Improving Machine Learning from Human Feedback

Erin Mikail Staples + Nikolai Lubimov

PyData DE 2023

Erin Mikail Staples (she/her)
Sr. Developer Community Advocate

Empowers the open source community through education, collaboration, and content creation.

Nikolai Liubimov (he/him)
CTO

Helps customers debug and adopt label studio usage best practices

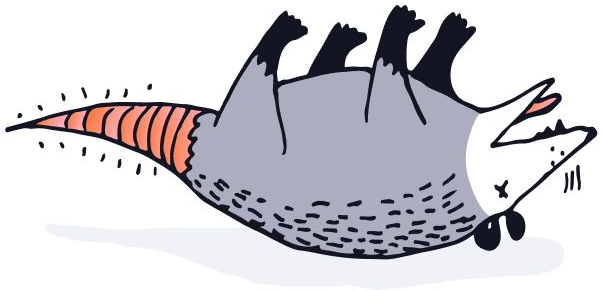
Large Foundational Models have hit the cultural zeitgeist


We will not be
creating Terminator here.

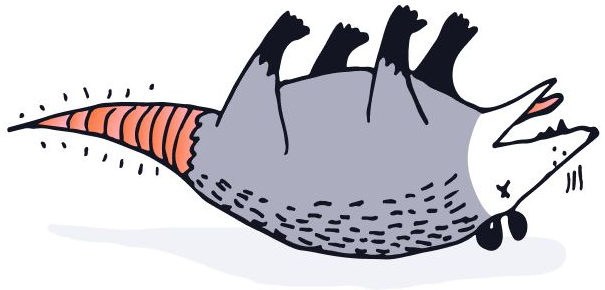
These large generative models are better with a human signal.

Why does this matter?

