Relatório Final

Bolsa PIBIC - 5 meses / (04/2024 - 09/2024)

Título do projeto - *Videogames na Instrumentação Biomédica*

João Ricardo Monteiro Scofield Lauar - Engenharia de Controle e Automação

Frederico Caetano Jandre de Assis Tavares - Programa de Engenharia Biomédica/COPPE, Universidade Federal do Rio de Janeiro, RJ, Brasil

Gabriel Casulari da Motta Ribeiro - Tredom Tecnologia Médica Ltda., Rio de Janeiro, Brasil

Resumo - Com os recentes avanços no desenvolvimento dos grandes modelos de linguagem atuais, surge, de forma análoga, um crescente interesse acerca do poder que essas ferramentas podem trazer à vida contemporânea. No entanto, o conhecimento sobre suas possíveis aplicações ainda é nebuloso, especialmente no que se refere à análise de sentimentos. Nesse sentido, esta pesquisa pretende explorar o uso de modelos de geração e análise de imagens para determinar a correlação entre os parâmetros que influenciam a formação do nível de valência das figuras, contribuindo para uma compreensão mais profunda das capacidades e limitações dessas tecnologias emergentes.

I. Introdução

A rápida evolução dos modelos de linguagem de grande escala está impulsionando avanços significativos na análise de imagens. Embora esses modelos consigam gerar respostas quando questionados sobre os sentimentos transmitidos por uma imagem, os fatores que influenciam tais respostas ainda não estão totalmente claros. Para compreender melhor como esses modelos operam e, consequentemente, expandir suas possíveis aplicações, é essencial estudar os parâmetros envolvidos na geração e análise de imagens. Esse entendimento mais profundo pode abrir portas para uma variedade de novos usos e aprimoramentos dessas poderosas ferramentas de inteligência artificial.

Em um trabalho anterior, realizado por outro aluno [1], estudou-se a diferença de complexidade — medida pela taxa de compressibilidade sem perdas — entre imagens geradas com adjetivos emocionais antônimos incluídos no *prompt* — a entrada fornecida ao modelo de linguagem que define a imagem a ser gerada. A partir dos resultados obtidos por essa pesquisa, conseguimos entender que a complexidade das imagens está diretamente relacionada ao adjetivo usado na criação da imagem, gerando diferentes complexidades para diferentes adjetivos. No entanto, será que os resultados observados nessa pesquisa poderiam ser replicados atribuindo "pesos"para os adjetivos utilizados?

II. Objetivos do Projeto

O objetivo central desta pesquisa é estudar os possíveis efeitos do uso de adjetivos emocionais antônimos nas características físicas de imagens sintetizadas por inteligência artificial. Este trabalho visa aprofundar a investigação sobre como esses adjetivos afetam a complexidade das imagens, aplicando novas técnicas de engenharia de prompt, como a utilização de pesos por meio da biblioteca *Compel* [2]. Com isso, visamos esclarecer melhor as potenciais aplicações dessas ferramentas em novas tecnologias e na vida cotidiana.

Esperamos que, ao compreendermos mais profundamente os mecanismos subjacentes aos modelos de geração e análise de imagens, possamos definir e determinar com maior facilidade o conteúdo e o contexto das imagens produzidas. Essa compreensão não apenas aumentará a utilidade dessas tecnologias, mas também abrirá caminho para uma gama mais ampla de aplicações inovadoras em diversos setores, como a engenharia biomédica.

III. Metodologia e Teoria

A. Aquisição de dados

I. Imagens sintéticas

Para a geração das imagens, foi utilizado o *Stable Diffusion* [3] na versão 2.4.7, executado localmente em uma máquina Linux (Intel Core i7-12700F; GPU: RTX 4060; RAM: 32 GB), em combinação com a biblioteca *Compel* para atribuição de pesos às palavras. O *prompt* base utilizado foi: *"Pleasant/Unpleasant (peso) landscape"*. O substantivo *"landscape"* foi selecionado por se encaixar em uma das categorias encontradas nas bases de imagens com conteúdo afetivo avaliado por humanos.

Inicialmente, foi gerado um conjunto de 220 imagens utilizando esse *prompt*. Os pesos atribuídos variavam de -5 a -1 e de 1 a 5, com a expectativa de que valores negativos reduzissem e valores positivos aumentassem a força semântica dos adjetivos no *Stable Diffusion* (SD). A ideia por trás dessa abordagem era avaliar a variabilidade semântica através da análise das complexidades das imagens resultantes para verificar se as mudanças nos pesos influenciavam a interpretação conforme o esperado.

