

Estudo de Caso de um Robô Social Integrado com LLM Interagindo em Ambientes Públicos

Marcos V. Cruz¹, Murillo R. Batista¹, Douglas A. de Sousa¹, Clarissa F. C. L. Loureiro¹,
Adriano L. Lima¹, Josue J. G. Ramos¹

¹ Centro de Tecnologia da Informação Renato Archer (CTI) – Campinas – SP – Brasil

{mvcruz, mbatista, dasousa, clima, alima, jgramos}@cti.gov.br

Abstract. *Social robots are an important aspect of Society 5.0 as they allow better collaboration between people and robots. The rise of Large Language Models (LLMs) may contribute significantly in conducting such interactions. This paper presents a case study of a social robot integrated to an LLM that is able to answer people with speech and body gestures. A preliminary pilot study tested both online and offline LLM settings, and case studies with integration on a Pepper robot were made on a research lab visit and a broader scope event with a large flow of people. The observed interaction results are presented and discussed, providing suggestions and guidelines for further uses of LLMs on robots for human interaction.*

Resumo. *Robôs com habilidades sociais são importantes para a Sociedade 5.0. O advento dos modelos de linguagem de grande escala (LLMs) pode representar um avanço significativo neste processo. Este trabalho contribui nesse contexto ao apresentar um estudo de caso envolvendo o uso de um robô social integrado a uma LLM capaz de responder a perguntas feitas por pessoas utilizando voz e gestos. Um estudo preliminar de desempenho de LLMs locais e remotas é seguido de estudos de caso realizados em dois ambientes distintos: um laboratório de pesquisa com visitantes e um evento com grande fluxo de pessoas. São apresentadas sugestões para futuras aplicações de LLMs na interação de robôs com pessoas.*

1. Introdução

A Sociedade 5.0 tem como pilar aspectos da Sociedade 4.0 como inteligência artificial, internet das coisas, robótica e computação em nuvem, acrescidas de diretrizes similares ao ESG (ambiental, social e governança)[SEBRAE 2025]. Um componente importante da Sociedade 5.0 é a robótica colaborativa, ou seja, o desenvolvimento de robôs com condições de interagir de forma proativa com as pessoas.

Busca-se na Sociedade 5.0 a interação entre homens e máquinas de forma mais frequente e natural, envolvendo autonomia, aspectos sociais e culturais. No caso específico de robôs físicos, diferentemente de aplicativos de smartphones ou computadores, há um elemento de incorporação do agente de software a uma estrutura mecânica capaz de atuar no ambiente. Isto dá à interação um grau de profundidade único na forma de antropomorfismo, ou seja, na percepção de que o robô tem personalidade e raciocínio próprios [Fink 2012].

A convivência entre humanos e robôs, assim como métodos de percepção e atuação relacionados, é estudada em uma área específica da Robótica, conhecida como Interação Humano-Robô (HRI). Um aspecto importante para uma interação que traga plenamente o antropomorfismo é como desenvolver a comunicação do robô, tanto verbal quanto não-verbal (como movimentação da cabeça e gestos com braços robóticos). uma comunicação pouco fluida ou excessivamente protocolar tende a diminuir a “ilusão” da personalidade. Sinais não-verbais têm a capacidade de fortalecer a expressividade e, consequentemente, a projeção de traços humanos do robô [Koike and Mutlu 2023].

O surgimento dos LLMs permitiu a existência de mecanismos de conversação mais complexos, capazes de dialogar de forma mais natural; é possível, também, passar instruções através de prompts para que um LLM já treinado e em funcionamento converse considerando certas informações prévias, como o nome e local de trabalho de um personagem a ser assumido por ele. Assim, é possível descrever uma personalidade e características prévias de conversação [Serapio-García et al. 2025], também, a um robô que se utilize de um LLM como ferramenta de comunicação não apenas verbal como, também, não-verbal.

O objetivo deste artigo é descrever o estudo de caso do robô Sophia, que utiliza a estrutura robótica do robô Pepper [SOFTBANK ROBOTICS 2025] em conjunto a uma máquina que executa um LLM para interagir com pessoas em ambientes públicos através de respostas e gestos escolhidos pelo LLM, que devem ser executados de forma coordenada e coerente, com o objetivo de permitir a aferição de antropomorfismo por parte do interlocutor.

Este artigo está organizado da seguinte maneira. Na Seção 2, as características de LLMs em HRI são apresentados, assim como trabalhos relacionados; na Seção 3, é descrita a metodologia, um teste piloto com os LLMs antes da integração ao robô, e explicado como é feita a integração do robô ao LLM; na Seção 4, os resultados são discutidos; o trabalho é concluído na Seção 5.

2. Aplicações de LLM em HRI

A HRI está evoluindo rapidamente com a integração de LLMs, que permitem que os robôs entendam e respondam a comandos de linguagem natural em tempo real. Este desenvolvimento está transformando a forma como os seres humanos e os robôs colaboram em domínios como indústria, segurança e interação social; no entanto, ele também apresenta novos desafios em termos de design, confiabilidade e experiência do usuário.

[Liu et al. 2024] apresentam como melhorar o desempenho da manipulação de robôs utilizando LLMs e modelos de linguagem, possibilitando o planejamento e o raciocínio de trajetórias complexas, e decompondo tarefas complexas em instruções simples. [Gkournelos et al. 2024] propõem um robô de manufatura baseado em LLM para melhorar a HRI na indústria de forma inteligente, permitindo interações intuitivas e programação rápida do sistema, resultando em melhorias significativas na colaboração e na eficiência.

