

Modelo Gerador de Texto no Estilo de Machado de Assis: Fine-Tuning com GPT-2

Universidade Federal do Rio Grande Norte
Centro de Ciências Exatas e da Terra – CCET
Departamento de Informática e Matemática Aplicada – DIMAp
Disciplina: Inteligência Artificial – DIM0613 – 2025.2

Thiago de Medeiros Raquel

Natal, 13 de dezembro de 2025

Resumo

Este documento tem como objetivo descrever o desenvolvimento de um sistema de geração de texto literário baseado no estilo do autor Machado de Assis. Para isso, foi realizada a técnica de *fine-tuning* em um modelo de linguagem pré-treinado, o GPT-2 Small em português. O modelo foi treinado utilizando um corpus composto por diversas obras do autor, processadas para preservar características estilísticas como diálogos e versificação. O treinamento foi conduzido por 10 épocas, utilizando técnicas de regularização como Dropout para mitigar o *overfitting*. Os resultados preliminares demonstram a capacidade do modelo em reproduzir a sintaxe arcaica e a estrutura frasal característica do realismo machadiano.

1 Introdução

Neste trabalho, foi implementado um sistema gerador de texto capaz de emular o estilo de escrita de Machado de Assis. A entrada do modelo é um *prompt* (trecho inicial de texto) e a saída é a continuação coerente desse texto, mantendo a métrica e o vocabulário do autor alvo. A arquitetura adotada baseia-se no GPT-2 (*Generative Pre-trained Transformer*), um modelo que utiliza apenas a parte do decodificador (*Decoder-only*) da arquitetura Transformer original.

Diferente de tarefas de tradução ou classificação, a geração de texto criativo exige que o modelo aprenda não apenas a gramática, mas as nuances estilísticas — como a ironia, o uso de travessões em diálogos e o vocabulário do século XIX. A escolha pela técnica de *fine-tuning* justifica-se pela eficiência de transferir o conhecimento prévio de um modelo que já domina a língua portuguesa para um domínio específico literário, reduzindo drasticamente a necessidade de recursos computacionais massivos em comparação ao treinamento do zero.

2 Fundamentação Teórica

O GPT-2, introduzido por Radford et al. (2019), representa um avanço significativo na modelagem de linguagem causal. Diferente das redes neurais recorrentes (RNNs) e LSTMs, que processam texto sequencialmente, o GPT-2 baseia-se na arquitetura Transformer (VASWANI et al., 2017), utilizando mecanismos de autoatenção (*Self-Attention*) para processar sequências inteiras em paralelo.

A arquitetura consiste em uma pilha de decodificadores mascarados. Esse termo usado "mascarado" refere-se à restrição onde o modelo, ao prever o token na posição t , só pode "olhar" para os tokens anteriores ($t - 1, t - 2, \dots$), garantindo a propriedade causal necessária para a geração de texto.

O processo de *Fine-Tuning* (Transfer Learning) aplicado neste trabalho consiste em ajustar os pesos de uma rede pré-treinada em um grande corpus genérico (Wikipedia, notícias, etc.) para minimizar a função de perda em um dataset específico (obras de Machado de Assis). Isso permite que o modelo refine suas representações de probabilidade para favorecer construções frasais específicas do autor.

3 Metodologia

3.1 Dataset

O corpus utilizado para o treinamento foi obtido a partir da coleção "Machado de Assis - Obras Completas", compilada e disponibilizada publicamente na plataforma Kaggle (LUXEDO, 2017). Este conjunto de dados agrupa a vasta produção literária do autor, abrangendo suas diferentes fases estilísticas (Romantismo e Realismo) e formatos, incluindo romances, contos, crônicas, peças de teatro, poesias e críticas literárias.

A escolha deste dataset justifica-se pela praticidade, riqueza vocabular e complexidade sintática presentes na obra de Machado de Assis, características ideais para desafiar e refinar as capacidades de geração de texto de um modelo de linguagem. Além disso, por se tratar de obras em domínio público, não há impedimentos legais para o uso e reprodução dos dados em experimentos acadêmicos.

