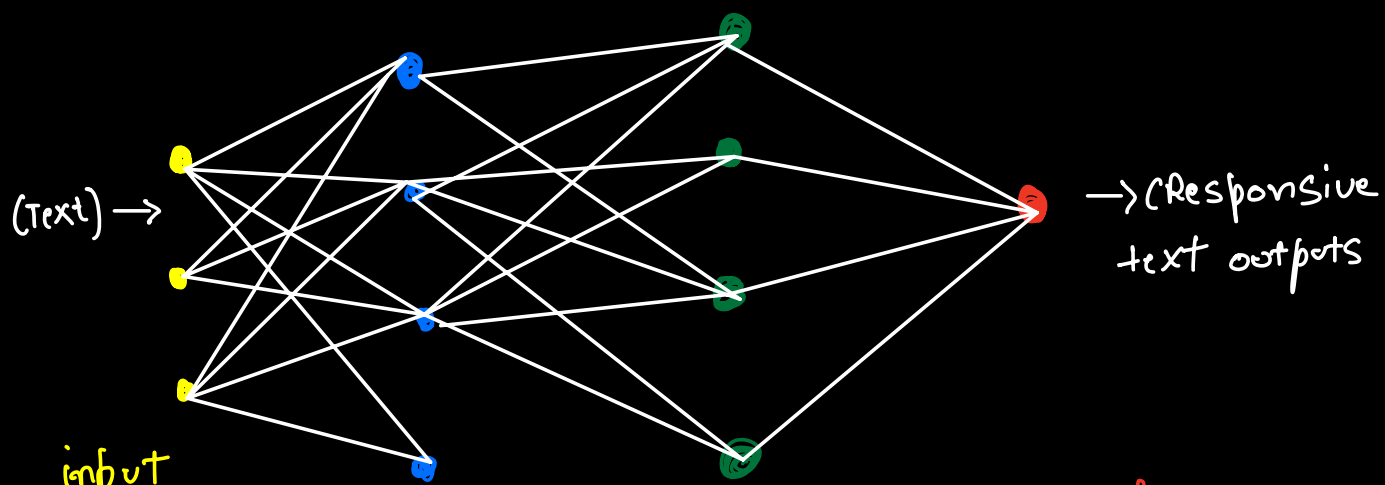
# Heart of Chat GPT

* LLM (GPT-3.5) ... > GPT 4 (latest LLM)



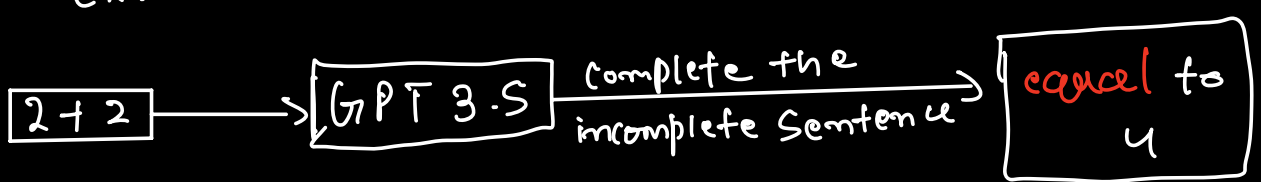

input layer

input layer 1

input layer 2

output layer

ex:




