

**EMIL Seminar - 04/23/2024**

**Saman Khamesian - AI News**

# Introducing Meta Llama 3

- Last week, Meta introduced its newest AI assistant called Llama 3
- It's been trained on
  - A custom-built 24K GPU clusters on over 15T token of data.
  - A training dataset 7x larger than that used for Llama 2, including 4x more code.
- It supports a 8K context length that doubles the capacity of Llama 2

# Introducing Meta Llama 3

Meta Llama 3 Instruct model performance

	Meta Llama 3 8B	Gemma 7B - It Measured	Mistral 7B Instruct Measured
MMLU 5-shot	68.4	53.3	58.4
GPQA 0-shot	34.2	21.4	26.3
HumanEval 0-shot	62.2	30.5	36.6
GSM-8K 8-shot, CoT	79.6	30.6	39.9
MATH 4-shot, CoT	30.0	12.2	11.0

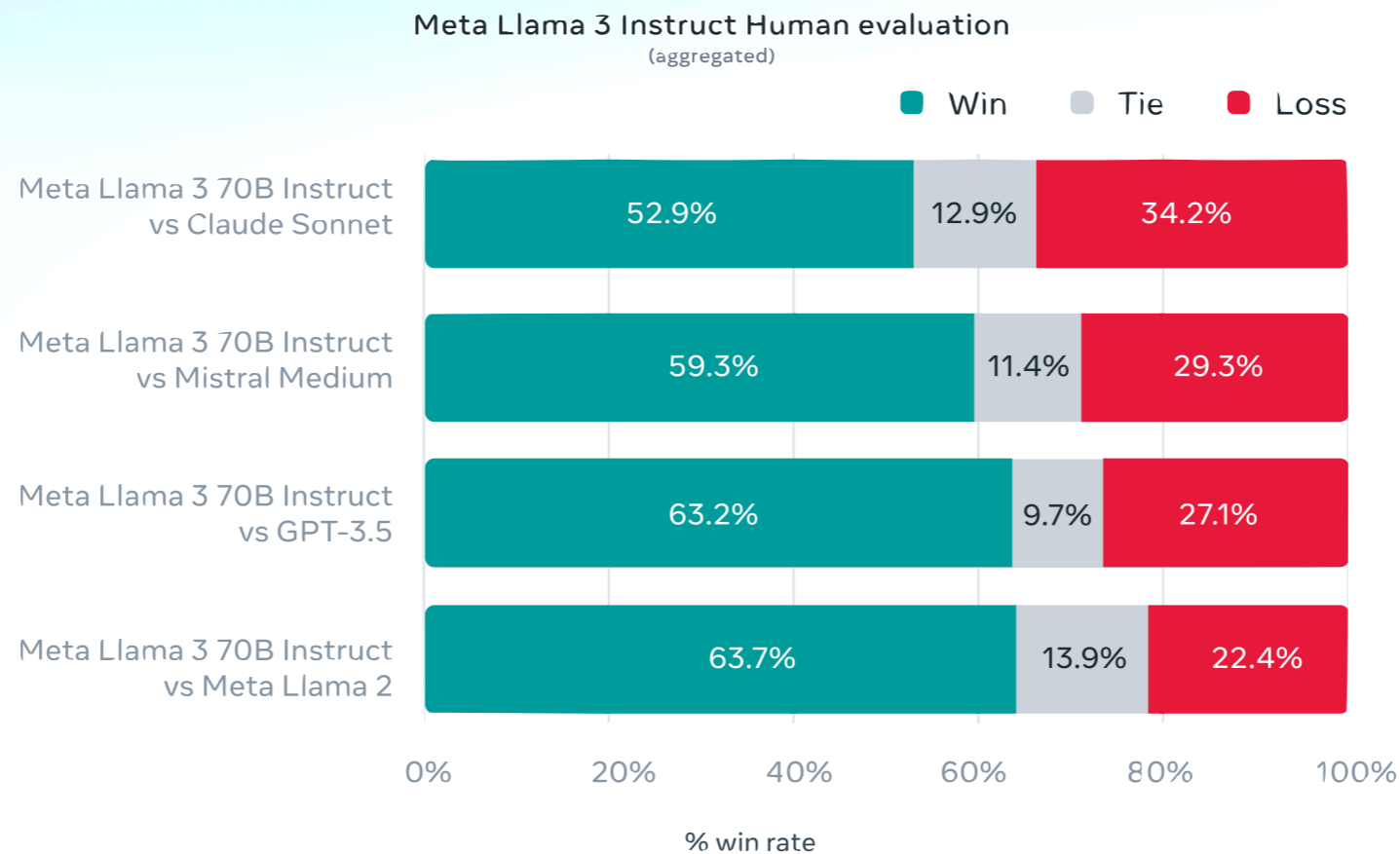
	Meta Llama 3 70B	Gemini Pro 1.5 Published	Claude 3 Sonnet Published
MMLU 5-shot	82.0	81.9	79.0
GPQA 0-shot	39.5	41.5 CoT	38.5 CoT
HumanEval 0-shot	81.7	71.9	73.0
GSM-8K 8-shot, CoT	93.0	91.7 11-shot	92.3 0-shot
MATH 4-shot, CoT	50.4	58.5 Minerva prompt	40.5

# Introducing Meta Llama 3

- They developed a new high-quality human evaluation set.
- This evaluation set contains 1,800 prompts that cover 12 key use cases: asking for advice, brainstorming, classification, closed question answering, coding, creative writing, extraction and etc.
- To prevent accidental overfitting of their models on this evaluation set, even their own modeling teams do not have access to it.

# Introducing Meta Llama 3

- The chart below shows aggregated results of their human evaluations across of these categories and prompts against Claude Sonnet, Mistral Medium, and GPT-3.5.



# Introducing Meta Llama 3



# Introducing Meta Llama 3

- Meta used this AI assistance across all its other platforms like Instagram, WhatsApp and Facebook.

