

< Chapter: Overview >

GPT-2 (small)

nano-gpt

GPT-2 (XL)

GPT-3



Table of Contents

Intro

Introduction

Preliminaries

Components

Embedding

Layer Norm

Self
Attention

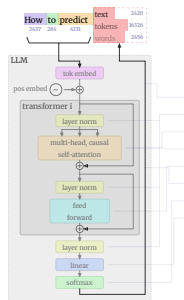
Projection

MLP

Transformer

Softmax

Output



Welcome to the walkthrough of the GPT large language model! Here we'll explore the model *nano-gpt*, with a mere 85,000 parameters.

Its goal is a simple one: take a sequence of six letters:

Press Space to continue

and sort them in alphabetical order, i.e. to "ABBBCC".



We call each of these letters a *token*, and the set of the model's different tokens make up its *vocabulary*.

Continue

Skip

nano-gpt
n_params = 85,584