Bigger ≠ Better

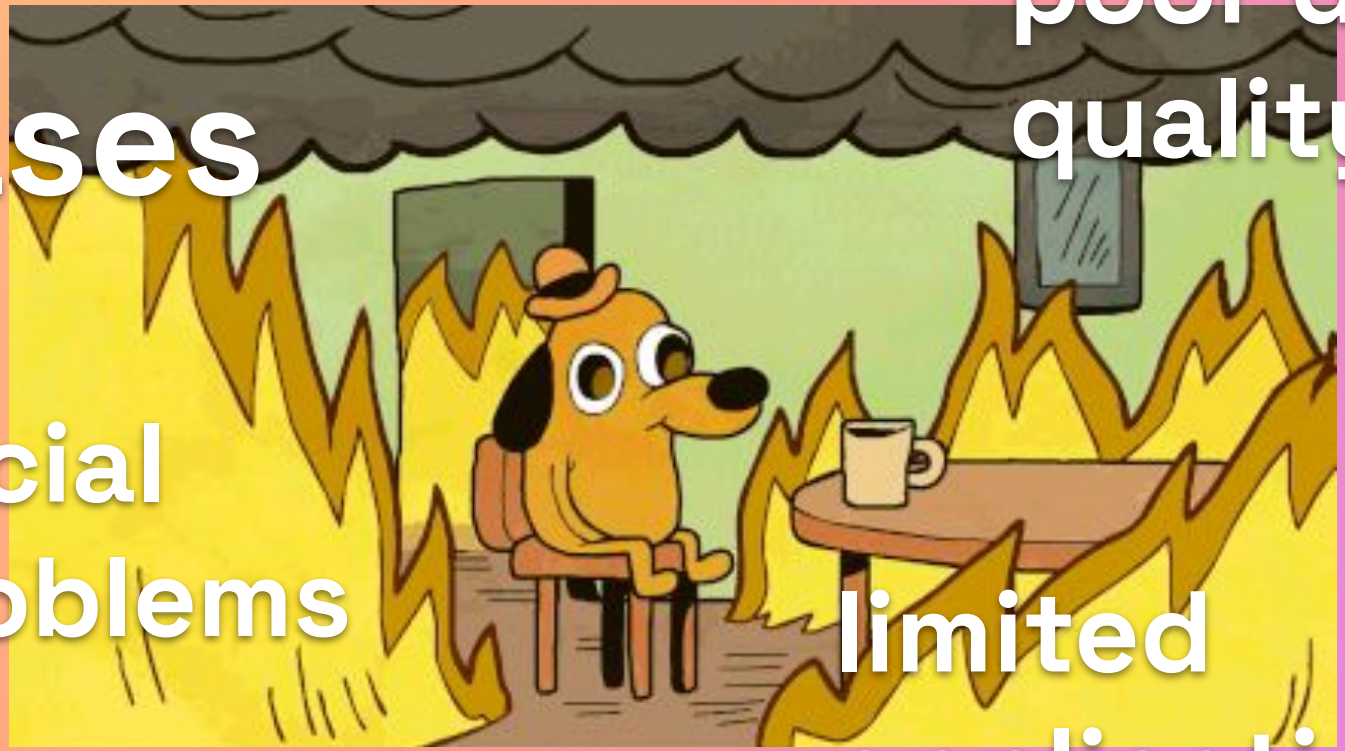


Internet-trained models bring with them internet-scaled biases.



biases

poor data quality



social problems

limited applications

Power of Reinforcement Learning

Andrej Karpathy
@karpathy

The hottest new programming language is English

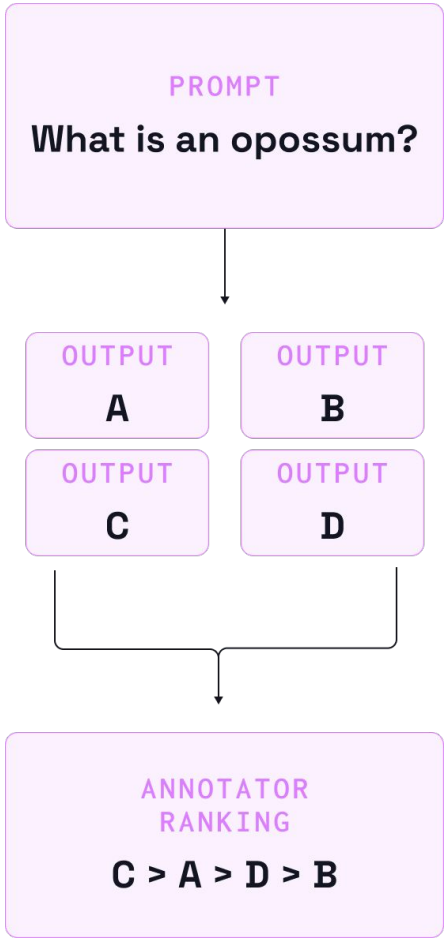
9:14 PM · Jan 24, 2023 · **2.2M** Views

Reinforcement Learning with Human Feedback helps to adjust for problems that tend to come with large-scale foundational models.

Reinforcement Learning

Goal-oriented model that seeks to identify similar actions or sequence of actions that would maximize future rewards.

Able to select the **best output among a series of outputs.**



Unsupervised Learning and Prompt Engineering focuses on adapting to an existing model's limitations.

Known limitations include:

- Harmful Speech
- Overgeneralized Data
- Out-of-Date Data
- Contain racial, gender, and religious biases
- Require large computational resources

Reinforcement Learning focuses on optimizing for the end goal by adapting the model itself to new and possibly uncertain information based on a human signal.

With RLHF one can align model output with one's specific needs while reducing bias at a fraction of the original training cost.

Open Source Tools

for Reinforcement Learning

- BLOOM
- ChatAlpaca
- OpenLlama
- CasperAI/TRLX
- PyTorch
- InstructGOOSE
- Label Studio
- Hugging Face

We're already seeing RLHF used in the wild

Announcements

Share

Introducing BloombergGPT, Bloomberg's 50-billion parameter large language model, purpose-built from scratch for finance

