UM ESTUDO DA TRANSFERÊNCIA DE APRENDIZADO EMPREGANDO MODELOS DE APRENDIZADO PROFUNDO: CONCEITOS, TÉCNICAS E APLICAÇÕES

Igor Vinicius de Jesus Gama da Silva[1] (IC), Isabela Neves Drummond (PQ)[1]
[1]*Universidade Federal de Itajubá.*

Palavras-chave: Aprendizado de máquina. *Fine-tuning*. GPT. Inteligência artificial.

Introdução

A Inteligência Artificial (IA) está cada vez mais presente em diversos contextos, desde assistentes virtuais e chatbots até sistemas educacionais. Em cada contexto, a capacidade de fornecer respostas precisas e contextualizadas é crucial para o sucesso e a eficácia dos sistemas de IA, portanto, os robôs inteligentes não se limitam apenas a substituir humanos em tarefas, mas também colaborar por meio de interfaces inteligentes homem-máquina, evolução que é especialmente relevante na era da Indústria 4.0 (BAHRIN et al., 2016).

No entanto, adaptar essas tecnologias a diferentes contextos educacionais ainda é um grande desafio, pois embora os Modelos de Linguagem de Grande Escala (LLMs), como o GPT, tenham uma capacidade para gerar texto coeso e compreender o contexto de interações, seu *fine-tuning* (ajuste fino) para aplicações específicas, exige um entendimento profundo das estratégias de interação humano-computador. Como mencionado por Kumar (2023), o sucesso da implementação dessas ferramentas depende da adaptação cuidadosa aos diferentes cenários de uso, garantindo que o modelo atenda às necessidades e expectativas dos usuários de forma eficaz.

Por outro lado, os Ambientes Virtuais de Aprendizagem (AVAs) já são amplamente utilizados para oferecer uma educação mais acessível e interativa, mas não sem dificuldades. Segundo Hamutoglu (2020), muitos alunos enfrentam barreiras ao tentar localizar informações relevantes nesses ambientes, o que pode impactar diretamente seu desempenho e engajamento, dificuldades que, frequentemente, resultam de interfaces pouco intuitivas ou da ausência de ferramentas eficazes para filtrar e organizar o vasto volume de conteúdo disponível.

Este estudo investiga o uso de modelos de aprendizado profundo com foco na técnica de transferência de aprendizado para otimizar assistentes em plataformas de ensino. O objetivo principal foi explorar como personalizar esses modelos, com foco na família GPT, para melhorar a experiência dos usuários em ambientes de aprendizado virtual, buscando personalizar o atendimento aos alunos, além disso, a pesquisa também analisou o impacto do hiperparâmetro de temperatura na qualidade das respostas geradas.

Metodologia

A metodologia empregada nesta pesquisa seguiu duas etapas principais, focadas no estudo dos modelos de aprendizado profundo da família GPT, com especial atenção ao *fine-tuning* e à avaliação do comportamento dos modelos em resposta a diferentes parâmetros.

Na primeira etapa, foi realizada uma análise detalhada dos modelos de linguagem da família GPT, começando pelo GPT-2. A escolha inicial desse modelo foi motivada por sua acessibilidade, uma vez que é de código aberto e amplamente utilizado para tarefas de Processamento de Linguagem Natural (NLP). O modelo foi carregado utilizando a plataforma *Hugging Face*, que oferece um vasto repositório de modelos para tarefas de processamento de Linguagem Natural e ferramentas para ajuste e integração em aplicações (WOLF, 2019).

Portanto, o modelo GPT-2 *Large*, com aproximadamente 774 milhões de parâmetros, foi inicialmente estudado devido à sua capacidade de gerar respostas mais coerentes e contextualizadas em comparação com submodelos menores (GPT-2 *Small* e *Medium*). Contudo, uma limitação encontrada foi a linguagem em que o modelo foi treinado, o inglês. Diante disso, foi necessário utilizar a biblioteca *Googletrans* para traduzir as interações de entrada e saída (de português para inglês e vice-versa), permitindo que as interações fossem realizadas em português, atendendo ao público-alvo do estudo.

