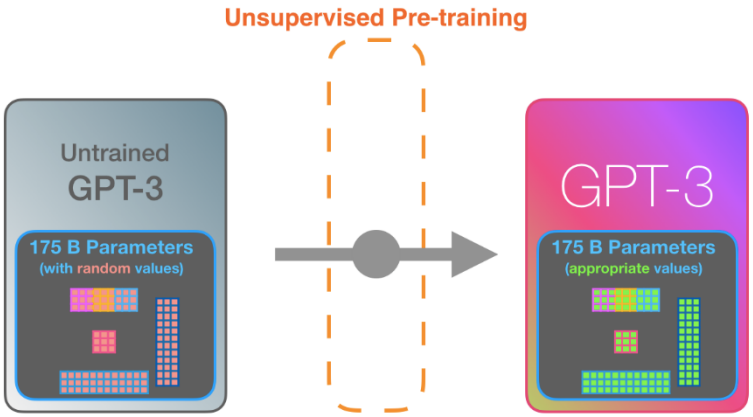
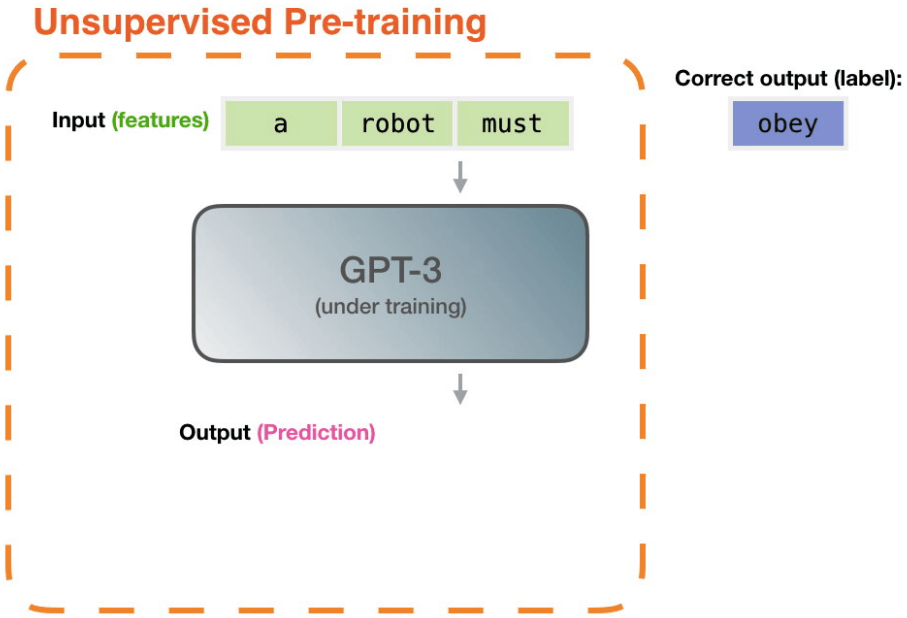
GPT-3: Word Prediction