March 30, 2023

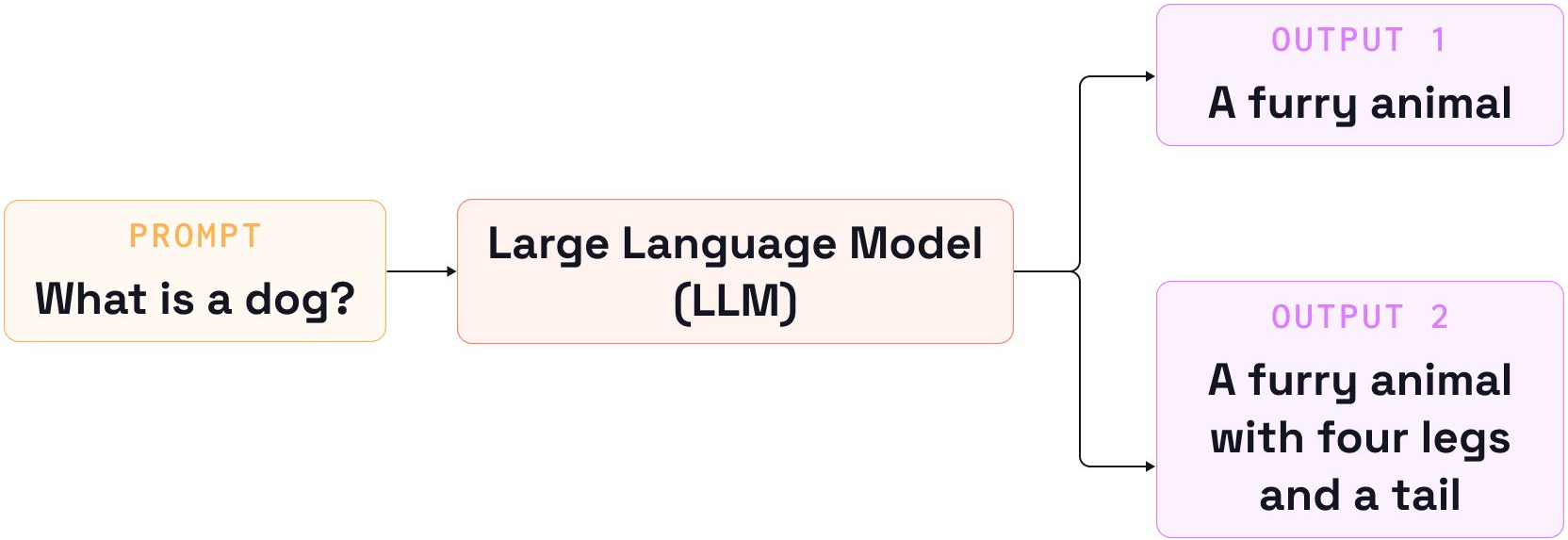
BloombergGPT outperforms similarly-sized open models on financial NLP tasks by significant margins – without sacrificing performance on general LLM benchmarks

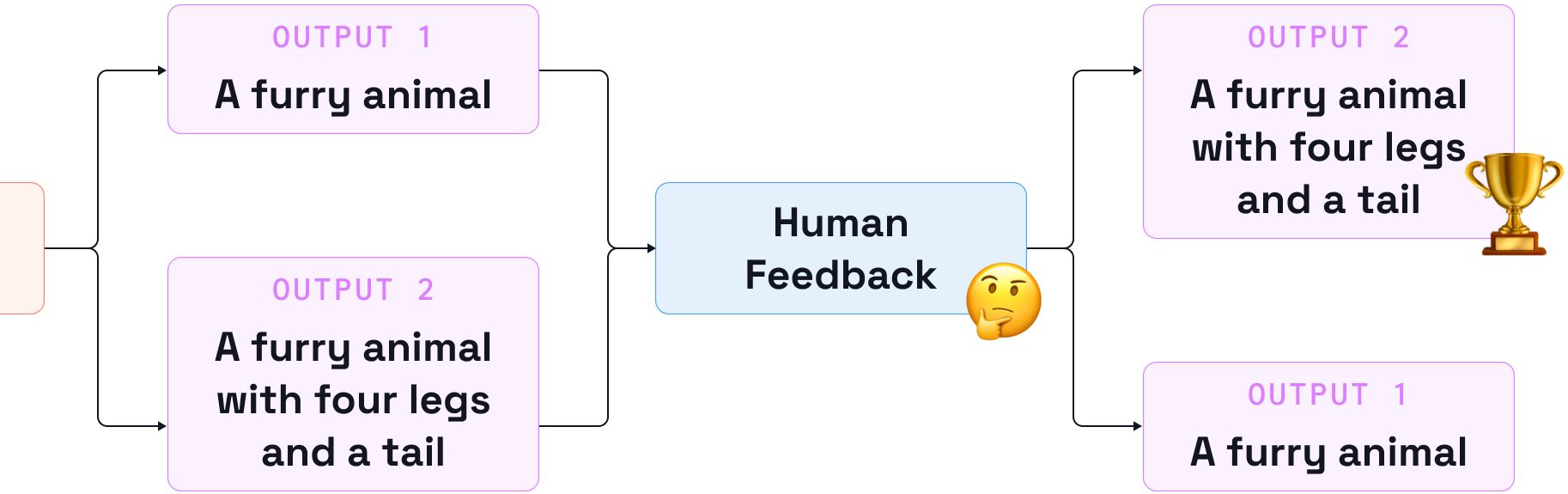
NEW YORK – Bloomberg today released a research paper detailing the development of BloombergGPT™, a new large-scale generative artificial intelligence (AI) model. This large language model (LLM) has been specifically trained on a wide range of financial data to support a diverse set of natural language processing (NLP) tasks within the financial industry.

