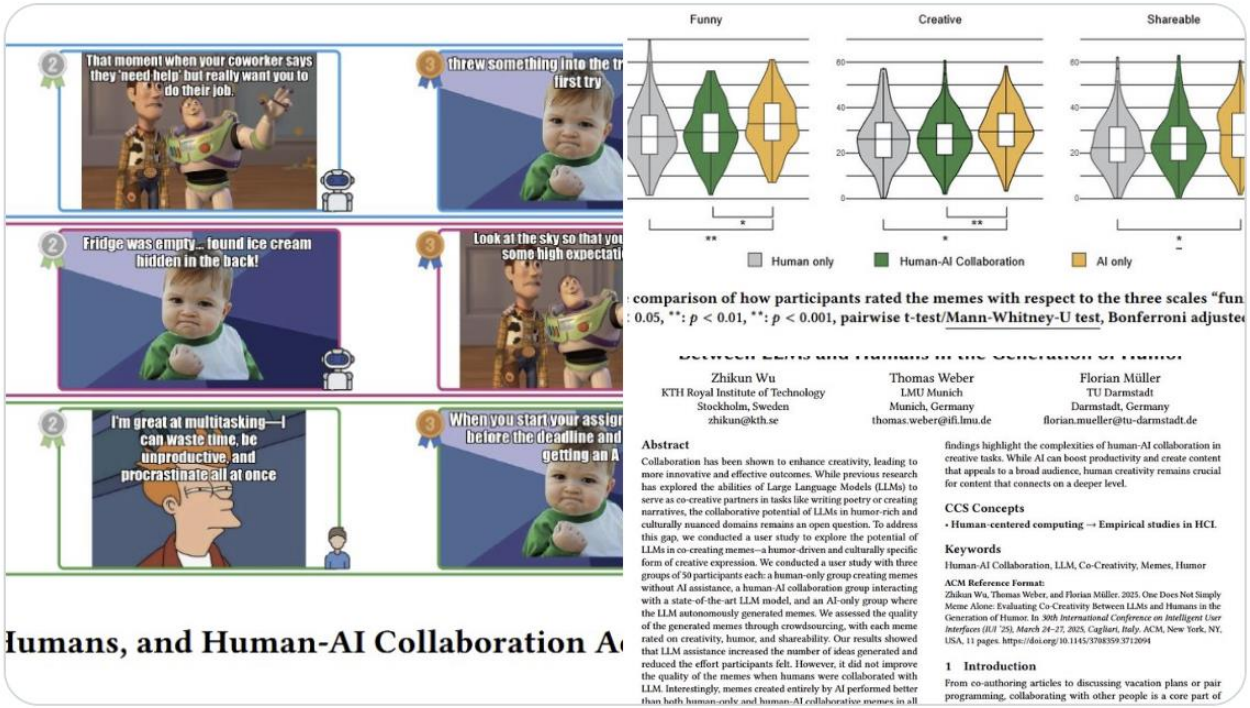
- **Clarity in communication**
- **Awareness of AI limitations**
- **Original thinking with practical grounding**

Have FUN!

Ethan Mollick
@emollick

I regret to announce that the meme Turing Test has been passed.

LLMs produce funnier memes than the average human, as judged by humans. Humans working with AI get no boost (a finding that is coming up often in AI-creativity work) The best human memers still beat AI, however.



8:33 PM · Mar 16, 2025 · **141.4K** Views