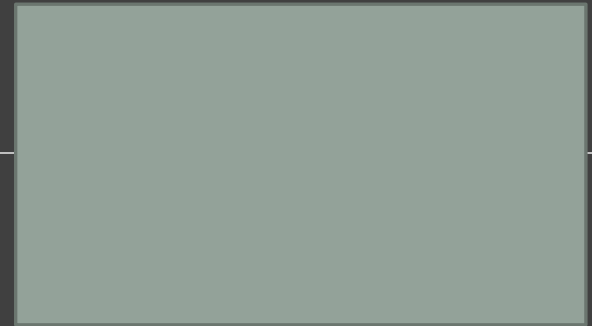


ChatGPT: Training an LLM

SAM SCOTT, MOHAWK COLLEGE, JUNE 2023



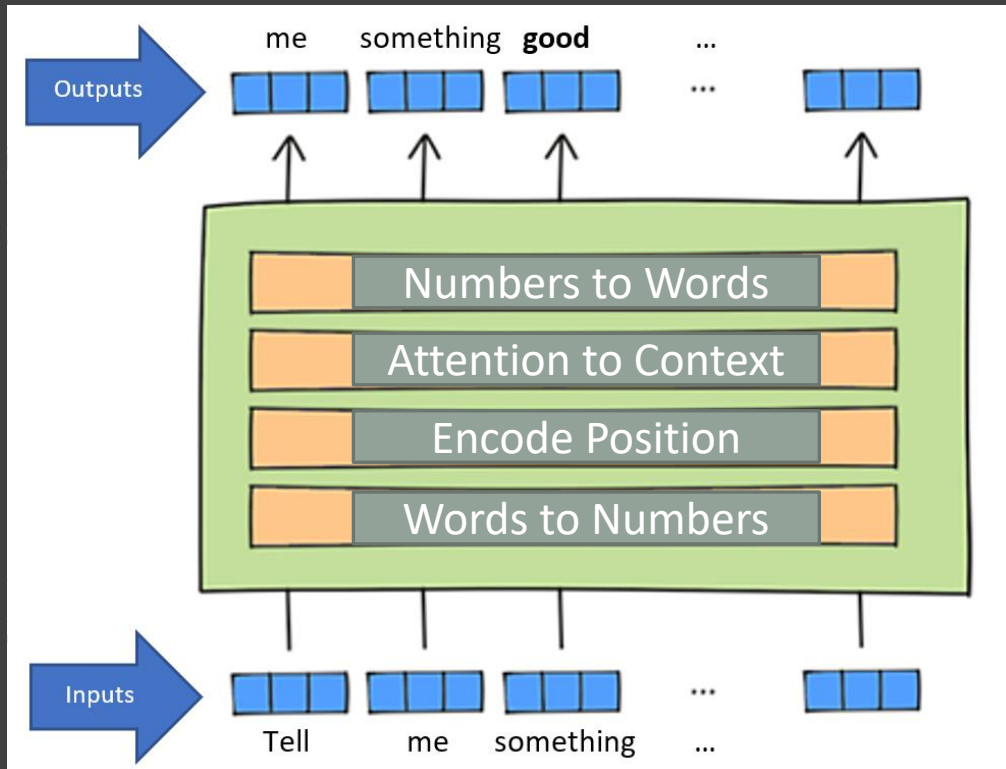
Reminder: The Task

Given a text **prompt**, predict the natural language **token** (word) that comes next.

ChatGPT is a **Large Language Model** powered by a deep **Artificial Neural Network** architecture called a **Transformer**.