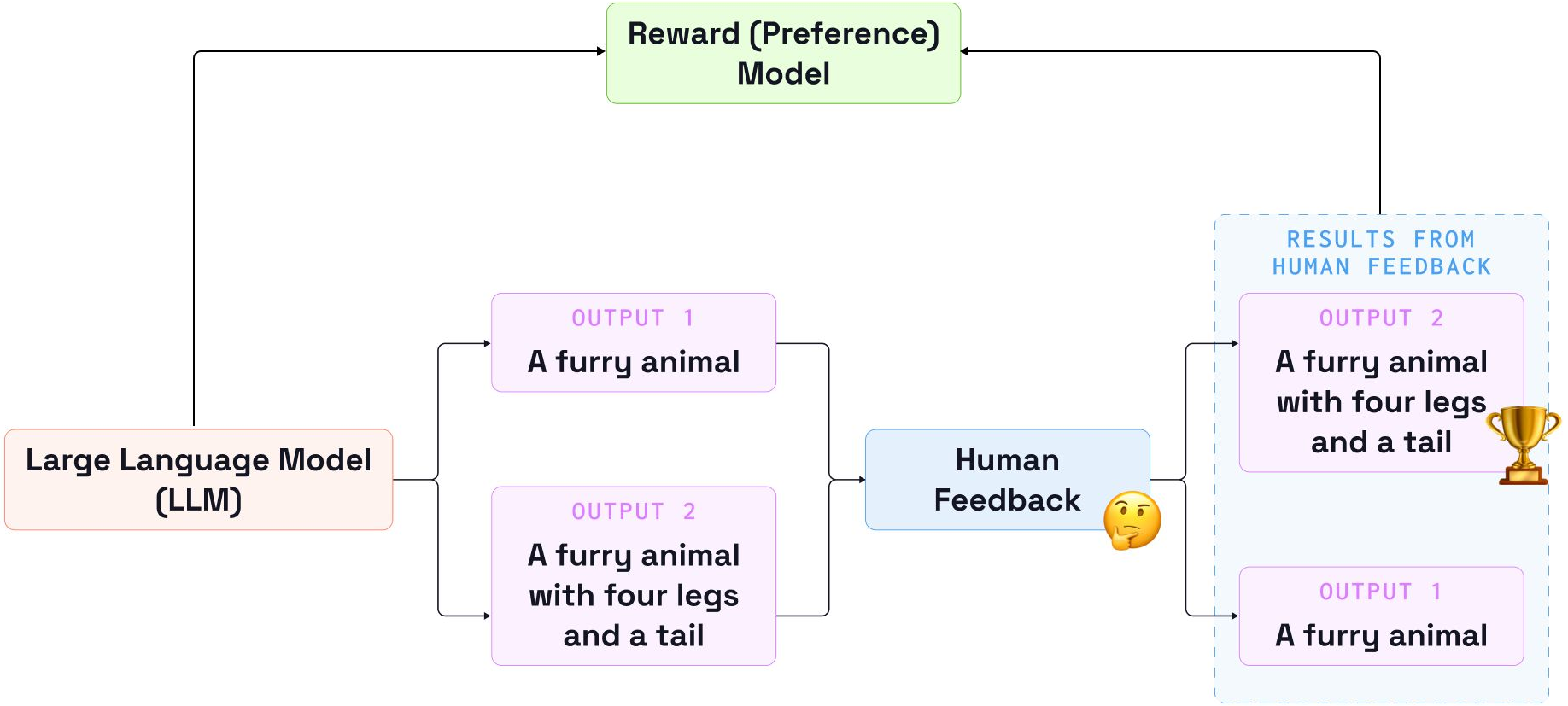
Recent advances in Artificial Intelligence (AI) based on LLMs have already demonstrated exciting new applications for many domains. However, the

So how did they do it?