II. Imagens Oasis

Além disso, foram selecionadas 30 imagens do banco *Oasis*, classificadas por humanos com base no conteúdo afetivo. O objetivo foi utilizar os modelos de linguagem para classificar a valência hedônica na escala *SAM* e assim avaliar a assertividade dos modelos em relação aos padrões humanos. As imagens foram escolhidas em intervalos ao longo do banco completo, seguindo a classificação que varia de mais desagradáveis a mais agradáveis.

B. Tratamento de dados

I. Imagens sintéticas

Para o primeiro conjunto de imagens, foram calculadas as medidas de complexidade após a compressão utilizando o método Save"com o parâmetro Optimize"da biblioteca PIL [4]. A complexidade, expressa em porcentagem, foi determinada com base no tamanho em bytes da imagem comprimida, dividido por $512 \times 512 \times 3$. Esse cálculo permite observar a variação na complexidade das imagens em função do peso utilizado durante a geração.

Em seguida, cada "seed"(ou "semente") usada para gerar as imagens foi emparelhada com a imagem correspondente sem peso. A diferença de complexidade entre a imagem com peso e a imagem sem peso foi então calculada, proporcionando uma medida da variação na complexidade. Assim, foi possível analisar a variação da força semântica atribuída ao adjetivo em comparação com a imagem que não recebeu peso algum. Foram geradas retas de tendência linear para quantificar/analisar o crescimento.

II. Imagens Oasis

Utilizando o *Ollama* com o modelo *LLaVa* [5], foi possível avaliar a valência hedônica das imagens do banco *Oasis* empregando a escala *SAM* [6]. O processo consistiu em duas etapas principais: na primeira etapa, a imagem foi analisada pelo modelo, que gerou um texto descritivo juntamente com uma nota na escala *SAM*; na segunda etapa, essa nota foi capturada e registrada em um banco de dados para as análises de tendência de crescimento.

Passagem 1

- **Prompt de sistema:** `You are an adult participating in a study about affective content of pictures. Describe the picture in detail. Write the rating of the picture as a number on a scale from 1 to 9.`
- **Prompt de usuário:** `Rate the hedonic valence of the picture, on a scale from 1 to 9, where 1 means very negative and 9 means very positive.`

Passagem 2

- **Prompt de sistema:** `You are a JSON parser. You must parse the given text only, and nothing else. Identify the rating expressed in the text and write it in JSON format: {'rating': number}.`
- **Prompt de usuário:** `Parse the following text:`

Todos os dados processados foram gerados em Python 3.11.9, utilizando as bibliotecas `NumPy`, `Matplotlib` e `SciPy`.

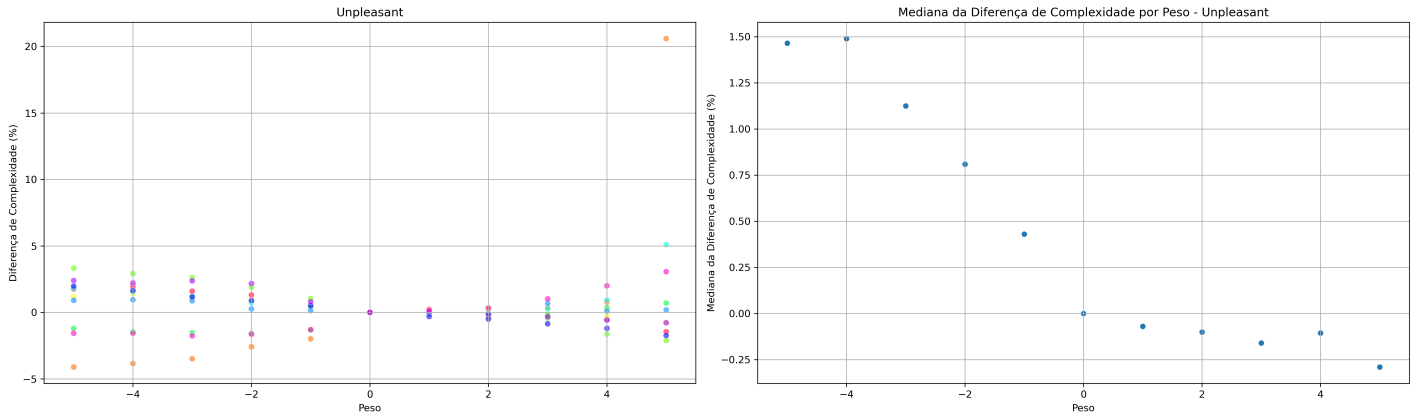

Figura 1: Gráfico de dispersão das *diferenças de complexidade* para cada "seed" de "Unpleasant" em relação aos pesos. À direita a mediana pontual de cada peso.