Além disso, foram realizados testes com o GPT-2 *Small*, a partir de um modelo treinado diretamente em português, eliminando a necessidade de tradução e permitindo uma

análise mais fluida. Entretanto, por possuir apenas 124 milhões de parâmetros, esse modelo apresentou algumas limitações em termos de robustez e adequação linguística.

Dessa forma, ainda com as limitações, o modelo GPT-2 *Small* mostrou ser mais eficiente nos experimentos, devido a possíveis erros de tradução com o modelo do GPT inglês, dessa forma, o *fine-tuning* foi realizado no modelo em português com base em um conjunto de dados customizado, utilizando a função *Trainer* e *TrainingArguments* da biblioteca *Transformers* no *Google Colab*, plataforma que oferece recursos como GPUs (Unidades de Processamento Gráfico) e TPUs (Unidades de Processamento Tensorial) para acelerar o processo de treinamento, permitindo que usuários prototipem soluções de IA sem precisar de configuração local (BISONG et al., 2019). Após o treinamento, o modelo foi testado por meio de uma interface de chat.

Simultaneamente ao estudo dos modelos, foi conduzido um experimento sobre o efeito do hiperparâmetro temperatura na geração de respostas pelos modelos GPT. Esse hiperparâmetro é fundamental no controle da aleatoriedade do texto gerado, portanto, foram testadas diferentes configurações de temperatura, variando entre 0.1 e 2.0, onde em temperaturas mais baixas, as respostas devem ser previsíveis e focadas, enquanto temperaturas mais altas geram maior variabilidade, mas podem comprometer a coerência das respostas.

O impacto da temperatura foi avaliado com base em três métricas principais:

- Perplexidade: Mede a confiança do modelo nas suas previsões. Valores baixos indicam maior fluência no texto gerado.
- Similaridade de Cosseno: Avalia a semelhança semântica entre textos gerados pelo modelo e um texto de referência, considerando o alinhamento do conteúdo.
- Distância de Edição Normalizada: Quantifica o número mínimo de operações necessárias para transformar uma resposta em outra, sendo útil para avaliar a variabilidade entre as respostas geradas.

Essas métricas foram aplicadas em diferentes perguntas para avaliar a resposta dos modelos em termos de criatividade, coerência e precisão, como nas perguntas: "*What do you imagine technology will be like in 100 years?*" (criatividade), "*How do you define artificial intelligence?*" (definição técnica), e "*Who was Leonardo da Vinci?*" (informação factual). Assim os testes foram conduzidos no *Google Colab*, que permitiu a criação de scripts para automatizar o processo de avaliação das respostas geradas com diferentes temperaturas.

Na segunda etapa, o foco foi realizar uma aplicação real utilizando modelos mais recentes, como o GPT-3.5 e GPT-4, disponíveis na plataforma *OpenAI*. O primeiro objetivo foi realizar o *fine-tuning* dos modelos mais antigos para avaliar suas capacidades, como o *Davinci-002* e o *Babbage*, além do GPT-3.5, para testar sua aplicabilidade no contexto educacional, onde a principal função do *fine-tuning* foi adaptar esses modelos para gerar feedback personalizado em ambientes virtuais de aprendizado (AVA).

Além do *fine-tuning*, nessa etapa a pesquisa também avançou para a implementação de assistentes virtuais na plataforma da *OpenAI* utilizando modelos mais recentes, como o GPT-3.5 e o GPT4. Em vez de limitar o treinamento a perguntas e respostas pré-formatadas, foi adotada uma abordagem mais flexível, com documentos específicos em diversos formatos (PDF, DOCX, etc.) sendo inseridos no sistema.

Desse modo, foi aplicada a técnica de engenharia de *prompt*, fornecendo ao modelo instruções detalhadas sobre como agir e responder com base nos documentos fornecidos, diante disso, foi crucial a utilização da funcionalidade de *file search* em conjunto com o *retrieval*, um processo de recuperação de informações relevantes a partir de uma base de dados, o que permite ao modelo buscar informações diretamente dos arquivos carregados, sem depender exclusivamente da base de conhecimento pré-existente do modelo.