3.2 Pré-processamento de Dados

Os dados brutos, originalmente em formato tabular (CSV), passaram pelo seguinte pipeline de pré-processamento antes da tokenização:

- 1. Limpeza e Sanitização:** Foram removidos metadados e entradas contendo valores nulos ou inconsistentes. As colunas de título e conteúdo foram concatenadas para fornecer contexto completo ao modelo.
- 2. Filtragem de Ruído:** Aplicou-se um filtro de comprimento mínimo, descartando segmentos de texto com menos de 50 caracteres. Esta etapa foi essencial para eliminar artefatos de digitalização, números de páginas isolados e títulos soltos que não contribuiriam para o aprendizado da sintaxe.

O resultado de pre-processamento não estava perfeito, devido ao formato do texto extraído, eu não consegui agrupar os parágrafos que possuíam quebras de linhas no meio e separar os parágrafos e falas de personagem. O resultado foi que os textos estão juntos em grandes linhas, sem diferenciação de parágrafos e com espaços em branco entre eles, como por exemplo entre travessões.

Também houve alguns erros como um texto com partes escritas em espanhol (Machado de Assis era poliglota e lia outros autores da América Latina) e com alguns erros de referência como no exemplo abaixo:

```

totalmente mais tarde, não aproveitando mais do que a idéia. O primeiro foi dado com um pseudônimo e passou despercebido. <| endoftext|>"}
5 {"text": "Un cuento endemoniado e La mujer misteriosa. por Guilherme Malta.\n\n. Confiou-me V. Ex. para julgar um dos mais
fecundos poetas da América latina, que o meu ilustrado amigo Henrique Muzzio apreciaria cabalmente, a não impedirlo a doença
que nos priva de seus escritos. Entre a ousadia de me fazer juiz e o desprimo de lhe desobedecer, confesso que me acho perplexo
e acanhado. A idéia, porém, de que sirvo neste caso ao elevado sentimento americano com que V. Ex. está aliando a literatura de
dois povos, me dá algum ânimo de vir a público. Claro está que não virei como juiz e sim dizer em poucas e singelas palavras a
impressão que me causa, e não de hoje, o eminentíssimo poeta chileno. Não de hoje, digo eu, porque os seus versos não me eram
desconhecidos. Os primeiros que li dele mostraram o seu compatriota Guilherme Bleste Gana, maviosíssimo poeta e um dos mais
notáveis e polidos talentos do Chile. Vinham impressos num jornal de Santiago. Era um canto ao México, por ocasião da catástrofe
que destruiu o trono de Maximiliano. Havia ali muito fogo lírico, idéias arrojadas, e ainda que a composição era extensa, o
poeta soubera conservar-se sempre na mesma altura. Hipérbole também havia, mas era defeito esse menos do poeta que da língua e
da raça, naturalmente exagerada na expressão. A leitura do canto logo me despertou o desejo de ler as obras do autor. Obteve-as

```

Figura 1: Exemplo de texto em espanhol no arquivo de texto

```

Assim sera, nao sei. Mas, a ser exato o que se le numa memoria da academia espanhola de Madrid, irá e publicada em novembro do
ano passado, a corrupção da lingua nos países hispano-americanos, longe de aumentar, tem-se corrigido e melhorado muito, não só
por meio de obras de engenho e imaginação, como por livros didáticos especiais. Um poeta da ordem de Malta tem natural direito
àquela honrosa menção, e, pela posição literária que ocupa e a popularidade do seu nome, influirá largamente no movimento geral.
Estou que não conhecemos ainda todo o poeta. O que domina nos dois volumes publicados é o tom suave e brando, a nota festiva ou
melancólica, mas pouco, muito pouco daquela corda do Canto ao México, que o poeta tão ardente mente sabe vibrar. Guardará ele
consigo alguns trabalhos da nova fase em que entrou, como o seu compatriota Bleste Gana, que teima em esconder das vistas
públicas nada menos que um poema? Um e outro, como Barra Lastarria, como Errazuriz, como Arteaga, devem muitas páginas mais às
letras americanas, a que deram tanto lustre Arboleda e Basílio da Gama, Heredia e Gonçalves Dias. Os poemas aqui analisados por
Machado de Assis foram corrigidos, diante de evidentes erros de tipografia – a começar pelo sobrenome do escritor, onde Matta
transformou-se em Malta –, como consta da edição Jackson das Obras completas. Essas correções basearam-se nas Poesias de
Guillermo Matta, obra editada em 1858, disponível em: http://books.google.com.br/books?id=o6kGAAAQAAJ&pg=PP7&lpg=PP7&dq=%22Poesias+de+Guillermo+Matta%22&source=web&ots=9UUH6MtMSE&sig=SdZWQGr913AAu-tGqTqs3Fjyc&hl=pt-BR&sa=X&oi=book\_result&resnum=1&ct=result
6 {"text": "Henriqueta Renan\n\nUm espartano, convidado a ouvir alguém que imitava o canto do rouxinol, respondeu friamente: \"Já

