Hello Colloquium

A demo presentation with math, code, and figures

Nathan Lambert

2026-02-22

What is Colloquium?

- Agent-native slide creation tool for research talks
- Markdown-based and git-friendly
- AI agents can drive it programmatically
- Single self-contained HTML output

LaTeX Math Support

The loss function for RLHF with a KL penalty:

$$\mathcal{L}(\theta) = -\mathbb{E}_{x \sim D}\left[\log \sigma\left(r_\theta(x_w) - r_\theta(x_l)\right)\right]$$

Inline math works too: $\nabla_\theta J(\theta) = \mathbb{E}_{\pi_\theta}[R \cdot \nabla_\theta \log \pi_\theta(a|s)]$

Code Highlighting

```
from colloquium import Deck

deck = Deck(title="My Talk", author="Researcher")
deck.add_title_slide(subtitle="A research presentation")
deck.add_slide(
    title="Key Results",
    content="Our method achieves **state-of-the-art** performance.",
)
deck.build("output/")
```

Two Equal Columns

- Point one
- Point two
- Point three

- Result A: **95.2%**
- Result B: **87.4%**
- Result C: **91.8%**

Asymmetric Columns (60/40)

This wider column has the main explanation text. The 60/40 split gives more room to the primary content while keeping a sidebar for supplementary info.

- Key finding one
- Key finding two

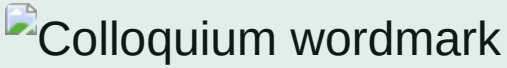
> *“Supplementary details go in the narrower column.”*

Supporting points:

1. Evidence A
2. Evidence B

Centered Image

Image with Text



Colloquium wordmark

Colloquium supports images in any layout. Here the wordmark sits in the wider column alongside explanatory text.

Images auto-scale to fit their container while maintaining aspect ratio.

Key Results

Experimental Results

Model	Accuracy	F1 Score	Training Time
Baseline	82.1%	79.3%	2h
Ours (small)	91.4%	89.7%	4h
Ours (large)	**95.2%**	**93.8%**	12h

> *“The results demonstrate significant improvements across all metrics.”*

Centered & Large Text

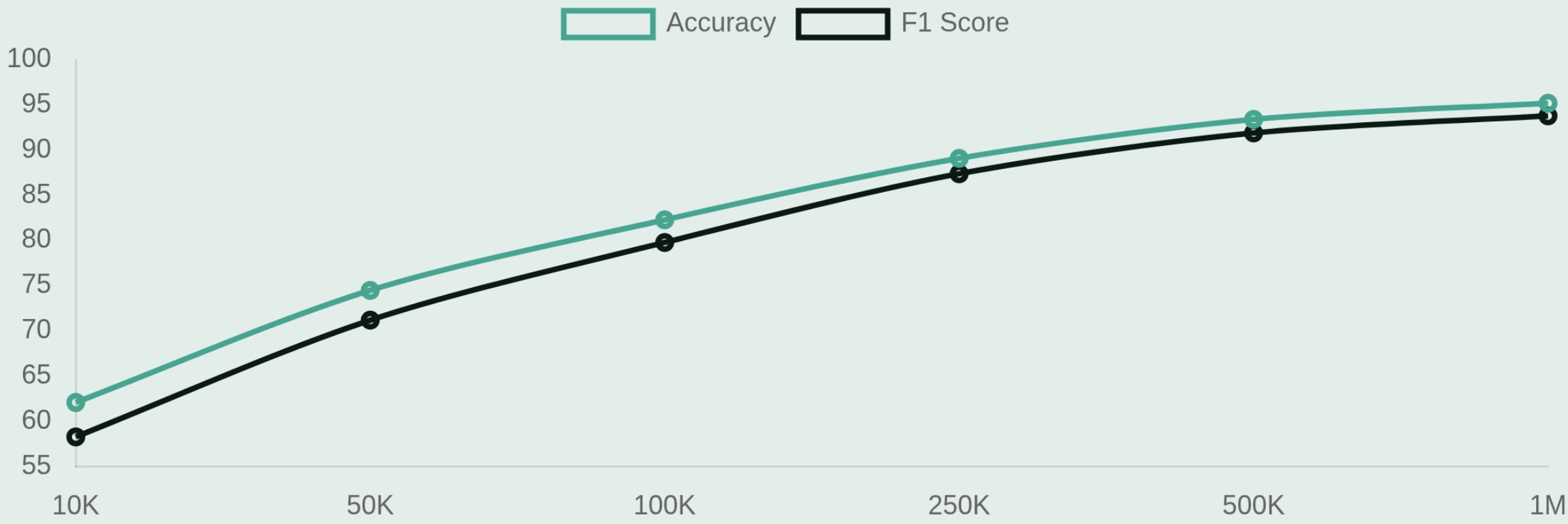
This slide uses the `align` and `size` directives to center all text and increase the font size.

