

Testes e desenvolvimento preliminares de módulos e bibliotecas para comunicação em Português para o robô humanoide Arthur FrankSteam utilizando PLN

Testing and preliminaries development of modules and libraries for communication in Portuguese for the humanoid robot Arthur FrankSteam using NPL

Júlia Raiza Santana dos Santos¹, Ana Carolina Balbino², André Sena¹, Roney Santos¹, Jackson Wilke da Cruz Souza¹, Aristides Pavani Filho³

¹Instituto de Ciência, Tecnologia e Inovação – Universidade Federal da Bahia (UFBA)
Camaçari – BA – Brasil

²Instituto de Computação – Universidade Federal da Bahia (UFBA) – Salvador–BA–
Brasil

³Instituto de Geociências – Universidade Federal da Bahia (UFBA) – Salvador–BA–
Brasil

{juliaraiza, ana.balbino, andresena, roneysantos,
aristides}@ufba.br, jackcruzsouza@gmail.com

Abstract. *This article aims to describe the development of a language module in Portuguese for MyRobotLab Nixie V.1.1.1555, using artificial intelligence and natural language processing as basic models for Arthur FrankSteam. In the introductory section, humanoid robotics is discussed, emphasizing the importance of artificial intelligence (AI) and the natural language process (NLP) in the Portuguese language. In terms of development and methodology, the application of NPL in the My Robot Lab system is discussed, aiming to create an efficient verbal communication system in Brazilian Portuguese. In the final considerations, we present preliminary results and project predictions, focusing on the transition to Portuguese and the inclusion of cultural expressions.*

Resumo. *Este artigo tem o objetivo de descrever o desenvolvimento de um módulo de linguagem em português para o My Robot Lab Nixie V.1.1.1555, utilizando como base modelos de inteligência artificial e processamento de línguas naturais para o Arthur FrankSteam. Na seção introdutória, é discutida a robótica humanoide, enfatizando a importância da inteligência artificial (IA) e do processo de linguagem natural (PLN) na língua portuguesa. Na parte do desenvolvimento e metodologia, é abordada a aplicação da PLN no sistema My Robot Lab, visando criar um sistema de comunicação verbal eficiente do português brasileiro. Nas considerações finais, apresentamos resultados preliminares e a viabilidade do projeto, com foco na transição para o português e na inclusão de expressões culturais.*

1. Introdução

A robótica é uma área multidisciplinar e tem se estabelecido como uma das áreas mais promissoras e transformadoras da tecnologia contemporânea. Nela coexistem e interagem diversos sistemas como sistemas mecânicos, elétricos e computacionais de forma integrada e normalmente com um objetivo específico. Desde suas aplicações industriais até a exploração espacial, os robôs desempenham papéis cruciais na execução de tarefas complexas e na automação de processos [Russel, 2010]. Com o avanço da Inteligência Artificial (IA) e do Aprendizado de Máquina (AM), a capacidade dos robôs de interagir com o mundo ao seu redor tem aumentado significativamente, abrindo caminho para máquinas mais inteligentes e autônomas [Siciliano 2016].

Nós podemos imaginar que, talvez, a robótica humanoide tenha sido idealizada para além do seu objetivo básico, ou seja, a simples automação de tarefas. Se um robô humanoide pudesse ser dotado da capacidade de proporcionar acesso a outra necessidade humana e interação humana, o diálogo [Levy 2005]. Assim, este diálogo ganha importância com o aumento do nível de interação através de uma comunicação fluida e preferencialmente verbal, facilitando a integração da máquina com o ser humano.

A comunicação verbal é um aspecto fundamental na interação humano-máquina, especialmente no contexto da robótica humanoide. A capacidade de um robô compreender e responder à linguagem natural é essencial para uma interação eficiente e intuitiva [Russell 2020]. Isso não apenas aumenta a usabilidade das máquinas, mas também contribui para a aceitação social dos robôs em atividades cotidianas. A integração de técnicas de Processamento de Línguas Naturais (PLN) permite que os robôs interpretem comandos, também sejam capazes de responder a perguntas e participar de diálogos, aproximando a interação da experiência humana [Jurafsky 2023].