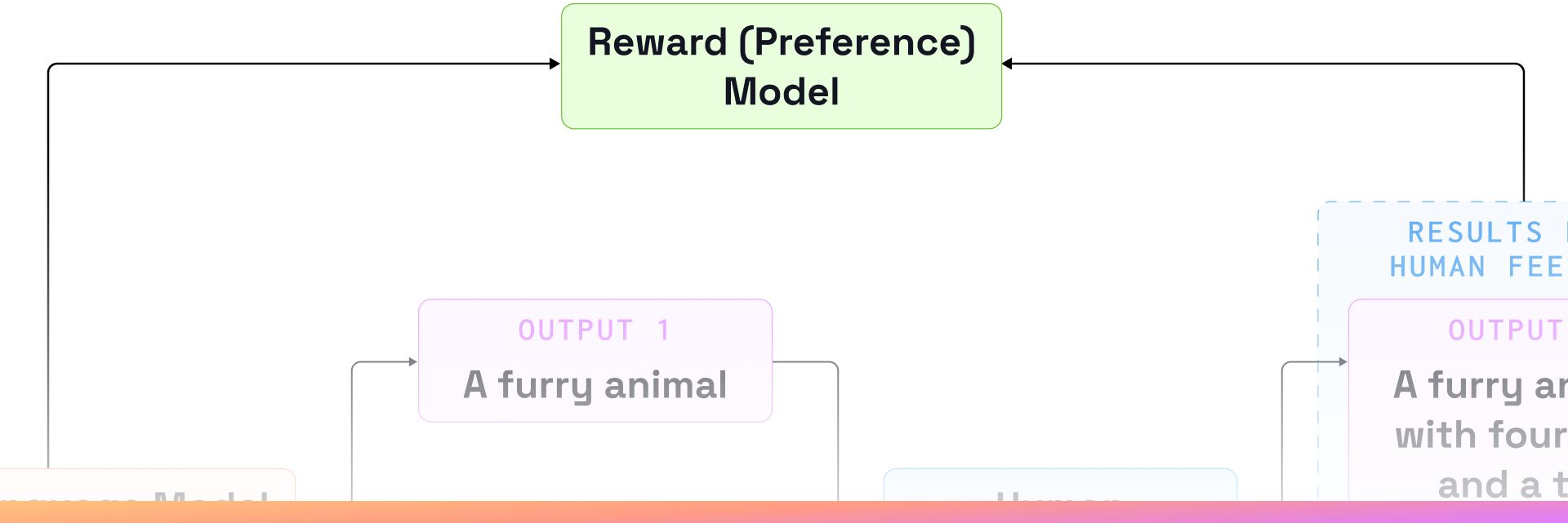
Large Language Model (LLM)


PROMPT
What is a dog?
Large Language Model (LLM)
OUTPUT 1
A furry animal
OUTPUT 2
A furry animal with four legs and a tail