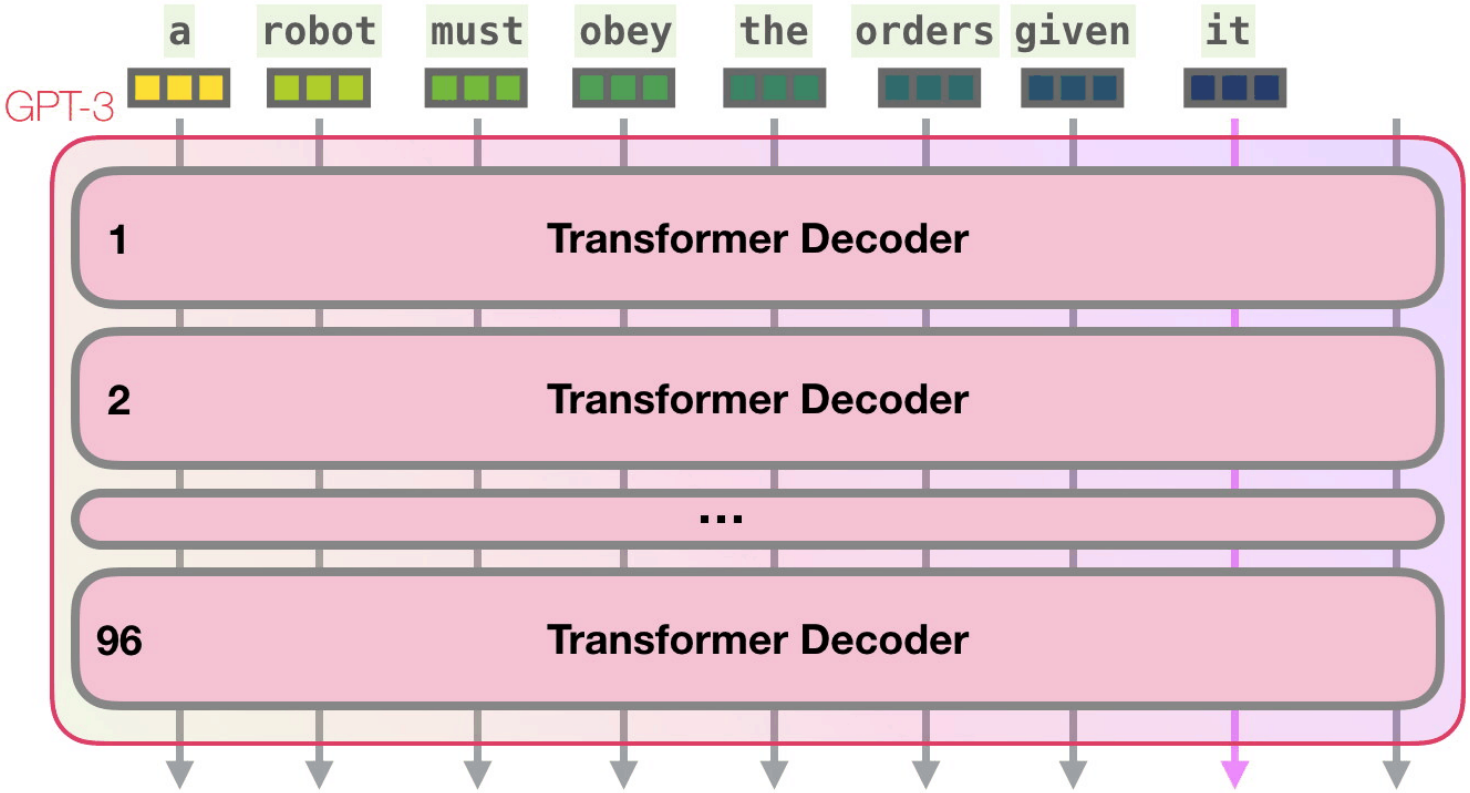




Word Prediction: Repeat millions of times.

How GPT3 Works - Visualizations and Animations, 2020
jalammar.github.io/how-gpt3-works-visualizations-animations/