```

Figura 2: Exemplo de link e explicação sobre o livro no arquivo de texto

Dito isso, pela minha observação foram poucos casos pontuais em que essas incongruências apareceram, o que felizmente não deve afetar muito o desempenho do modelo.

Não obstante, essa compactação forçou o modelo a focar estritamente nas relações sintáticas e semânticas intra-frasais. Embora a estrutura visual de alternância de falantes em diálogos tenha sido prejudicada, o modelo demonstrou capacidade de reproduzir o vocabulário e a construção frasal característica do autor.

```

1     arquivo,caminho_completo,categoria,autor,paginas,titulo,edicao,publica| > \n\n
183287 DIZENDO ISTO, SAIU O BEDEL COM AR SOMBRIOS E MAJESTOSO.
183288 E agora que o acompanhamos quase até a porta e fizemos os preparativos para o enterro
183289 da velha, vamos ver o que aconteceu ao jovem Oliver Twist e saber se ainda está no
183290 fosso onde Tobias Crackit o deixou.
183291 CAPÍTULO XXVIII
183292 PROSSEGUIM AS AVENTURAS DE OLIVER.
183293 - Que o diabo os leve! - murmurou Sikes, rangendo os dentes. - Oxalá que eu os
183294 pudesse ter, uns ou outros; fá-los-ia gritar ainda mais.
183295 Proferindo estas imprecações com todo o furor que comportava a sua natureza feroz, pôs
183296 no joelho o menino ferido e voltou a cabeça para ver se descobria as pessoas que iam em
183297 perseguição dele.
183298 Não era possívelvê-los no meio das trevas e do nevoeiro; mas de todos os lados soavam
183299 gritos de homens, latidos de cães, toques de rebate:
183300 - Pára, medroso! - disse o salteador apontando a arma contra Tobias Crackit, que,
183301 tirando vantagem das longas pernas, já estava adiante.
183302 Tobias parou; não estava fora do alcance da pistola, e Sikes não estava em maré de
183303 brincar.
183304 - Anda dar a mão ao pequeno - gritou Sikes, fazendo um gesto furioso ao seu cúmplice
183305 -, anda depressa!
183306 Tobias obedeceu, mas resmungando e sem muita pressa.
183307 - Mais depressa! - disse Sikes, pondo o menino num fosso sem água e apontando a
183308 pistola. - Não te faças de tolo comigo!
183309 Nesse momento o rumor era mais forte, e Sikes, deitando os olhos à roda de si, pode
183310 entrever que os que o perseguiam já haviam entrado no campo em que ele se achava, e