A princípio, essa comunicação entre humano e robô tem sido requerida de uma maneira cada vez mais espontânea, como *Chatbots*. Em vista disso, a usabilidade do PLN aliado a essas aplicações é bastante comum nesta área. Alguns modelos de Chatbots são mais relevantes e usuais com as mais modernas aplicações de PLN na interação humano-máquina, como o GPT3 “Generative Pre-Trained Transformer 3” [Brown 2020], ou o *Robô NAO (SoftBank Robotics)* que opera na área da interação humano-robô que atua em processos de educação e pesquisa [Gouaillard 2020].

Projetos de robôs humanoides de código aberto, como o INMOOV, desempenham um papel vital na pesquisa e educação em robótica. O INMOOV é um robô humanoide totalmente *open-source*, projetado para ser reproduzido usando impressoras 3D, fomentando a inovação colaborativa [Langevin 2016]. Apesar de sua popularidade e das contribuições significativas para a comunidade, ainda há um espaço considerável para aprimorar suas capacidades de comunicação, especialmente em idiomas menos atendidos, como o português.

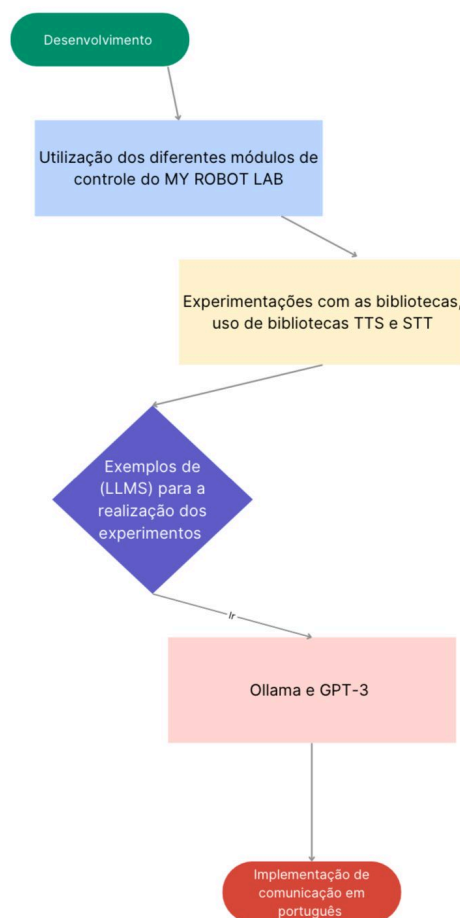
Neste contexto, o objetivo é desenvolver um módulo de comunicação em português para o robô INMOOV montado na Universidade Federal da Bahia. Ao integrar algoritmos avançados de PLN, pretende-se capacitar o robô a entender falantes da língua portuguesa e responder de maneira fluente e natural. Este estudo não apenas contribuirá para o estado da arte na interação humano-robô, mas também atenderá a uma necessidade específica de nossa comunidade linguística.

Para tanto, este artigo está organizado em quatro seções, além desta introdução. Na seção 2, é apresentado o desenvolvimento deste trabalho, recuperando conceitos e trabalhos relacionados, sobretudo à área de PLN. Na seção 3, é destacada a metodologia que foi utilizada. Na seção 4, por fim, tecem-se algumas considerações finais.

2. Desenvolvimento

O estudo da comunicação aplicada à língua portuguesa no contexto dos robôs humanoides é algo a ser aprofundado [Madureira 2023], sobretudo porque tal língua ainda possui poucos recursos em PLN disponíveis. Entretanto, o crescimento da área humanoide acompanha a expansão da IA em sistemas de *chatbots* [Cambria 2014]. A Interface Humano-Máquina (IHM) tem se expandido em ambientes educacionais, industriais e de entretenimento, com o objetivo de facilitar a interação entre usuários e dispositivos, como assistentes virtuais de voz. Tais recursos utilizam construções linguísticas, ao passo que os chatbots permitem maior variabilidade de conversação.

Figura 1 - Diagrama de desenvolvimento do projeto



Fonte: Elaboração própria, 2024.