A partir disso, um experimento controlado foi realizado envolvendo dois grupos de alunos: o grupo experimental, que utilizou o assistente virtual configurado, e o grupo de controle, que respondeu às mesmas perguntas sem o auxílio da IA. Vale ressaltar que ambos os grupos de alunos não estavam matriculados na disciplina escolhida para a realização do experimento, com o intuito de evitar qualquer viés nas respostas.

Portanto, com o uso da técnica de engenharia de *prompt* junto ao recurso de *retrieval*, crucial para trazer respostas com feedback personalizado, permitiu que a recuperação de informações fosse realizada diretamente de arquivos da disciplina carregados no sistema, conforme explicado anteriormente, o assistente permite a inserção de arquivos em diferentes formatos, que são utilizados para ampliar a

base de dados da IA, fornecendo respostas mais precisas e contextualizadas com o que foi entregue.

A avaliação do assistente virtual foi baseada em dez perguntas operacionais referentes ao contexto da disciplina, diante disso, os alunos responderam às perguntas utilizando o assistente virtual e, posteriormente, as mesmas perguntas foram apresentadas ao grupo de controle sem a IA. As respostas foram avaliadas em termos de tempo de resposta, precisão e qualidade, com o uso de *logs* para acompanhar a interação dos alunos com o assistente e análise das respostas geradas.

Resultados e discussão

Os resultados da pesquisa foram obtidos por meio dos diferentes experimentos realizados com os modelos da família GPT, com as etapas de *fine-tuning*, como também a avaliação do impacto do hiperparâmetro de temperatura e da aplicação com a utilização dos assistentes virtuais implementados com os modelos GPT-3.5 e GPT-4 em ambientes educacionais.

Desse modo, o estudo da temperatura nos modelos GPT-2 revelou que esse hiperparâmetro tem um impacto direto na qualidade das respostas geradas. Em temperaturas baixas (entre 0.1 e 0.5), as respostas apresentaram maior coerência e previsibilidade, sendo adequadas para tarefas que exigem precisão, como o fornecimento de informações factuais em ambientes educacionais. No entanto, conforme a temperatura aumentava (acima de 1.0), as respostas se tornaram mais criativas, mas menos consistentes, mostrando uma variação maior na similaridade semântica e na estrutura textual.

Por exemplo, para perguntas como "*What do you imagine technology will be like in 100 years*?" temperaturas mais altas (próximas de 2.0) resultaram em respostas mais diversas, mas com uma queda na similaridade de cosseno e um aumento na distância de edição, sugerindo que para tarefas que exigem criatividade, o ajuste da temperatura pode ser vantajoso, mas em situações que demandam respostas precisas, como em assistentes virtuais educacionais, valores baixos são mais apropriados.

A análise das métricas, incluindo Perplexidade, Similaridade de Cosseno e Distância de Edição Normalizada, confirmou essa tendência. A Perplexidade aumentou com a elevação da temperatura, enquanto a Similaridade de Cosseno entre as respostas e os textos de referência diminuiu, indicando maior aleatoriedade nas saídas.

Já durante o processo de *fine-tuning*, os modelos *Davinci-002* e *Babbage* apresentaram resultados pouco satisfatórios, pois embora os modelos tenham demonstrado a capacidade de gerar respostas coerentes em alguns casos, houve instabilidade em relação a respostas factuais, especialmente quando se tratava de datas e informações específicas. Esse comportamento foi observado no experimento com perguntas relacionadas a datas de avaliações e atividades acadêmicas, onde o modelo frequentemente gerava respostas incorretas ou inventadas.

Por outro lado, o modelo GPT-3.5, após o *fine-tuning*, apresentou uma melhora significativa em termos de consistência e precisão, além de ter um melhor custo benefício do que o modelo GPT-4, o qual tinha um gasto superior por requisições. Portanto, o uso de prompts detalhados e a técnica de *retrieval* para buscar informações diretamente de documentos melhorou a capacidade do assistente virtual de fornecer respostas baseadas em conteúdo específico, o que foi fundamental para a aplicação no contexto educacional.