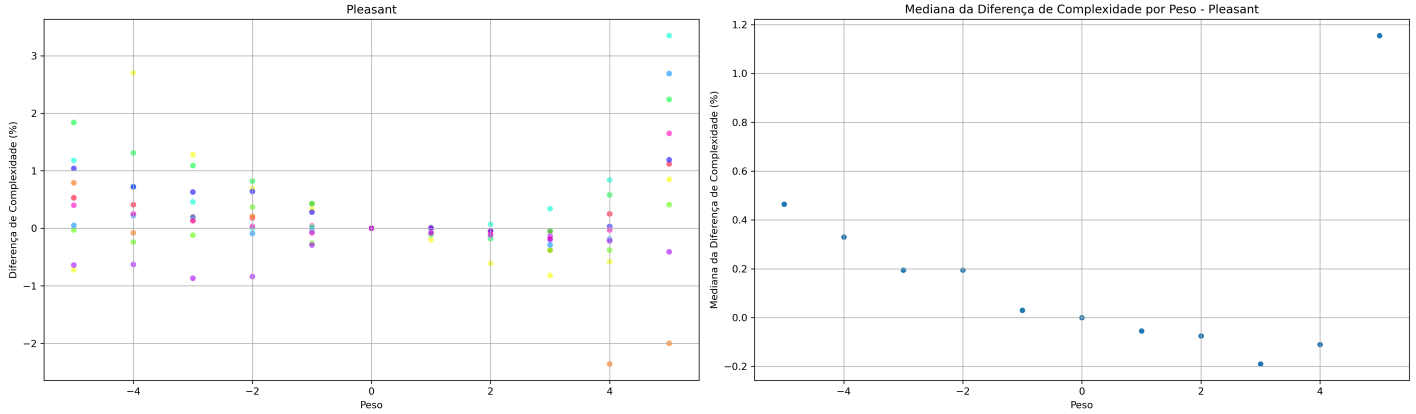

Figura 2: Gráfico de dispersão das *diferenças de complexidade* para cada "seed" de "Pleasant" em relação aos pesos. À direita a mediana pontual de cada peso.

IV. Resultados Obtidos

A. Imagens sintéticas

A partir do gráfico de peso versus complexidade, observou-se que a complexidade da imagem varia conforme o peso atribuído aos adjetivos se afasta de zero. Isso indica uma relação entre os pesos atribuídos aos prompts descritivos das imagens e a complexidade visual resultante (Fig. 1).

O uso da biblioteca *Compel* para a atribuição de pesos mostrou-se satisfatório, evidenciando uma progressão clara do prompt conforme definido. Em termos qualitativos, as imagens "pleasant" tendem a apresentar cores mais quentes e maior presença de vegetação, enquanto aquelas "unpleasant" mantêm a mesma topografia, mas exibem menos vegetação e cores mais frias. Com a variação dos pesos, é possível perceber a preservação do cenário, com essas alterações mais ou menos acentuadas de acordo com o peso aplicado (Figura 2).

Seguindo a ordem crescente dos pesos (-5 a 5), as medianas das diferenças de complexidade para as imagens *Pleasant* e *Unpleasant* foram, respectivamente: (Tabela 1)