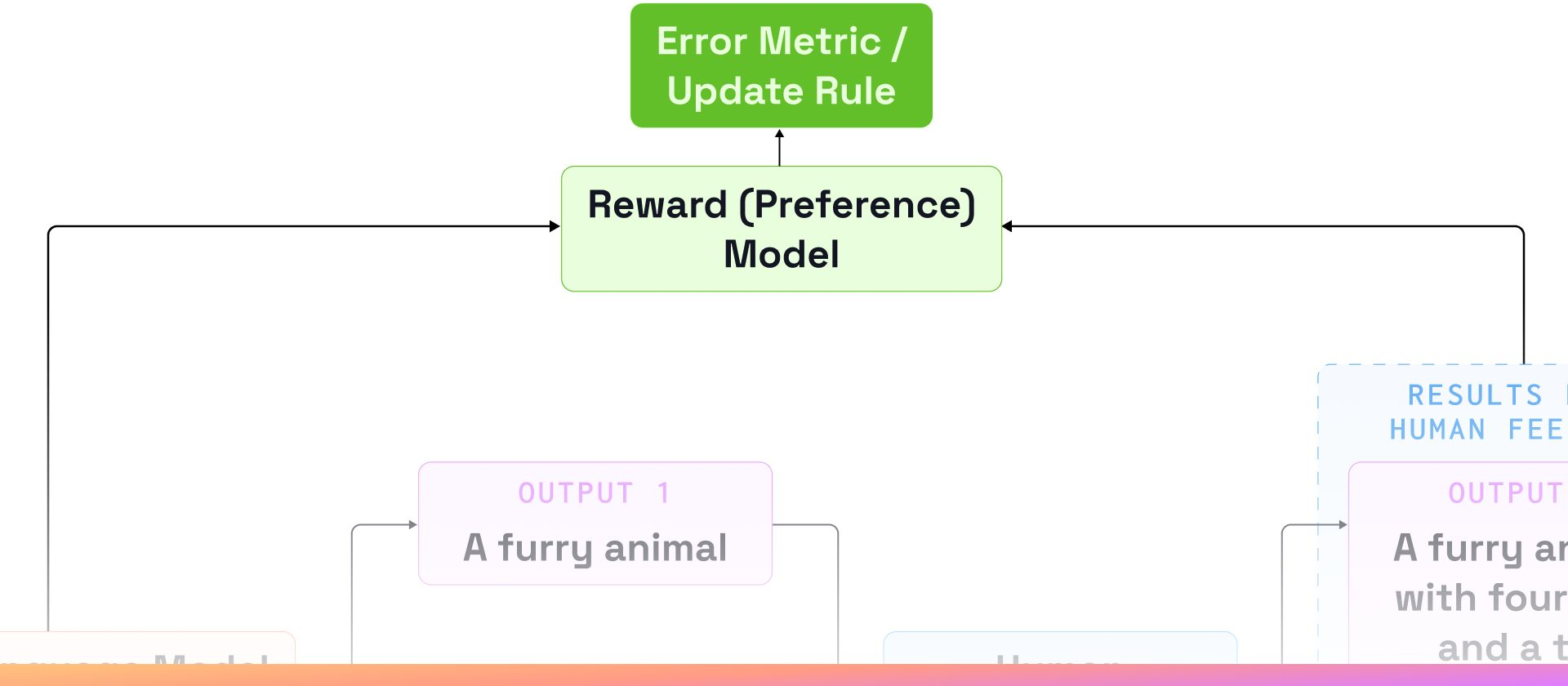

OUTPUT 1
A furry animal
OUTPUT 2
A furry animal with four legs and a tail
Human Feedback
OUTPUT 2
A furry animal with four legs and a tail
OUTPUT 1
A furry animal


Reward (Preference) Model
Large Language Model (LLM)
OUTPUT 1
A furry animal
OUTPUT 2
A furry animal with four legs and a tail
Human Feedback
RESULTS FROM HUMAN FEEDBACK
OUTPUT 2
A furry animal with four legs and a tail
OUTPUT 1
A furry animal

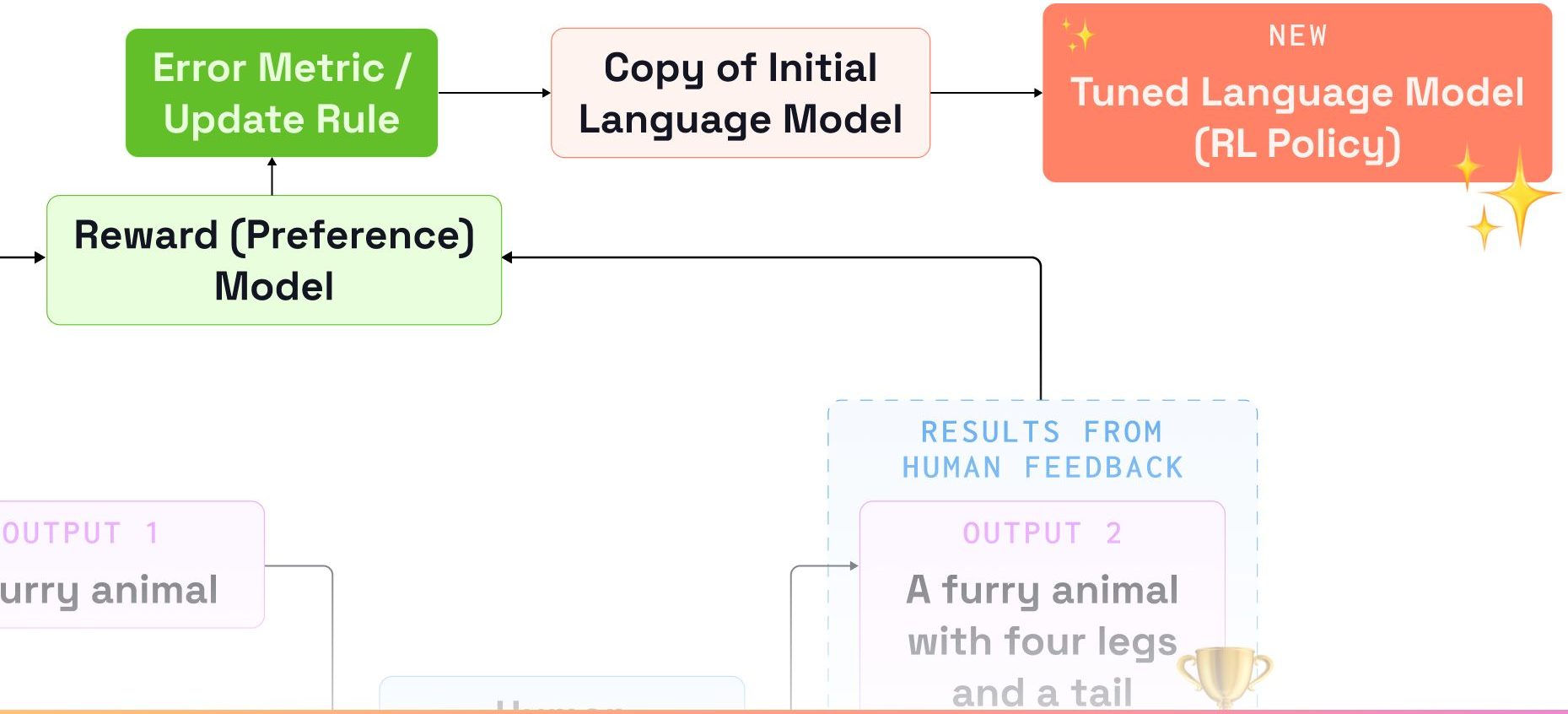
The Importance of the Reward (Preference) Models