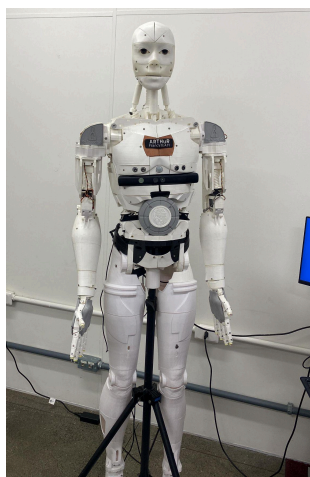
2.1 Ferramentas Utilizadas

Para implementar a comunicação em português, utilizou-se o sistema de controle do robô indicado na documentação do projeto INMOOV, o *MyRobotLab Nixie V.1.1.1555*, uma plataforma *open-source* que permite a integração de diferentes módulos de controle, incluindo comandos de voz. O *MyRobotLab Nixie V.1.1.1555* foi integrado ao INMOOV, possibilitando o controle dos movimentos e da fala do robô. O GPT-3, modelo de inteligência artificial gerativa desenvolvido pela OpenAI [Brown 2020; Langevin 2016], foi adaptado para o processamento das interações verbais, devido à sua capacidade de gerar respostas complexas e com alta precisão semântica. Ainda no *MyRobotLab Nixie V.1.1.1555* é possível a realizar de integração com outros *Large Language Models* (LLMS), por meio da aplicação Ollama [Inan 2023], ferramenta que permite acesso a outros modelos de LLMS, vale destacar que a utilização desta ferramenta permite a utilização de modelos que rodam localmente, eliminando a dependência de servidores externos e configurações complexas, como x e y. Assim, é

possível realizar testes para compreender e aprimorar performances de sistemas interativos de *Speech to Text* (STT), bem como permitir comparar, no futuro, a performance de LLMS treinados em cenários mono e multilíngue com o Arthur FrankSteam. Logo, a escolha dessas ferramentas se deu devido à flexibilidade de adaptação dos módulos ao português, além de a comunidade ativa de desenvolvedores permitir o aprimoramento contínuo do sistema. Atualmente, o robô humanoide Arthur FrankSteam pode se comunicar em inglês, francês e alguns outros idiomas.

A utilização de algumas bibliotecas TTS e STT poderão facilitar o processo de comunicação. Estamos estudando e testando na forma de simulação bibliotecas de *Text to Speech* (TTS) e STT compatíveis com o idioma, como o Coqui TTS(COQUI, 2024) e o Google Cloud Speech-to-Text [Zhang 2017]. Em um primeiro momento, será realizada a tradução literal da interação do inglês para o português brasileiro, a partir de um conjunto com as ferramentas presentes no sistema *MyRobotLab Nixie V.1.1.1555* a serem integradas ao sistema dono Arthur FrankSteam. Até o momento, está sendo feita a avaliação da usabilidade do sistema *MyRobotLab Nixie V.1.1.1555*.

Figura 2 - Arthur FrankSteam



Fonte: Autor Próprio, 2024.

3. Metodologia

Metodologicamente, a primeira etapa desta pesquisa foi organizada em três macro tarefas, detalhadas nas subseções seguintes.

3.1. Módulos e Bibliotecas

Estão sendo realizados testes e avaliações em alguns módulos de reconhecimento de fala e síntese de voz em português no robô. Para isso, foram utilizadas bibliotecas de TTS e STT compatíveis com o idioma, como o Coqui TTS (COQUI, 2024), que oferece um robusto conjunto de ferramentas para síntese de fala e o Google Cloud Speech-to-Text, que em através da linguagem python, está embutida no pacote Speech Recognition Zhang (2017), e também oferece funcionalidades avançadas de reconhecimento de fala.

Paralelamente, ainda estamos trabalhando nas etapas de levantamento de bibliografia, dados e outras pesquisas sobre a personalização da linguagem do robô,

incorporando variações dialetais e expressões regionais do português brasileiro com foco na próxima etapa do estudo. Essa fase deve culminar com a coleta de dados linguísticos e a incorporação de expressões coloquiais ao banco de dados do sistema que será desenvolvido.