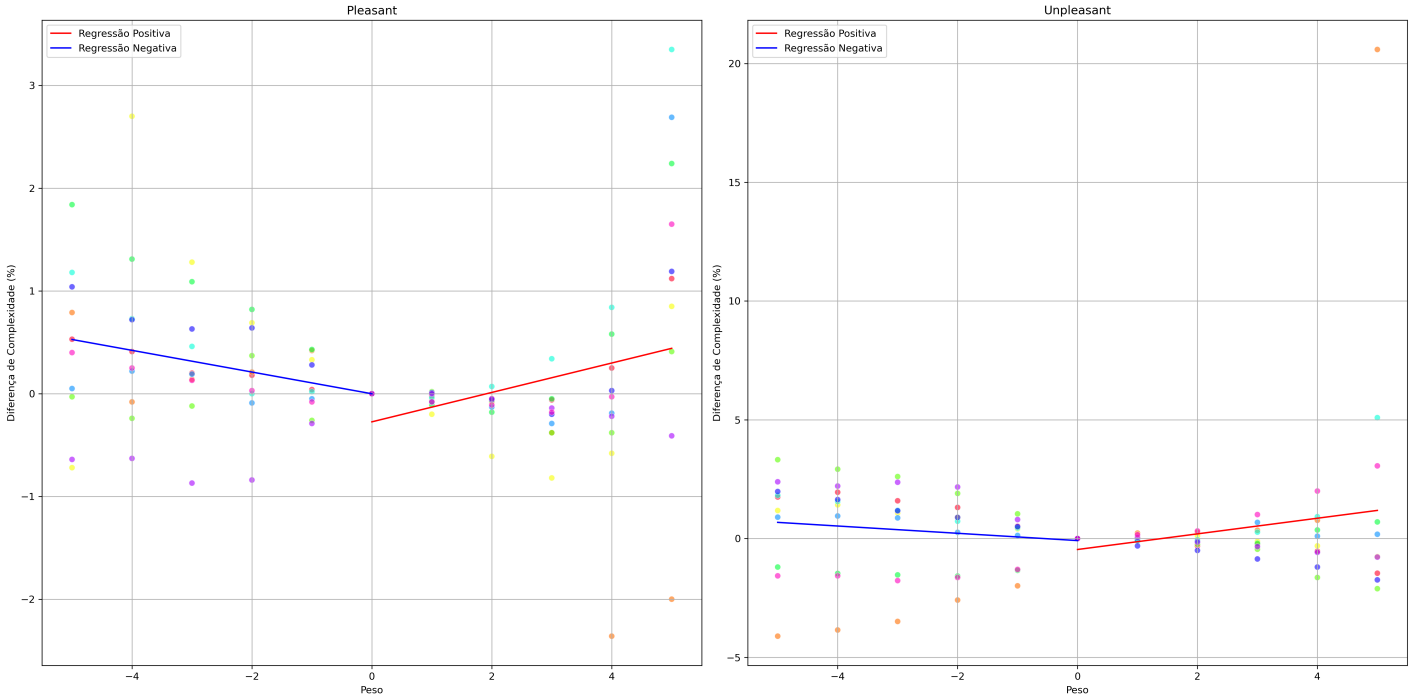

Figura 3: Gráficos de dispersão com retas de tendência para "Pleasant"e "Unpleasant"em função do peso.

Tabela 1: Medianas das diferenças de complexidade para as imagens *Pleasant* e *Unpleasant*.

Peso	**Pleasant (%)**	**Unpleasant (%)**
-5	0,465	1,465
-4	0,330	1,490
-3	0,195	1,125
-2	0,195	0,810
-1	0,030	0,430
1	-0,055	-0,070
2	-0,075	-0,100
3	-0,190	-0,160
4	-0,110	-0,105
5	1,155	-0,290

Para as imagens *Pleasant*, a inclinação da reta de regressão foi de 0,329 (p = 0,126) para os pesos positivos e -0,153 (p = 0,222) para os pesos negativos. Para as imagens *Unpleasant*, as inclinações foram 0,143 (p = 0,024) para os pesos positivos e -0,105 (p = 0,021) para os pesos negativos. (Figura 3)

B. Imagens Oasis

A análise das imagens do *Oasis* revelou que as classificações do modelo seguiram a lógica progressiva estabelecida pelo banco de dados. No entanto, foram identificadas duas medições cujos valores estavam fora da escala *SAM*. Observou-se que as notas de valência aumentaram progressivamente, apresentando uma inclinação de 1,96 e um intercepto de -3,39, com um coeficiente de correlação $r = 0,742$.

É importante destacar que, como as notas de valência humana variavam de 1 a 7, enquanto a classificação do modelo utilizava uma escala diferente, a inclinação da função identidade deveria ser $\frac{9}{7}$. Ao comparar essa inclinação com a obtida pela regressão linear, identificamos uma discrepância de aproximadamente 52% entre as duas inclinações.