LLM -> large Lang Model

* Characterized by size

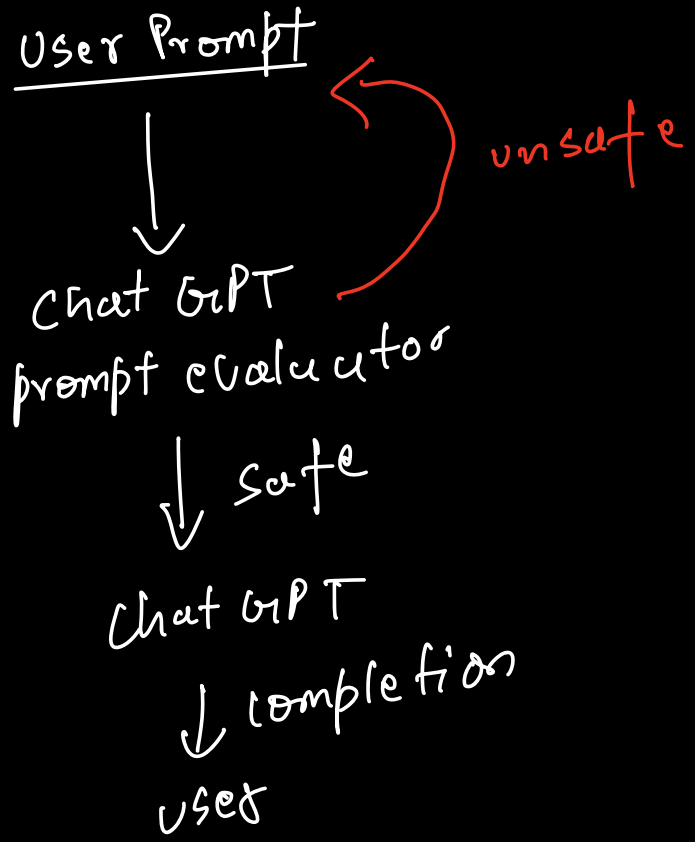
GPT 3.5 has 175 billion parameters with 96 layers

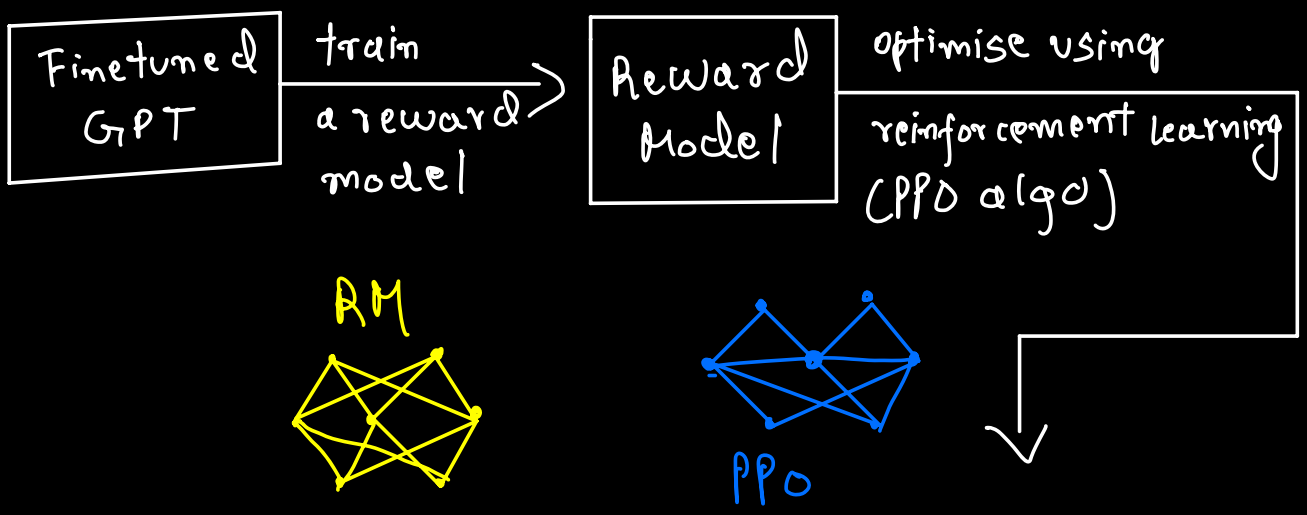
Text: My fav color is red

Token Ids: [3666, 4004, 3124, 318, 2266, 13]

Source Data: 500 billion tokens (100s of billions of words on the net)

* Model is trained to predict tokens, based on input tokens.






How the model is used to answer a prompt?

1. CPI

build context → Context
- India
- PM

ex: Prompt: PM of India

Answer: The PM of " is Modi

↳ build context → Context
- India
- PM
- Modi

* Chat GPT UI provides context from previous responses to keep building context

↳ conversational Prompt Injection

## 2. Primary Prompt Engineering

→ Build prompts

Tone: Soft
- Rhyme
- Long: Eng

Prompt: Complete the rhyme:

Roses are Red, oceans is blue

Sugar is sweet, and so are you,

→ Mode: happy

* These prompts are invisible to user

## 3. Moderation API:

Build constraints → - stock
- finance

Content moderation

↓

pre-trained model → N → Template response generation

↓

Response ← return results