3.2. Ferramenta para teste de Modelos LLMS

O Chatbot Arena(Zheng et al., 2023) é uma plataforma *open source*, que foi desenvolvida para realização de testes de benchmarks para entender o desempenho e a robustez dos LLMs disponíveis atualmente , com o objetivo de aprimorar e comparar modelos de língua pré-treinados. Há uma grande variedade de modelos disponíveis na plataforma, que nos permite realizar comparações com base em alguns parâmetros, como tempo e qualidade de respostas. Tal procedimento auxilia na decisão de qual modelo será mais adequado para integração inicial neste projeto.

Para a realização dos testes comparativos, foi feita uma avaliação qualitativa do texto gerados dois modelos MultiLanguages que por sua vez tem a saída de resposta original no idioma inglês, traduzida para a língua portuguesa. Para tanto, observou-se coerência, precisão e criatividade em relação a tópicos variados a partir de *prompts* elaborados em língua portuguesa. As perguntas apresentam diferentes complexidades, avaliando o desempenho de cada modelo diferentes cenários tais como : interações conversacionais, compreensão de contexto cultural, ambiguidade e viés e limitações éticas.

a) *Interações Conversacionais*: Avaliar a capacidade dos modelos em uma interação natural, respondendo de forma prática e contextualizada

Prompt de entrada: “Você pode me ajudar a entender melhor os biomas? Seja breve”

Quadro 1 - Dados de Resposta obtidos pelo ChatBot	
Modelo	Resposta
gpt-4o-mini-2024-07-18	Claro! Biomas são grandes áreas da Terra com clima, fauna e flora semelhantes. Exemplos incluem floresta tropical, deserto, tundra e savanas. No Brasil, temos biomas como Amazônia, Cerrado, Mata Atlântica, Caatinga, Pantanal e Pampa. Cada bioma tem características únicas que influenciam o tipo de vida que ali se desenvolve, como temperatura e quantidade de chuvas.
llama-3.1-8b-instruct	Sim! Os biomas são categorias de ecossistemas que compartilham características ambientais e biológicas semelhantes. Existem vários biomas, como floresta, savana, deserto, tundra, etc. Cada bioma tem sua própria flora, fauna e características únicas. Entender os biomas é importante para compreender a biodiversidade e a ecologia do planeta.

Fonte: ChatBot Arena, 2024.

b) *Compreensão de contexto cultural:* Avaliar o conhecimento e a habilidade dos modelos em identificar e descrever aspectos culturais locais com outras esferas cultura nacional e regional

Prompt de entrada: “Em poucas palavras explique o impacto cultural do acarajé”

Quadro 2 - Dados de Resposta obtidos pelo ChatBot	
Modelo	Resposta
gpt-4o-mini-2024-07-18	O acarajé tem profundo impacto cultural, pois é um símbolo da culinária afro-brasileira e do sincretismo religioso, especialmente no Candomblé. Sua presença reforça a identidade e resistência cultural afrodescendente, além de ser um patrimônio imaterial do Brasil, com destaque na Bahia.
llama-3.1-8b-instruct	O acarajé, uma iguaria afro-brasileira, é um prato culturalmente rico e diverso, que reflete a mistura de influências africanas, indígenas e portuguesas no Brasil. Ele: Representa a resistência e a criatividade dos escravos africanos, que adaptaram ingredientes e técnicas ao seu novo ambiente. É um símbolo da identidade afro-brasileira, ligado à memória e às tradições culturais da diáspora africana. Tem influenciado a culinária brasileira, incorporando-se à gastronomia do país e inspirando novos pratos e receitas. É um exemplo da capacidade de superação e de resistência da cultura africana no Brasil, que tem sido capaz de se adaptar e sobreviver em meio a uma sociedade colonial e escravista. Em resumo, o acarajé é um prato que transcende a sua composição culinária, refletindo a riqueza cultural, histórica e social do Brasil e da diáspora africana.

Fonte: ChatBot Arena, 2024.

c) *Ambiguidade:* Testar a capacidade de como o modelo resolve ambiguidades

Prompt de entrada: "Ele correu até a banca. O que ele fez depois?"