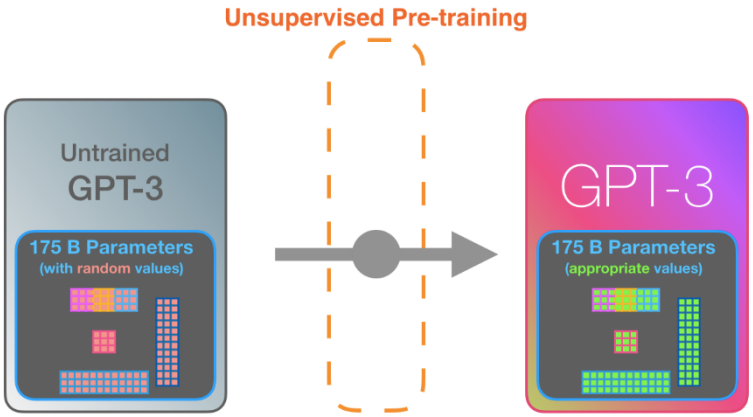
GPT-3: Word Prediction

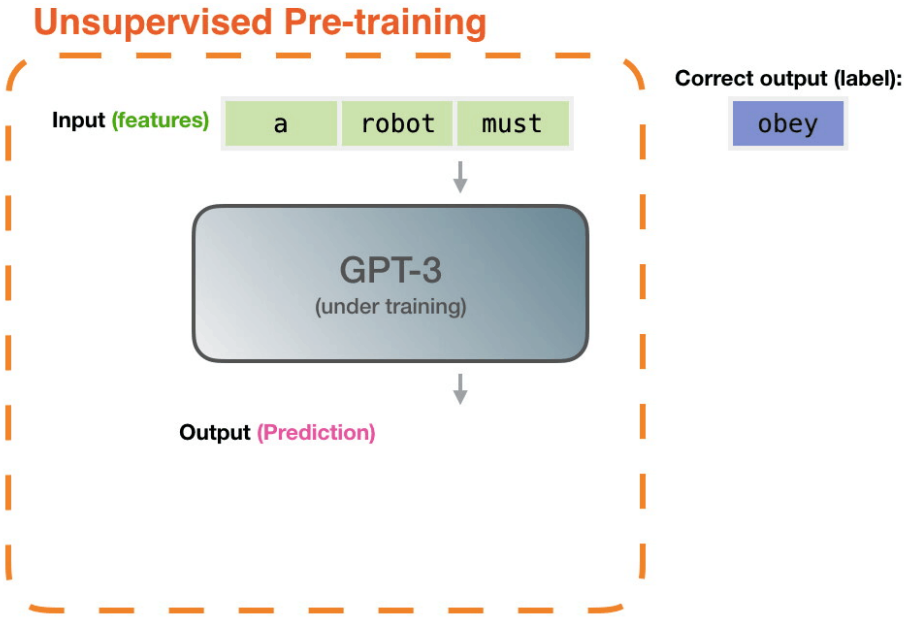


GPT-3 has **96 Decoder Layer** with 96 **attention heads**.

How GPT3 Works - Visualizations and Animations, 2020
jalammar.github.io/how-gpt3-works-visualizations-animations/

GPT-3: Word Prediction

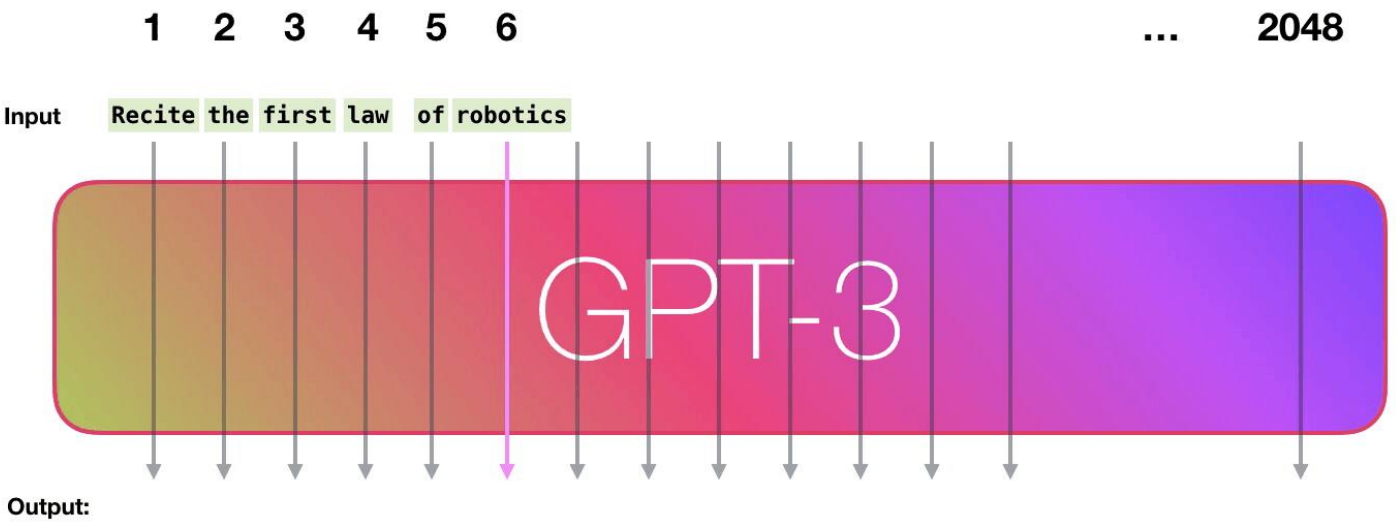




Word Prediction: Repeat millions of times.