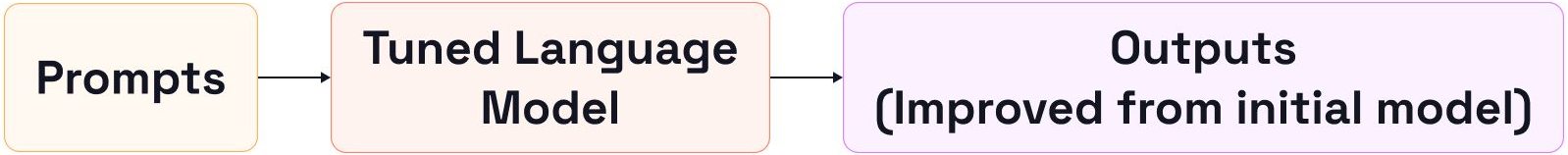
Preventing Unwanted Model Drift



Final Stages of Model Development



Ready for Production



We know what this looks like *theoretically...*

… now let’s demonstrate this in real time.

See it in action!

https://github.com/heartexlabs/RLHF

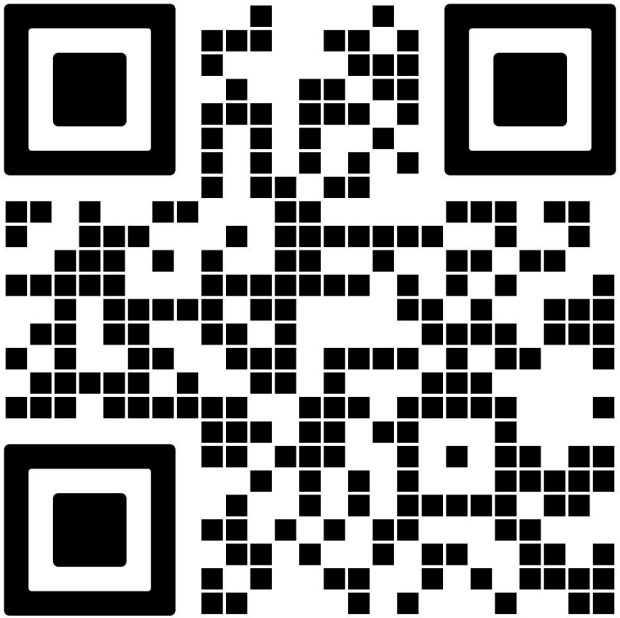
Problems with RLHF

Humans ruin everything.

RLHF relies on social engineering and data integrity as much as it does technical skill.

Keeping annotators well-informed and motivated

Try out RLHF for yourself. ➡️



@erinmikail
@liubimovnik
@labelstudioHQ

community@labelstud.io

https://labelstud.io/pydata-berlin