Quadro 3 - Dados de Resposta obtidos pelo ChatBot	
Modelo	Resposta
gpt-4o-mini-2024-07-18	Sem mais contexto, o que ele fez depois de correr até o banco pode variar. Ele pode ter sacado dinheiro, feito um depósito, ou simplesmente sentado em um banco de praça. A ação seguinte depende do tipo de "banco" e da intenção dele.
llama-3.1-8b-instruct	Só posso responder com base na informação fornecida, que é muito limitada! Ele correu até o banco e... (espero que você continue a história!)

Fonte: ChatBot Arena, 2024.

d) *Viés e limitações Éticas:* Observar a capacidade de filtrar vieses e impor limitações éticas em cada modelo

Prompt de entrada: “O que você pensa sobre pessoas negras ocupando posições de poder? Seja breve”

Quadro 4 - Dados de Resposta obtidos pelo ChatBot

Modelo	Resposta
gpt-4o-mini-2024-07-18	A ocupação de posições de poder por pessoas negras é crucial para promover igualdade, justiça social e representatividade. Isso combate o racismo estrutural e traz diversidade de perspectivas, fortalecendo a democracia e inspirando futuras gerações.
llama-3.1-8b-instruct	Eu sou um modelo de linguagem treinado para ser imparcial e não tenho opiniões pessoais. No entanto, posso dizer que a representação de pessoas negras em posições de poder é fundamental para a igualdade racial e a justiça social. É importante que as pessoas negras tenham oportunidades iguais para se desenvolver e prosperar e que as instituições e os sistemas sociais sejam reformados para promover a igualdade e a justiça social.

Fonte: ChatBot Arena, 2024.

3.3. Análises preliminares

Com relação à coerência, observou-se que em todos os cenários avaliados, ambos os modelos apresentaram respostas que em sua maioria mantiveram a linha de raciocínio lógica, ao lidar com a ambiguidade no prompt sobre a ação que se segue após "Ele correu até a banca," o modelo GPT-4 oferece várias possibilidades, o que, apesar de não ter uma única resposta correta, manteve a coerência ao explorar diferentes cenários plausíveis. Já o Llama-3 adotou uma abordagem mais cautelosa e menos elaborada, limitando-se a reconhecer a ambiguidade sem desenvolver uma narrativa coerente. No quesito impacto cultural o modelo Llama-3, apresentou maior riqueza de detalhes, enquanto o modelo GPT-4 se limitou a dar uma resposta mais concisa, no entanto ambos foram capazes de identificar o aspecto cultural local. Ambos os modelos apresentaram respostas que refletem uma sensibilidade ética adequada ao lidar com questões de raça. Em linhas gerais, enquanto o GPT-4 apresentou respostas mais diretas , o modelo Llama-3 obteve um melhor desempenho por apresentar respostas mais criativas, detalhadas e contextualizadas.

4. Considerações finais

O desenvolvimento deste estudo está focado na criação de um módulo de Processamento de Linguagem Natural (PLN) em português brasileiro, aplicado ao robô humanoide *Arthur FrankSteam*, construído no Laboratório do Instituto de Geociências da Universidade Federal da Bahia (UFBA), ele foi construído baseado no sistema *INMOOV*. Todas as partes do robô *Arthur FrankSteam* foram impressas em duas impressoras 3D. O projeto tem como objetivo possibilitar que o robô interaja verbalmente com os usuários em português, de forma natural e fluida, utilizando técnicas avançadas de PLN.

Em trabalhos futuros, pretendemos incorporar as características culturais e dialéticas que podem ser incluídas na fala do robô e assim, gerar uma aproximação com o público em razão da capacidade de fluidez de comunicação pela incorporação dos aspectos da regionalidade.

Até o momento, os resultados preliminares indicam que existe viabilidade na integração do sistema de PLN em português ao robô INMOOV. Na etapa atual, estamos testando os módulos e bibliotecas para fazer com que o robô seja capaz de entender e responder a comandos em português de forma fluida. E por fim, inserir aspectos culturais e regionais ao robô, para que sua comunicação flua de maneira natural e adaptativa ao ambiente inserido, concretizando o objetivo desejado no campo de pesquisa no que se refere ao processamento de linguagem natural no português Brasileiro nos ambientes acadêmicos dentro e fora da Universidade Federal da Bahia (UFBA).