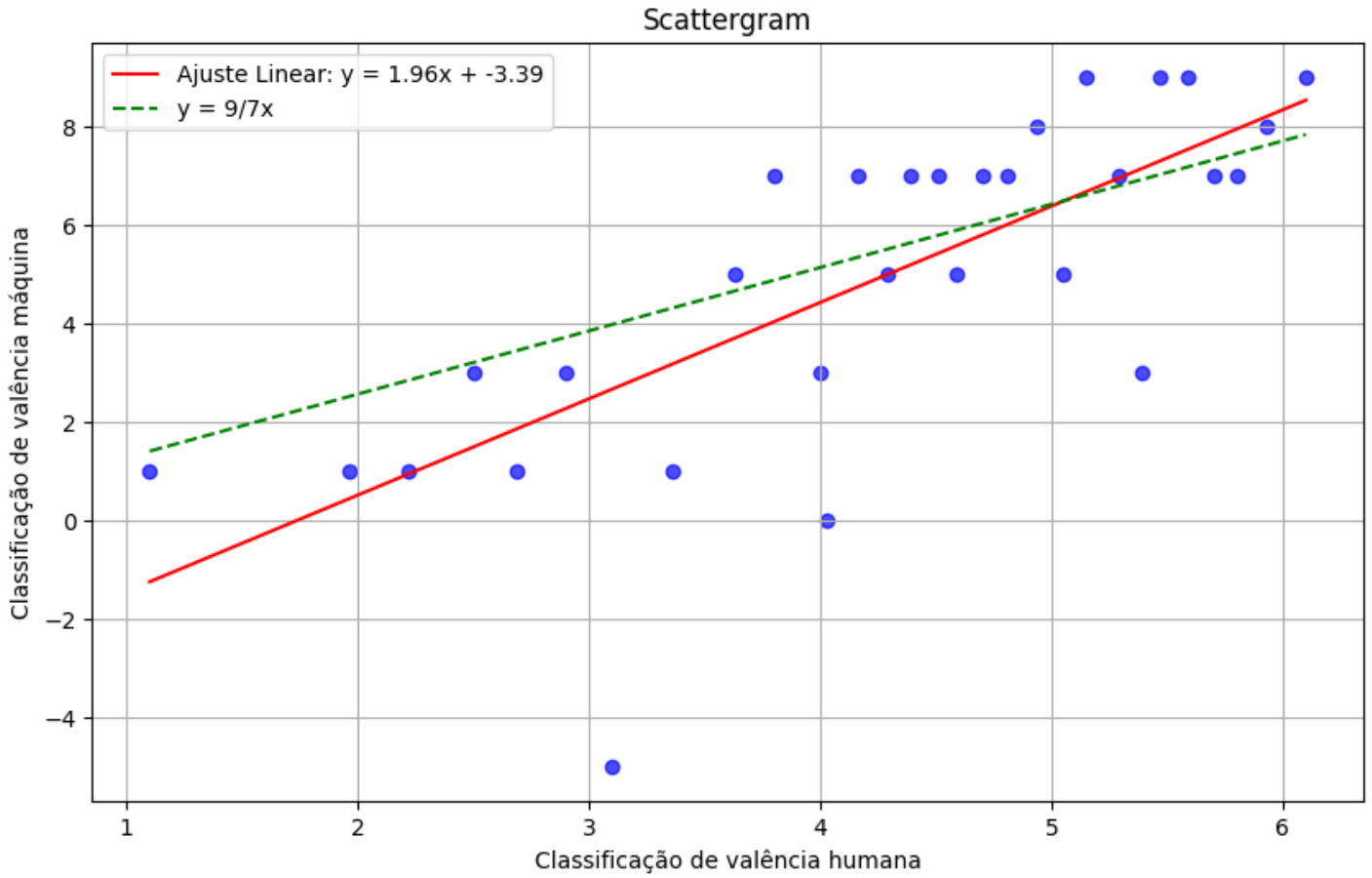

Figura 4: Gráficos de dispersão com ajuste linear para imagens do banco Oasis. Slope (a): 1,96 Intercept (b): -3,39 Coeficiente de Correlação (r): 0.742

(a) *Pleasant +3.*

(b) *Pleasant -3.*

(c) *Unpleasant +3.*

(d) *Unpleasant -3.*

Figura 5: Imagens com pesos obtidas após compressão

V. Conclusões

Os resultados observados no gráfico sugerem que tanto as valências hedônicas extremas, positivas e negativas, estão associadas a uma maior riqueza de detalhes visuais, o que pode estar relacionado à forma como os modelos de geração de imagem processam e representam essas emoções.

Além disso, o uso do modelo *LLaVa* para a avaliação de valência hedônica mostrou ser promissor, embora com algumas limitações. As inconsistências na classificação de algumas imagens indicam que pode não ser completamente confiável para tarefas que exigem um baixo índice de erros.