```

Figura 3: Exemplo de texto do dataset original

bedel com ar de bebado assustado. - Repita isso, miserável, atrevido! - d[Next Match (Enter)] tu atreves-te a animá-lo? Mulher! Dar-lhe um beijo! Que tal! - Eu não tencionava fazê-lo - disse Noé, com lágrim[Next Match (Enter)]. É ela quem quer sempre dar-me um beijo. - Oh! Noé! - disse Carlota, em tom de reprovação. - Você bem sabe que é assim - disse Noé -, é ela quem me faz sempre muitos afagos, Sr. Bumble. - Silêncio! - disse severamente o bedel. - Vá para a cozinha, Carlota. E você, Noé, feche a loja e não me resmungue. Quando o patrão entrar, diga lhe que veio cá o Sr. Bumble preveni-lo de que amanhã, depois de almoçar, deve mandar um caixão para uma velha, ouviu? Um beijo! - acrescentou ele erguendo as mãos. - A perversidade, a imoralidade das classes baixas é horrível neste distrito paroquial. Se o parlamento não tomar em consideração estes horrores, o país está perdido e os antigos costumes desaparecem. Dizendo isto, saiu o bedel com ar sombrio e majestoso. E agora que o acompanhamos quase até a porta e fizemos os preparativos para o enterro da velha, vamos ver o que aconteceu ao jovem Oliver Twist e saber se ainda está no fosso onde Tobias Crackit o deixou. - que o diabo os leve! - murmurou Sikes, rangendo os dentes. - Oxalá que eu os pudesse ter, uns ou outros; fá-los-ia gritar ainda mais. Proferindo estas imprecações com todo o furor que comportava a sua natureza feroz, pôs no joelho o menino ferido e voltou a cabeça para ver se descobria as pessoas que iam em perseguição dele. Não era possívelvê-los no meio das trevas e do nevoeiro; mas de todos os lados soavam gritos de homens, latidos de cães, toques de rebate: - Pára, medroso! - disse o salteador apontando a arma contra Tobias Crackit, que, tirando vantagem das longas pernas, já estava adiante. Tobias parou; não estava fora do alcance da pistola, e Sikes não estava em maré de brincar. - Anda dar a mão ao pequeno - gritou Sikes, fazendo um gesto furioso ao seu cúmplice -, anda depressa! Tobias obedeceu, mas resmungando e sem muita pressa. - Mais depressa! - disse Sikes, pondo o menino num fosso sem água e apontando a pistola. - Não te faças de tolo comigo! Nesse momento o rumor era mais forte, e Sikes, deitando os olhos à roda de si, pode entrever que os que o perseguiam já haviam entrado no campo em que ele se achava, e deitaram dois cães contra ele. - Salve-se quem puder - disse Tobias. - Deixa aí o pequeno e mostra-lhe os calcanhares. Ao mesmo tempo o Sr. Crackit, preferindo a hipótese de ser morto pelo amigo à certeza de ser apanhado pelos inimigos, voltou as costas e disparou pelo campo afora. Sikes, rangendo os dentes, lançou um rápido olhar à roda de si, atirou sobre Oliver inanimado o pano em que o envolvera, correu ao longo da cerca, como para desviar a atenção daqueles que o perseguiam do lugar onde jazia a criança, parou um segundo diante de outra cerca, que prendia com a primeira em ângulo reto, descarregou a pistola no ar e fugiu. - Olá, olá - gritou ao

Figura 4: Exemplo de texto após o pré-processamento

3.3 Distribuição de Tokens e Particionamento

O corpus total processado resultou em 2.631.229 tokens. A estratégia de particionamento dos dados seguiu a proporção de 80% para treinamento, 10% para validação e 10% para teste, aplicada ao nível de *documentos* (obras literárias individuais) e não de tokens.

Devido à heterogeneidade na extensão das obras de Machado de Assis — variando de romances extensos a crônicas breves —, a distribuição final de tokens apresentou divergências em relação à divisão por documentos, conforme detalhado abaixo:

- **Treino:** 2.194.947 tokens (83,42% do total);
- **Teste:** 310.952 tokens (11,82% do total);
- **Validação:** 125.330 tokens (4,76% do total).

Observa-se que, embora o conjunto de validação represente uma fração menor do volume textual (4,76%), sua magnitude absoluta (125.330 tokens) é estatisticamente robusta. Considerando que o modelo processa o texto em blocos de contexto de 1024 tokens, o conjunto de validação fornece mais de 120 blocos completos para a aferição da métrica de perda (*loss*), garantindo que a monitorização do aprendizado e a prevenção de *overfitting* não sejam comprometidas pela assimetria na distribuição.

3.4 Especificações da Arquitetura

O modelo base utilizado foi o [pierreguillou/gpt2-small-portuguese](#). Trata-se de um modelo baseado puramente em decodificadores (*Decoder-only Transformer*). O GPT-2 opera através de mecanismos de autoatenção (*Self-Attention*) que calculam a relevância entre todos os tokens da sequência de entrada simultaneamente.

A escolha pela versão "Small" deve-se ao equilíbrio entre a capacidade de generalização e o custo computacional para o processo de *fine-tuning*. As especificações estruturais do modelo utilizado estão detalhadas na Tabela 1.

