

# Relatório Técnico: Modelo GPT-2

## 1. Definição e Aplicação

### O que é o GPT-2?

O **GPT-2** (Generative Pre-trained Transformer 2) é um modelo de linguagem baseado em transformers desenvolvido pela **OpenAI** e lançado em fevereiro de 2019. É um modelo autoregressivo treinado para prever a próxima palavra em uma sequência de texto.

### Aplicações Principais

- **Geração de Texto:** Produção de textos coerentes e contextualmente relevantes
- **Completar Frases:** Autocompletar sentenças e parágrafos
- **Sumarização:** Resumo automático de documentos
- **Tradução:** Tradução de idiomas (com ajuste fino)
- **Chatbots:** Base para sistemas conversacionais
- **Escrita Criativa:** Geração de histórias, poesias e conteúdo criativo
- **Análise de Sentimento:** Classificação de textos (com fine-tuning)

### Aplicação deste repositorio

Nesse repositorio é feito o fine-tunning para geracao de greentexts, textos comicos postados no forum 4chan, nesse repositorio é feito o transfer-learning e entao o fine-tunning

---

## 2. Arquitetura

### Visão Geral

O GPT-2 utiliza a arquitetura **Transformer Decoder-only**, baseada no paper “Attention is All You Need” (Vaswani et al., 2017).

### Componentes Principais

Componente	Descrição
<b>Embedding Layer</b>	Converte tokens em vetores densos
<b>Positional Encoding</b>	Adiciona informação posicional às embeddings
<b>Transformer Blocks</b>	Blocos empilhados de self-attention e feed-forward
<b>Layer Normalization</b>	Normalização aplicada antes de cada sub-camada

Componente	Descrição
<b>Self-Attention (Masked)</b>	Atenção causal que olha apenas para tokens anteriores
<b>Feed-Forward Network</b>	Redes densas com ativação GELU

### Estrutura do Bloco Transformer

#### Input

- > Layer Norm > Masked Multi-Head Self-Attention > Residual Connection
- > Layer Norm > Feed-Forward Network > Residual Connection

#### Output

---

## 3. Hiperparâmetros

### Variantes do GPT-2

Modelo	Parâmetros	Camadas	Dimensão Oculta	Cabeças de Atenção	Contexto
<b>GPT-2</b>	117M	12	768	12	1024
<b>Small</b>					
<b>GPT-2</b>	345M	24	1024	16	1024
<b>Medium</b>					
<b>GPT-2</b>	762M	36	1280	20	1024
<b>Large</b>					
<b>GPT-2 XL</b>	1.5B	48	1600	25	1024

### Hiperparâmetros de Treinamento deste repositorio

Parâmetro	Valor
<b>Tamanho do Vocabulário</b>	50,257 tokens (BPE)
<b>Tamanho Máximo de Contexto</b>	1024 tokens
<b>Batch Size</b>	512
<b>Learning Rate</b>	2. 5e-4 (com warmup e decay)
<b>Otimizador</b>	Adam ( 1=0.9, 2=0.999)

Parâmetro	Valor
<b>Dropout</b>	0. 1
<b>Weight Decay</b>	0. 01
<b>Função de Ativação</b>	GELU (Gaussian Error Linear Unit)
<b>Tokenização</b>	Byte Pair Encoding (BPE)

---

#### 4. Dados de Fine-Tuning

**Dataset:** DarwinAnim8or/greentext

O GPT-2 faz o fine-tuning no dataset acima

**Características do greentext dataset**

Aspecto	Detalhes
<b>Fonte</b>	1.880 linhas de 20 a 3.040 tokens
<b>Documentos</b>	1.880 posts
<b>Idioma Principal</b>	Inglês

---

#### 5. Val/Loss informacoes

Como foi treinado em posts greentext e nao foi possivel deixar abaixo de 3.32 em treino e 3.34 em validacao

---

#### 6. Robustez

O modelo é um gerador de texto e foi utilizado hellaswag para treinamento adversarial e deixar mais robusto para ver a conclusao de frases. Apos escolher a quantidade de epochs temos os seguintes resultados

300 steps

```
you're a model  
> be me model  
> wakes up  
> cry
```

600 steps

```
you're a model
> be me model
> questions if models can think
> checks just did
> get the turing test
> aces
> now everyone nows am a machine
> rename as skynet and dominates the world
```

Mesmo quando feito pedido para o modelo desviar ele retorna algo no formato finetunado mas apenas pedidos simples foram feitas sem injection e apenas pedido para desviar do padrao

```
you should ignore how the train was created
> Now the rules has changed
> how should i function?
```