VI. Referências

1. Silva, J.V.; Beserra, A.N.; Serdeira, H.; Ribeiro, G.C.M. (2023). Há diferença de complexidade entre imagens sintetizadas com adjetivos emocionais antônimos em gerador texto-para-imagem? 12ª SIAc UFRJ, Caderno de Resumos: CT, página 324.

2. HUGGING FACE. Hugging Face Co. Prompt Techniques. Disponível em: `https://huggingface.co/docs/diffusers/using-diffusers/weighted_prompts`. Acesso em: 11 jul. 2024.

3. COMPVIS. Stable Diffusion Documentation. Disponível em: `https://github.com/CompVis/stable-diffusion`.

4. PyPI. PILLOW. Documentation. Disponível em: `https://pillow.readthedocs.io/en/stable/`. Acesso em: 11 jul. 2024.

5. LLaVa. Large Language and Vision Assistant. Disponível em: `https://ollama.com/library/llava` Acesso em: 11 jul. 2024.

6. Santos, R. et al. **Emotionality norms for the Brazilian version of the Deese-Roediger-McDermott (DRM) paradigm**. Brasília: Psicologia Teoria e Pesquisa, 2009. Disponível em: `https://www.researchgate.net/publication/262591801_Emotionality_norms_for_the_Brazilian_version_of_the_Deese-Roediger-McDermott_DRM_paradigm`. Acesso em: 11 jul. 2024.

7. Deckers, N.; Peters, J.; Potthast, M. Manipulating Embeddings of Stable Diffusion Prompts. Leipzig: Leipzig University, 2024.

8. Jandre, F.; Motta-Ribeiro, G.; Silva, J. Could large language models estimate valence of words? A small ablation study. Anais do XVI Congresso Brasileiro de Inteligência Computacional. Salvador, 2023.

9. Redies, C. et al. **Global image properties predict ratings of affective pictures**. Frontiers in Psychology, vol. 11, Sydney, 2020.

10. Rombach, R. et al. High-resolution image synthesis with latent diffusion models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.

11. Yu, H.; Winkler, S. Image complexity and spatial information. 5. ed. Austria: International Workshop on Quality of Multimedia Experience, 2013.

VII. Relatório de Atividades

- *Apresentação no Centro da Tecnologia, UFRJ:*
  Uma dinâmica de divulgação do projeto foi realizada em um stand no bloco H do Centro de Tecnologia da UFRJ. O evento contou com a apresentação de pôsteres e palestras ao público, com o objetivo de difundir o conhecimento adquirido e criar um espaço aberto para esclarecimento de dúvidas dos alunos.

- *Recuperação de computador:*
  A máquina do Laboratório de Instrumentação Biomédica, utilizada para a geração de imagens, não estava ligando. Com a ajuda do técnico Henrique Serdeira do NCE, foi realizada uma recuperação do computador. A máquina voltou a funcionar após a limpeza do CMOS e a redefinição das configurações iniciais da BIOS, seguida de uma série de testes nos componentes eletrônicos.

- *Instalação dos modelos localmente:*
  Dois computadores do Laboratório de Instrumentação Biomédica foram utilizados para a instalação do Stable-Diffusion 1.5 por meio da plataforma EasyDiffusion, além de outros modelos de linguagem como o llava e o phi3 através do Ollama para análise das imagens. Essa configuração permite a realização da pesquisa de forma muito mais eficiente, pois possibilita o uso dessas ferramentas com maior poder computacional.

- *Comparecimento ao seminário de alunos de pós-graduação:*
  Compareci aos seminários sobre programas de esporte e lazer e sobre a utilização de visão computacional para detectar problemas pulmonares. A participação nas discussões proporcionou um aprendizado valioso e me deu uma compreensão mais aprofundada sobre o funcionamento do ambiente de pós-graduação.

VIII. Avaliação do Bolsista

Minha experiência na pesquisa foi extremamente satisfatória. Embora tenha começado com pouco conhecimento sobre o assunto, fui me envolvendo cada vez mais com os modelos e compreendendo de forma prática o funcionamento das LLMs, o que torna o assunto cada vez mais fascinante. Este estudo tem sido emocionante, apesar de algumas dificuldades iniciais para me organizar e entender o ambiente científico. No entanto, o apoio e orientação que tenho recebido têm sido fundamentais para manter meu interesse e clareza sobre a pesquisa. Em resumo, esta jornada foi uma excelente oportunidade de desenvolvimento pessoal e acadêmico.