Great for emphasis slides.

Vertically Centered Title

All content on this slide is vertically centered, like a title slide but with `##` heading style.

Training Performance



Resource Usage

Text Sizes

text-4xl **text-3xl** **text-2xl**

text-xl — Key takeaways

text-lg — Callouts and introductions

text-base — Default body text

text-sm — Dense lists, supporting details

text-xs — Footnotes, references, fine print

LLM Conversation

USER

What is RLHF?

ASSISTANT

RLHF (Reinforcement Learning from Human Feedback) is a technique for aligning language models with human preferences using reward models trained on human comparisons.

Conversation in Columns

- RLHF uses human preferences to train reward models
- The reward model scores LLM outputs
- PPO optimizes the policy against the reward

SYSTEM

You are a helpful AI research assistant.

USER

What is RLHF?

ASSISTANT

RLHF is a technique for aligning language models with human preferences.

Multi-Turn Conversation

USER

Can you explain the RLHF training pipeline?

ASSISTANT

The RLHF pipeline has three main steps:

1. **SFT** — supervised fine-tuning on demonstrations
2. **Reward modeling** — train a reward model on human preferences
3. **PPO** — optimize the policy against the reward model

USER

What's the role of KL divergence?

ASSISTANT

The KL penalty prevents the policy from diverging too far from the SFT model. Without it, the model can exploit the reward model with degenerate outputs — this is called *reward hacking*.

Callout Boxes

Accent

High-contrast callout for key takeaways.

- Great for punchy emphasis
- Uses the deck accent color

Muted

Softer card for supporting notes.

- Good for side explanations
- Uses the code/surface background

Surface

Neutral bordered panel for references or caveats.

- Works when you want less visual weight
- Keeps strong contrast with body text

Title-only Box

Core idea

Key RLHF Papers

The foundational work on RLHF (Christiano et al., 2017) introduced learning reward models from human comparisons.

InstructGPT (Ouyang et al., 2022) scaled this approach to large language models, demonstrating significant alignment improvements.

For a comprehensive overview, see (Lambert, 2024).

Christiano et al., 2017; Ouyang et al., 2022

The RLHF Loss

$$\mathcal{L}_{\mathrm{RM}}(\theta) = -\mathbb{E}_{(x,y_w,y_l)\sim D}\left[\log \sigma\left(r_\theta(x, y_w) - r_\theta(x, y_l)\right)\right]$$

The reward model is trained with the Bradley-Terry preference model, where y_w is the preferred response and y_l is the rejected response.