Reminder: The Journey

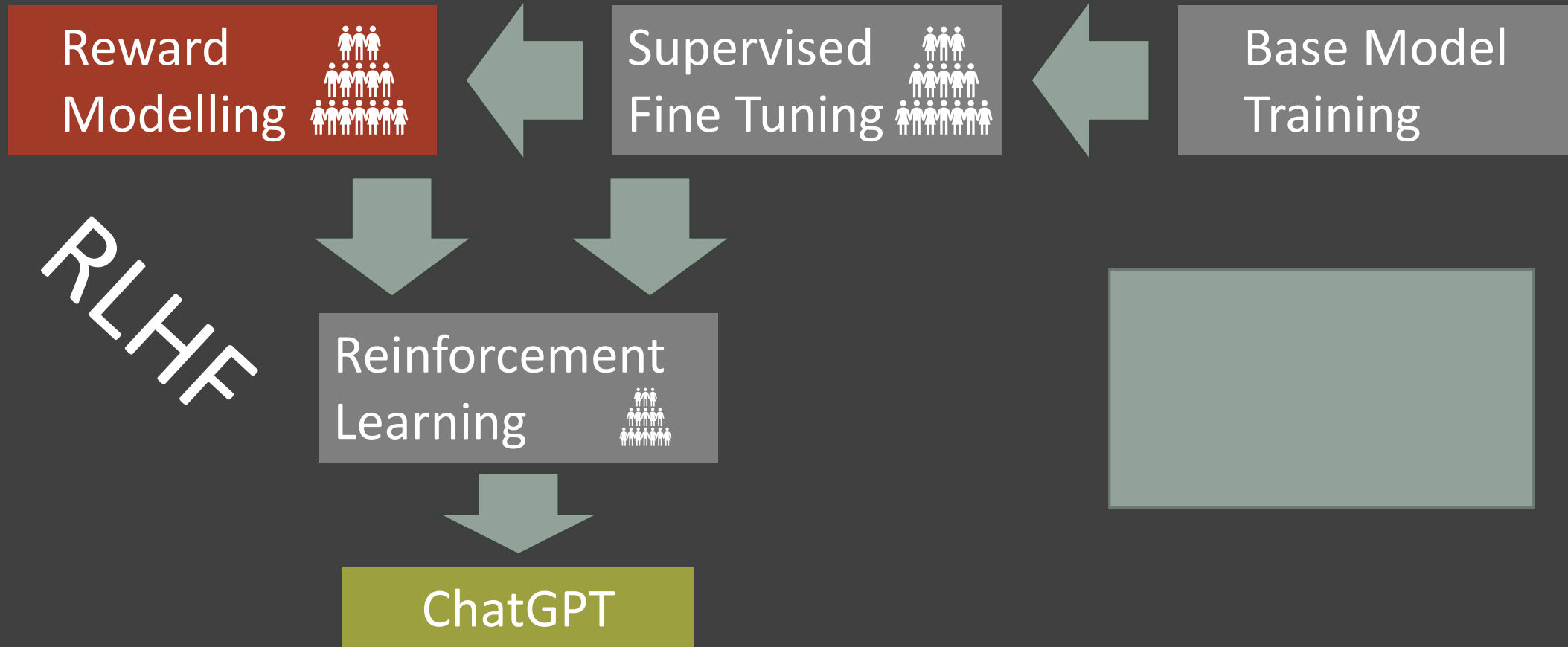


Adapted from <https://www.lavivienpost.com/how-chatgpt-works-architecture-illustrated/>

ChatGPT generates one word at a time based on a generalized context.



Training Methods for ChatGPT



Base Model Training

Typical Scope	Human Involvement	Base Models
A trillion tokens, 1000 GPUs, many months	None	davinci, curie, babbage, ada

The Study Archive file format is a data compression and archive format, based directly on the ZIP file format.[1] Study archives are used for the development and use of sets of interactive file cards, which may contain not only text but also images and audio files, for various applications.



System is gently nudged
in the right direction.



Supervised Fine Tuning

Typical Scope	Human Involvement	SFT Models
100 000 prompts and responses, 100 GPUs, one month	Prompts with responses	davinci-instruct-beta

Prompt: Can you give me some suggestions for classic 90's rock songs I would like? I'm a big fan of Nirvana and Pearl Jam but I know all their songs already.

Response: 1. Soundgarden – Rusty Cage, 2. Radiohead – Creep, 3. Smashing Pumpkins – 1979

Prompt: Hi, how are you today?

Response: As an AI, I do not have feelings or emotions, but I am here to help you. How can I assist you today?