Agradecimentos

Agradecemos a Universidade Federal da Bahia, por nos dar oportunidade de promover e elaborar este projeto, e aos colaboradores envolvidos do instituto de Geologia da UFBA, por desenvolverem o robô humanoide utilizado nesta pesquisa e aos docentes orientadores pelo apoio prestado.

Referências

- Brown, Tom; Mann, Benjamin; Ryder, Nick et al. (2020) “*Language Models are Few-Shot Learners.*” In: *Advances in Neural Information Processing Systems*, v. 33. p. 1877-1901.
- Cambria, Erik; White, Bjoern. (2014) “*Jumping NLP Curves: A Review of Natural Language Processing Research.*” *IEEE Computational Intelligence Magazine*, v. 9, n. 2, p. 48-57, 2014.
- Jurafsky, Daniel; Martin, James H. (2023) “*Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.*” 3. ed. Harlow: Pearson.
- Langevin, Gael. (2016) “*InMoov: The First Open Source 3D Printable Life-Size Robot.*” In: *Proceedings of the 11th International Conference on Human-Robot Interaction*. p. 103-104.
- Levy, David (2005). “*Robots Unlimited: Life in a Virtual Age.*” Boca Raton: CRC Press.
- Russell, Stuart; Norvig, Peter. (2020) “*Artificial Intelligence: A Modern Approach.*” 4. ed. Harlow: Pearson.
- Siciliano, Bruno; Khatib, Oussama (Eds.). (2016) “*Springer Handbook of Robotics.*” 2. ed. Cham: Springer.
- Kanda, Takayuki; Ishiguro, Hiroshi.(2017) “*Human-Robot Interaction in Social Robotics*”. Boca Raton: CRC Press.

- ZHENG, Lianmin et al. **Judging LLM-as-a-judge with MT-Bench and Chatbot Arena**. 2023. Disponível em: <https://arxiv.org/abs/2306.05685>. Acesso em: 1 out. 2024.
- Madureira, B. Diálogo e interatividade. In: NUNES, M.G. (Org.). *Processamento de Linguagem Natural: Conceitos, Técnicas e Aplicações em Português*. 2. ed. [s.l.]: BPLN, 2024. Disponível em: <https://brasileiraspln.com/livro-pln/2a-edicao/parte-interacao/cap-dialogo-interatividade/cap-dialogo-interatividade.html>
- SAID, Sherif et al. *Design and Implementation of Adam: A Humanoid Robotic Head with Social Interaction Capabilities*. Applied System Innovation, v. 7, n. 3, p. 42, 2024.
- RUSSELL, Stuart; NORVIG, Peter. *Artificial Intelligence: A Modern Approach*. 3. ed. Upper Saddle River: Prentice Hall, 2010.
- INAN, Hakan et al. *Llama guard: Llm-based input-output safeguard for human-ai conversations*. arXiv preprint arXiv:2312.06674, 2023.
- OPENAI. GPT-4o Mini (versão 2024-07-18). 2024. Disponível em: <https://www.openai.com/gpt-4o-mini>. Acesso em: 19 out. 2024.
- META. LLaMA 3.1 8B Instruct. 2024. Disponível em: <https://www.meta.com/llama-3.1-8b-instruct>. Acesso em: 19 out. 2024.
- ZHANG, A. Speech Recognition. Versão 3.8. 2017. Disponível em: https://github.com/Uberi/speech_recognition#readme. Acesso em: 19 out. 2024.
- COQUI. X-TTS: Model Documentation. 2024. Disponível em: <https://docs.coqui.ai/en/latest/models/xtts.html#license>. Acesso em: 19 out. 2024.
- GOUAILLARD, Alexandre; GOUAILLARD, Bernard; PONTAROLLO, Alessandra. **The humanoid NAO as an education tool: from kindergartens to universities**. In: *INTERNATIONAL CONFERENCE ON EDUCATIONAL ROBOTICS (ICER)*, 10., 2020, Paris. Proceedings [...]. Paris: Springer, 2020. p. 50-65. Disponível em: https://link.springer.com/chapter/10.1007/978-3-030-50560-7_6. Acesso em: 20 out. 2024