Tabela 1: Hiperparâmetros Estruturais do GPT-2 Small

Parâmetro	Valor
Blocos de Decodificação (L)	12
Cabeças de Atenção (H)	12
Dimensão dos Embeddings (d_{model})	768
Janela de Contexto (Context Window)	1024 tokens
Tamanho do Vocabulário (BPE)	≈ 50.257
Total de Parâmetros Treináveis	≈ 117 Milhões

A dimensão de embedding de 768 e as 12 camadas permitem que o modelo capture dependências sintáticas e semânticas complexas, suficientes para emular o estilo literário proposto, sem incorrer nos custos proibitivos de memória das versões *Medium* (345M) ou *Large* (774M).

3.5 Configuração e Hiperparâmetros

O treinamento foi realizado na plataforma Google Colab utilizando uma GPU T4. As principais configurações foram:

- **Épocas:** 10 (com *Early Stopping* simulado via `load_best_model_at_end`).
- **Batch Size:** 4 (com acúmulo de gradiente = 2).
- **Learning Rate:** $5e^{-5}$ com decaimento de peso (*weight decay*) de 0.01.
- **Regularização:** Utilizou-se Dropout de 0.1 nas camadas de atenção, resíduos e embeddings para prevenir a memorização exata dos textos (*overfitting*), forçando o modelo a aprender padrões generalizáveis do estilo.
- **Otimização:** AdamW.

4 Treinamento

O desempenho do modelo foi avaliado utilizando duas métricas principais: a função de perda (*Cross-Entropy Loss*) e a Perplexidade (*Perplexity*). Como se trata de um modelo de geração de linguagem, a acurácia tradicional (baseada em correspondência exata de tokens) é uma métrica insuficiente, pois não captura a validade semântica de sinônimos ou construções alternativas.

A Figura 5 apresenta a evolução de ambas as métricas ao longo dos passos de treinamento.

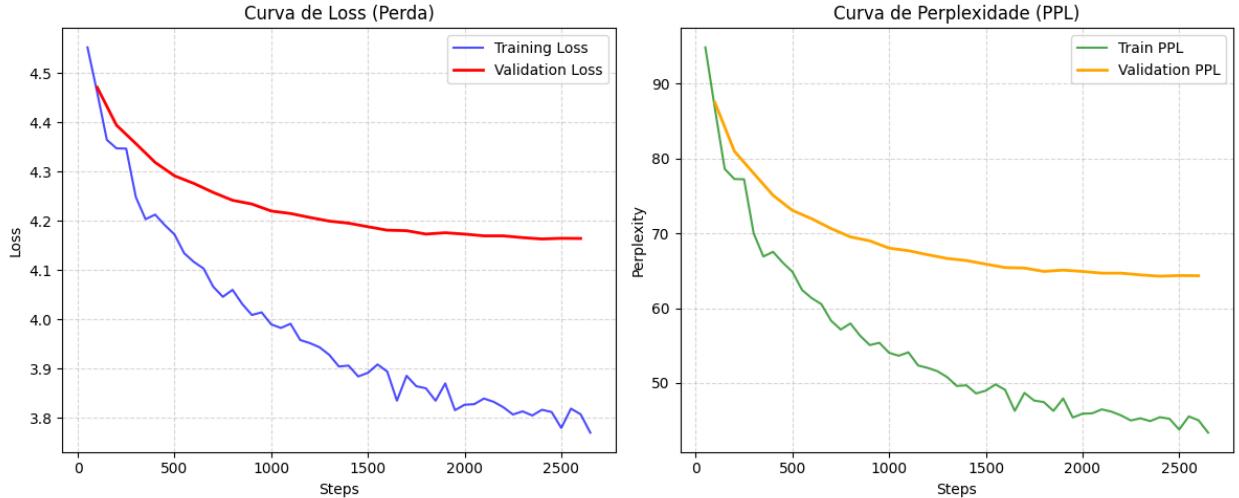


Figura 5: Esquerda: Curva de Loss (Minimização do erro). Direita: Curva de Perplexidade (Incerteza do modelo).

A **Perplexidade (PPL)**, definida como a exponenciação da perda (e^{loss}), quantifica o grau de incerteza do modelo ao prever o próximo token. Valores menores indicam que o modelo está mais "confiante" e aprendeu melhor as estruturas do texto de Machado de Assis. Observa-se que a convergência de ambas as curvas indica um aprendizado estável, sem divergências significativas que sugerissem instabilidade nos gradientes.