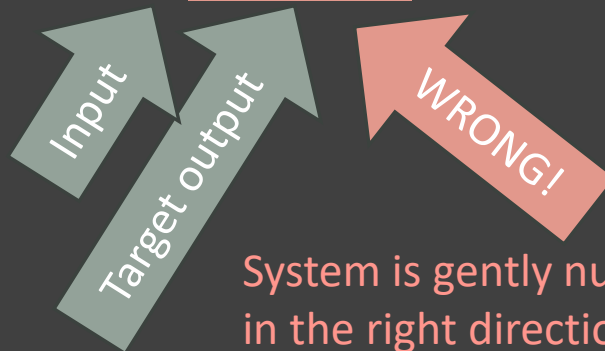


Supervised Fine Tuning

Typical Scope	Human Involvement	SFT Models
100 000 prompts and responses, 100 GPUs, one month	Prompts with responses	davinci-instruct-beta

Prompt: Can you give me some suggestions for classic 90's rock songs I would like? I'm a big fan of Nirvana and Pearl Jam but I know all their songs already.

Response: 1. Soundgarden – Rusty Cage, 2. Nirvana – Creep, 3. Smashing Pumpkins – 1979



Reward Modelling

Typical Scope	Human Involvement	Reward Models
A million response ratings, 100 GPUs, one month	Response ratings	N.A.

Prompt: Can you give me some suggestions for classic 90's rock songs I would like? I'm a big fan of Nirvana and Pearl Jam but I know all their songs already.

Response: 1. Nirvana – Drain You, 2. Pearl Jam – Jeremy, 3. Smashing Pumpkins – 1979 <|end|> <|-0.2|>

Prompt: Hi, how are you today?

Response: As an AI, I do not have feelings or emotions, but I am here to help you. How can I assist you today? <|end|> <|1.0|>



End

Reward

End

Reward

Reward Modelling

Typical Scope	Human Involvement	Reward Models
A million response ratings, 100 GPUs, one month	Response ratings	N.A.

Prompt: Can you give me some suggestions for classic 90's rock songs I would like? I'm a big fan of Nirvana and Pearl Jam but I know all their songs already.

Response: 1. Nirvana – Drain You, 2. Pearl Jam – Jeremy, 3. Smashing Pumpkins – 1979 <|end|> <|-0.2|>

Prompt: Hi, how are you today?

Response: As an AI I do not have feelings or emotions, but I am here to help you. How can I assist you today? <|end|> <|1.0|>

Input

Target output



Reward Modelling

Typical Scope	Human Involvement	Reward Models
A million response ratings, 100 GPUs, one month	Response ratings	N.A.

Prompt: Can you give me some suggestions for classic 90's rock songs I would like? I'm a big fan of Nirvana and Pearl Jam but I know all their songs already.

Response: 1. Nirvana – Drain You, 2. Pearl Jam – Jeremy, 3. Smashing Pumpkins – 1979 <|end|> <|-0.2|>

Prompt: Hi, how are you today?

Response: As an AI, I do not have feelings or emotions, but I am here to help you. How can I assist you today? <|end|> <|1.0|>



Reinforcement Learning

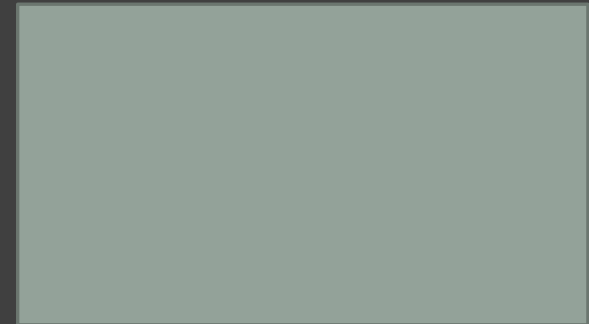
Typical Scope	Human Involvement	RLHF Models
30 000 prompts, 100 GPUs, one month	Prompts	text-davinci-003, ChatGPT

Prompt: Can you give me some suggestions for classic 90's rock songs I would like? I'm a big fan of Nirvana and Pearl Jam but I know all their songs already.