How GPT3 Works - Visualizations and Animations, 2020
jalammar.github.io/how-gpt3-works-visualizations-animations/

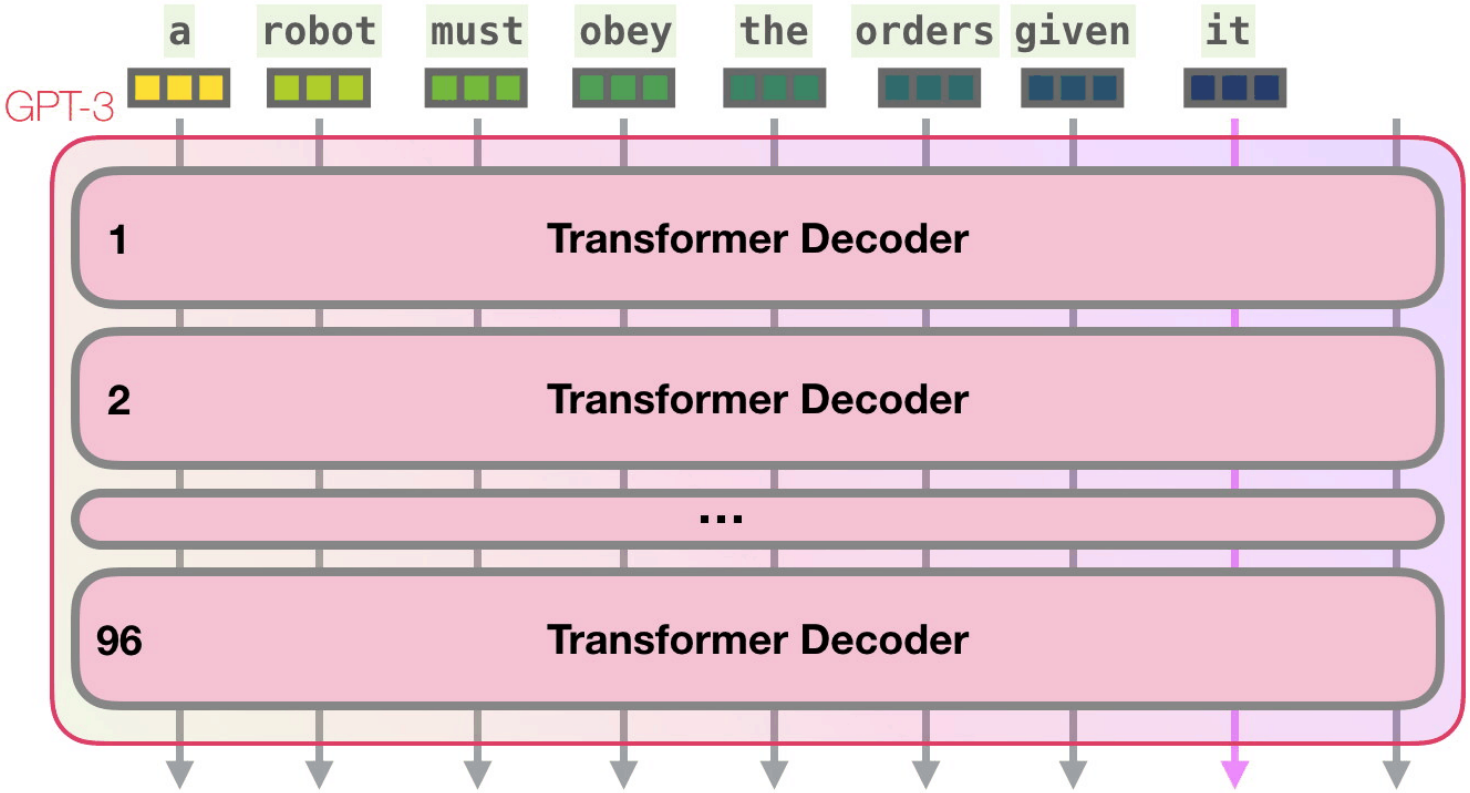
GPT-3: Word Prediction



GPT-3 is **2048 tokens** wide. That is its "**context window**". That means it has **2048 tracks** along which tokens are processed.

How GPT3 Works - Visualizations and Animations, 2020
jalammar.github.io/how-gpt3-works-visualizations-animations/

GPT-3: Word Prediction



GPT-3 has **96 Decoder Layer** with 96 **attention heads**.

How GPT3 Works - Visualizations and Animations, 2020
jalammar.github.io/how-gpt3-works-visualizations-animations/