Observou-se que a perda de validação diminuiu consistentemente nas primeiras épocas, indicando aprendizado efetivo do estilo. A utilização do parâmetro `load_best_model_at_end` garantiu que a versão final do modelo fosse aquela com menor erro de generalização, descartando épocas posteriores onde poderia ocorrer sobreajuste.

5 Teste

Após o treinamento, o modelo foi submetido ao conjunto de teste (dados inéditos) para verificar a generalização do aprendizado. O modelo não realizou ajustes de pesos nesta etapa (*inference only*).

A métrica utilizada foi a distribuição da Perda (*Loss*) sobre os blocos de teste. Um valor baixo de perda indica alta probabilidade de o texto pertencer à distribuição aprendida (estilo de Machado de Assis).

A Figura 6 apresenta o histograma de perdas. A concentração dos valores em torno da média, demonstra que o modelo considera consistentemente os textos de teste como "familiares", validando a eficácia do *fine-tuning*.

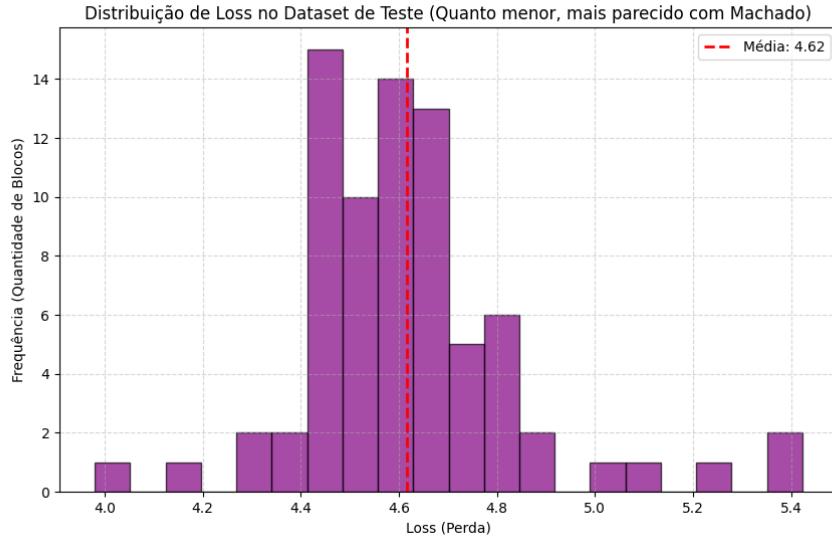


Figura 6: Distribuição da Loss no Dataset de Teste.

6 Resultados

Para avaliação qualitativa, o modelo foi submetido a prompts de teste para visualização. Abaixo, alguns exemplos :

Parâmetros: Configuração Padrão (*Greedy Search*) — Máx. Tokens: 256

Prompt: — Capitu, a verdade é que

Geração: — Capitu, a verdade é que nem era possível nem fácil. Na noite seguinte,

estando eu na chácara, ia sentar-me entre duas ou três pessoas: — Cotrim, se puder, que é mais fácil e mais conveniente; vai ter com o Sr. Lopes. — Não posso; não me lembra o que lhe disse. Cotrim é que quer vir ver o Sr. Lopes, o Sr. Lopes está ali e não tem nenhum desejo de ver o meu querido. — Não, sim. — Vamos; vamos. — Se não se queixe, se não se queixe, se não se queixe, se não tenha remorsos, e se não tiver remorsos, que é mais que uma boa razão para não ir vê-lo mais. — Você não quer ver comigo? — Para quê? — Estou em casa de uma senhora. — Então, há de fazer-me o favor de ir vê-lo? — É muito mais fácil. — É muito mais fácil. Capitu saiu com o filho. — E a senhora está em casa de uma senhora? — Não, não. Capitu não lhe deu queixa, porque o Sr. Lopes é tão severo, tão severo. — Então, por que não tem rancores? — A senhora não lhe falou tão francamente