Em vista disso, o experimento realizado mostrou que o grupo experimental, o qual utilizou o assistente virtual, apresentou uma precisão média de 86,25%, enquanto o grupo de controle, que respondeu às perguntas manualmente, obteve uma média de 77,5%.

Além disso, os tempos de resposta no grupo experimental foram consistentemente mais baixos, com uma dispersão menor, indicando que o assistente virtual foi capaz de agilizar o processo de busca e entrega de respostas precisas. A Figura 1 que mostra o boxplot gerado para análise do tempo de resposta reforça essa conclusão, mostrando que o uso da IA reduziu significativamente o tempo necessário para que os alunos encontrassem as respostas corretas.

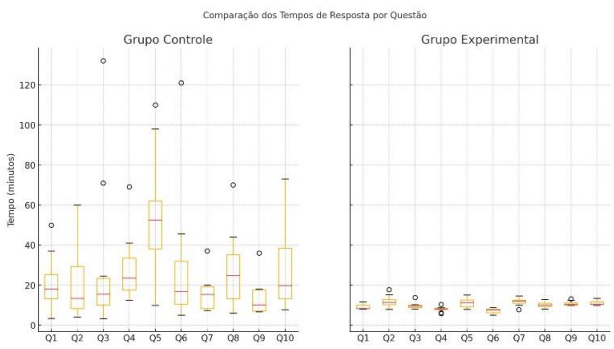

Figura 1 – Boxplot referente ao tempo de resposta dos alunos do Grupo Controle e do Grupo Experimental.

Conclusões

Esta pesquisa se insere no contexto da transferência de aprendizado, dentro de um projeto cujo foco é avaliar o potencial de modelos de aprendizado profundo pré-treinados na solução de problemas. Os resultados alcançados permitiram não apenas concluir as atividades propostas no plano de trabalho, mas também ir além, sugerindo a possibilidade de explorar uma aplicação na área educacional.

Ao longo da pesquisa, as técnicas de *fine-tuning* mostraram-se menos eficientes quando comparadas aos mecanismos de *retrieval* e engenharia de *prompt*, que se destacaram pela capacidade de gerar respostas mais precisas e personalizadas. O ajuste da temperatura também se provou crucial, com temperaturas mais baixas favorecendo consistência em cenários que exigem precisão, e temperaturas mais altas promovendo variabilidade para tarefas que requerem criatividade.

As ferramentas de *retrieval* foram eficazes ao proporcionar uma interação mais fluida entre o usuário e o assistente, melhorando o engajamento dos alunos. Em conclusão, a combinação dessas técnicas oferece uma abordagem promissora para personalizar o aprendizado em AVAs, simulando a interação individualizada de um professor, melhorando a relevância das respostas e tornando o ambiente educacional mais dinâmico e eficaz.

Agradecimentos

Agradeço profundamente aos meus pais e à minha irmã pelo apoio incondicional ao longo de minha jornada. Sou grato também aos meus amigos, que estiveram ao meu lado, oferecendo incentivo e companheirismo. Um agradecimento especial à minha orientadora, Isabela, ao coorientador Flávio e ao professor Adler, cuja confiança e orientação foram fundamentais para o desenvolvimento da minha pesquisa. Por fim, agradeço ao CNPq e à UNIFEI pela oportunidade da bolsa de iniciação científica, que contribuiu significativamente para minha formação.

Referências

BAHRIN, Mohd Aiman Kamarul et al. Industry 4.0: A review on industrial automation and robotic. Jurnal teknologi, v. 78, n. 6-13, 2016.

BISONG, Ekaba et al. Building machine learning and deep learning models on Google cloud platform. Berkeley, CA: Apress, 2019.

HAMUTOGLU, Nazire Burcin et al. Evaluating students experiences using a virtual learning environment: satisfaction and preferences. Educational Technology Research and Development, v. 68, n. 1, p. 437-462, 2020.

KUMAR, Harsh et al. Impact of guidance and interaction strategies for LLM use on Learner Performance and perception. arXiv preprint arXiv:2310.13712, 2023.

WOLF, T. Huggingface's transformers: State-of-the-art natural language processing. arXiv preprint arXiv:1910.03771, 2019.