Christiano et al., 2017

RLHF Timeline

- **2017**: Deep RL from human preferences (Christiano et al., 2017) and PPO (Schulman et al., 2017)
- **2019**: Fine-tuning LMs from human preferences (Ziegler et al., 2019)
- **2020**: Learning to summarize with human feedback (Stiennon et al., 2020)
- **2022**: InstructGPT (Ouyang et al., 2022) and Anthropic's HHH assistant (Bai et al., 2022)
- **2023**: DPO (Rafailov et al., 2023), IPO (Azar et al., 2023), Llama 2 (Touvron et al., 2023), Zephyr (Tunstall et al., 2023), Tülu 2 (Ivison et al., 2023), RLAIF (Lee et al., 2023)
- **2024**: KTO (Ethayarajh et al., 2024), AlpacaFarm (Dubois et al., 2024), and the RLHF Book (Lambert, 2024)

Conclusions

1. Colloquium makes slide creation **fast** and **reproducible**
2. Full LaTeX math support for academic presentations
3. Git-friendly markdown source files
4. AI agents can create and modify presentations programmatically

Thank You

Questions?

References (1/2)

Azar, M., Rowland, M., Piot, B., Guo, D., Calandriello, D., et al.. “*A General Theoretical Paradigm to Understand Learning from Human Feedback*.” *arXiv preprint arXiv:2310.12036*, 2023. [link]

Bai, Y., Jones, A., Ndousse, K., Askell, A., Chen, A., et al.. “*Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback*.” *arXiv preprint arXiv:2204.05862*, 2022. [link]

Christiano, P., Leike, J., Brown, T., Martic, M., Legg, S., et al.. “*Deep Reinforcement Learning from Human Preferences*.” *Advances in Neural Information Processing Systems*, 2017. [link]

Dubois, Y., Li, C., Taori, R., Zhang, T., Gulrajani, I., et al.. “*AlpacaFarm: A Simulation Framework for Methods that Learn from Human Feedback*.” *Advances in Neural Information Processing Systems*, 2024. [link]

Ethayarajh, K., Xu, W., Muennighoff, N., Jurafsky, D., and Kiela, D.. “*KTO: Model Alignment as Prospect Theoretic Optimization*.” *arXiv preprint arXiv:2402.01306*, 2024. [link]

Ivison, H., Wang, Y., Pyatkin, V., Lambert, N., Peters, M., et al.. “*Camels in a Changing Climate: Enhancing LM Adaptation with Tulu 2*.” *arXiv preprint arXiv:2311.10702*, 2023. [link]

Lambert, N.. “*Reinforcement Learning from Human Feedback*.” *Manning Publications*, 2024. [link]

Lee, H., Phatale, S., Mansoor, H., Lu, K., Mesnard, T., et al.. “*RLAIF: Scaling Reinforcement Learning from Human Feedback with AI Feedback*.” *arXiv preprint arXiv:2309.00267*, 2023. [link]

References (2/2)

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., et al.. “*Training Language Models to Follow Instructions with Human Feedback*.” *Advances in Neural Information Processing Systems*, 2022. [link]

Rafailov, R., Sharma, A., Mitchell, E., Ermon, S., Manning, C., et al.. “*Direct Preference Optimization: Your Language Model is Secretly a Reward Model*.” *Advances in Neural Information Processing Systems*, 2023. [link]

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O.. “*Proximal Policy Optimization Algorithms*.” *arXiv preprint arXiv:1707.06347*, 2017. [link]

Stiennon, N., Ouyang, L., Wu, J., Ziegler, D., Lowe, R., et al.. “*Learning to Summarize with Human Feedback*.” *Advances in Neural Information Processing Systems*, 2020. [link]

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., et al.. “*Llama 2: Open Foundation and Fine-Tuned Chat Models*.” *arXiv preprint arXiv:2307.09288*, 2023. [link]

Tunstall, L., Beeching, E., Lambert, N., Rajani, N., Rasul, K., et al.. “*Zephyr: Direct Distillation of LM Alignment*.” *arXiv preprint arXiv:2310.16944*, 2023. [link]

Ziegler, D., Stiennon, N., Wu, J., Brown, T., Radford, A., et al.. “*Fine-Tuning Language Models from Human Preferences*.” *arXiv preprint arXiv:1909.08593*, 2019. [link]