Response: 1. Nirvana – Drain You, 2. Pearl Jam – Jeremy, 3. Smashing Pumpkins – 1979 <|end|>

Reward Model: -0.4

System is nudged away from producing these tokens.



Reinforcement Learning

Typical Scope	Human Involvement	RLHF Models
30 000 prompts, 100 GPUs, one month	Prompts	text-davinci-003, ChatGPT, etc.

Prompt: Hi, how are you today?

Response: As an AI, I do not have feelings or emotions, but I am here to help you. How can I assist you today? <|end|>

Reward Model: 1.3

System is nudged towards producing these tokens.

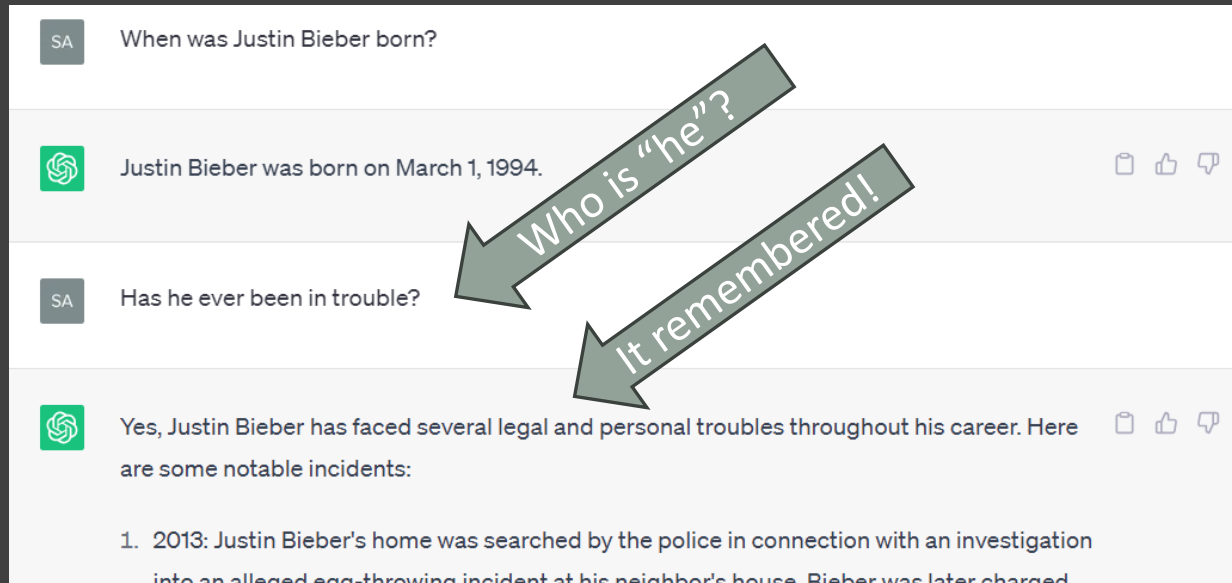


Training Methods Recap

Type		Scope	Humans
Base Model Training		1 000 000 000 000 tokens, 1000 GPUs, months of training	None
Supervised Fine Tuning		100 000 prompts and responses, 100 GPUs, weeks of training	Prompt and response generation
RLHF	Reward Modelling	1 000 000 response ratings, 100 GPUs, weeks of training	Response ratings
	Reinforcement Learning	30 000 prompts, 100 GPUs, weeks of training	Prompt generation



From Completion to Chat



ChatGPT seems to track the conversation

- Prompt Engineering

ChatGPT acts like a helpful assistant.

- SFT and RLHF
- Prompt constraints
- A “system” message

Can we trust ChatGPT?

Is it moderated by humans?

- No

Does it memorize true information or have some kind of database?

- No

Is it ok to cite ChatGPT as a source?

- No!!!
- There's no author
- There's no built-in drive to tell the truth
- The “hallucination” problem is not going to go away



Up Next...

- The Discovery of Prompt Engineering
- Social, Ethical, Cognitive Implications