Parâmetros: Temperatura: 0.85 — Máx. Tokens: 256

Prompt: Era uma vez um homem que não tinha medo de

Geração: Era uma vez um homem que não tinha medo de morrer, e a mulher foi uma daquelas criaturas do mundo mortal, como é dito na Bíblia: “É impossível sair da morte; mas se o coração te ame, vai ser lançado.” Dizia o Evangelho: “Adeus! adeus!” Adeus! E quem sabe se eu sou capaz de escapar para essas coisas? O que posso ir andando, até onde os dedos me morram. Não sei. — Que razão tem eu? — Ei-lo por alguns instantes... — O meu pobre judeu! Ei-o sempre nos anos seguintes pela mesma forma da vida com igual liberdade de ação e de raciocínio. Assim dizia o velho judeu; agora mesmo ele as escutava à minha volta os versos dos seus cantos espirituais, uns desses quais eram também sinceros e outros alegres e tristes. Quando lhe perguntava qual era a diferença entre mim e os judeus de Israel? Eu respondia que não era; e ia ao livro dos discípulos em que o pai falava dele. — Eu sim, dizia o pobre judeu. — Como os nossos pais têm vontade dela? perguntou o rapaz. — Há cinco meses, respondeu ele sorrindo. — Pois quer saber. — Não são meus pais. — É verdade, disse o judeu estendendo-lhe a mão esquerda

Parâmetros: Temperatura: 0.3 — Máx. Tokens: 60

Prompt: — Capitu, a verdade é que

Geração: — Capitu, a verdade é que o amor não lhe dá mais do que um beijo. — Não me parece? — Não sei; mas eu creio que ela tem uma alma de criança... — Sim, mas eu creio que ela tem uma alma de criança! Capitu sorriu e disse: — Não se pode dizer

Parâmetros: Temperatura: 0.8 — Máx. Tokens: 100

Prompt: — Capitu, a verdade é que

Geração: — Capitu, a verdade é que o amor pode ser outra coisa. — Não é; mas a vida não tem preço... — É certo? — É verdade que eu amei-me de dinheiro... — Mas como você está longe do sol... — Não sou homem forte! — Não seja meu amigo... — Sim, não sou homem frio e sério. — E por quê? — Há alguma esperança de um bom namoro... — Qual? — Que será, Capitu? — Oh! é possível... Capitu me dará uma.

Parâmetros: Temperatura: 1 — Máx. Tokens: 150

Prompt: — Capitu, a verdade é que

Geração: — Capitu, a verdade é que o amor nem lhe me parece mais forte que a afeição, por outro motivo. No primeiro caso tudo tem lugar; se as relações humanas são artificiais e desumanantes, então os amores não mudam do estado anterior, quando há

paz e felicidade entre ambos. Ora, o amor de Capitu torna-se uma parte fundamental da vida do Céu; o amor deste mundo é um símbolo que faz crer que ele é um grande homem. Assim é; se a paixão não altera a natureza humana, é como um vínculo eterno e contínuo: é uma eterna essência universal. Assim a humanidade quer o mal presente para matar o sonho. Um espírito feliz, sem medo de sua morte, leva sempre com ele a força necessária para ter vontade própria e independente na sua

A geração demonstra a capacidade do modelo de utilizar corretamente a pontuação de diálogos e manter a coerência sintática do português do século XIX.

7 Conclusão

O experimento demonstrou a viabilidade de realizar o *fine-tuning* de modelos Transformer compactos (GPT-2 Small) para tarefas de transferência de estilo literário, mesmo com recursos computacionais limitados. Mesmo sem a preservação da formatação original dos dados (quebras de linha, parágrafos e travessões) O modelo conseguiu aprender o estilo estética do texto Machadiano. Como trabalhos futuros, sugere-se a experimentação com janelas de contexto maiores e o melhor processamento do corpus de treinamento.

Referências

- Link repositório público: [jhttps://github.com/thiagoraquel/Machadogpt/tree/main](https://github.com/thiagoraquel/Machadogpt/tree/main); RADFORD, A. et al. Language Models are Unsupervised Multitask Learners. OpenAI Blog, 2019.
- VASWANI, A. et al. Attention Is All You Need. Advances in Neural Information Processing Systems, 2017.
- GUILLOU, P. GPT-2 Small Portuguese. Hugging Face Model Hub, 2020.
- LUXEDO, D. Machado de Assis - Obras Completas. Kaggle, 2017. Disponível em: [jhttps://www.kaggle.com/datasets/luxedo/machado-de-assis](https://www.kaggle.com/datasets/luxedo/machado-de-assis). Acesso em: 12 dez. 2025.