As LLMs facilitam a colaboração entre humanos e robôs de forma mais intuitiva e eficiente, permitindo que os robôs interpretem instruções de linguagem complexas. [Asuzu et al. 2025] demonstrou que os LLMs podem efetivamente gerar o planejamento

de execução de tarefas de robôs de alta complexidade usando linguagem natural e modelos de linguagem de visão, resultando em um aumento de desempenho de 8,6% nas avaliações simuladas e um aumento de 14% na avaliação física. Seu estudo mostrou que o LLM pode melhorar o desempenho em tarefas de manipulação, montagem e navegação, especialmente quando combinado com feedback ou demonstrações humanas.

Outros estudos fornecem interfaces ou operadores amigáveis baseados em LLM, permitindo uma melhor comunicação e uma programação rápida do sistema em manufatura, como [Kim et al. 2024], que integraram efetivamente o LLM à colaboração entre humanos e robôs, reduzindo as interrupções repetitivas de busca de ferramentas e melhorando a eficiência na manufatura inteligente centrada no ser humano.

Outra possível aplicação de LLMs em HRI é em aprendizagem incremental; os robôs equipados com LLMs podem aprender de forma incremental com as interações humanas, melhorando seu comportamento ao longo do tempo e reduzindo o número de erros repetidos. Entretanto, [Schreiter et al. 2025] apontaram que o uso da estrutura de LLMs reduz a capacidade de reconhecer dinamicamente esses estados de se adaptar e se recuperar para retomar a interação. Além disso, o LLM aprimora as interações entre humanos e robôs, mas as interações com scripts têm desempenho semelhante em termos de eficiência e foco, podendo ter uma vantagem em termos de latência de resposta e consumo de energia.

Com base nestes estudos, resume-se que o uso de LLMs em HRI apresenta diversos benefícios, como maior capacidade de adaptação a situações distintas, maior riqueza de comportamentos do robô, facilidade de supervisão e fácil integração ao sistema robótico. Entretanto, o LLM é passível de falhas, especialmente quando utilizados modelos de menor capacidade, e também pode apresentar problemas de latência de resposta. Um desafio para a aplicação que deve ser considerado é o de preparação do modelo para reconhecimento do contexto, seja através de ajuste-fino (fine-tuning) ou de prompts prévios.

Não é apenas na indústria que há exemplos de LLMs aplicados em HRI. [Zhang et al. 2023] mencionam que há uma série de aplicações de LLMs em HRI na vida cotidiana em muitos campos diferentes, como ambientes educacionais, assistência médica e atendimento ao cliente. Em especial, os LLMs têm a capacidade de processar a linguagem humana ambígua e convertê-la em respostas acionáveis. Além disso, eles podem lidar com entradas multimodais (som, imagem), o que os torna mais atraentes para a HRI.

Propomos um robô social equipado com modelos LLM pré-treinados que servem como base de conhecimento e interação, permitindo que ele responda a perguntas que exijam habilidades em vários domínios. De modo geral, este projeto destaca que o LLM aprimora significativamente as habilidades de conversação dos robôs sociais, tornando suas interações mais naturais e eficazes, e pode ser aplicado em muitos cenários diferentes.

A principal contribuição deste estudo é a criação de um robô social nacional capaz de interagir em português brasileiro, emancipando e dando importância à pesquisa nacional na área de HRI. Além disso, inclui estudos de caso em situações reais.

3. Metodologia

Definimos um estudo qualitativo baseado em um estudo de caso exploratório [Yin 2017] para identificar como um robô com capacidades sociais integradas com um LLM se comporta em um ambiente público não controlado. Escolhemos esse estudo empírico pois ele se enquadra em um cenário no qual estamos realizando uma investigação particular, que não é generalizável e se aplica a uma situação específica, sem obstruções no ambiente, ou seja, o ambiente não é controlado e não adicionamos elementos artificiais a ele. O estudo busca observar alguns temas para responder às seguintes perguntas de pesquisa (RQs):

1. **RQ:** Quais são as principais características que influenciam para que uma interação entre humano e robô integrado com um LLM seja eficiente e precisa?
 - (a) **RQ:** O LLM fornece ferramentas adequadas para a implementação de um robô social?
2. **RQ:** Quais habilidades um robô com LLM precisa ter para interagir com pessoas em um ambiente público?
 - (a) **RQ:** Que tipo de dificuldade um robô social enfrenta em ambientes não controlados?
 - (b) **RQ:** Qual a aceitabilidade de um robô social operado por LLM?
3. **RQ:** Como os elementos do robô devem ser combinados para que haja coerência na interação?

Para responder a estas perguntas, usamos o método de pesquisa-ação [Lewin 1946], pois, ao mesmo tempo em que realizamos um estudo investigativo, faremos uma reflexão sobre a abordagem implementada nesse cenário; assim, aplicamos teoria e prática simultaneamente. Nossas observações são baseadas nos principais aspectos de um robô social (Funções com Orientação Social; Modalidades de Interação; Capacidades Cognitivas e Adaptativas; Presença Social; Capacidades de Comunicação), segundo [Goodrich et al. 2008]. A Figura 1 mostra a relação entre as RQs e os estudos de caso, ao ilustrar como os dados foram coletados, analisados e interpretados.

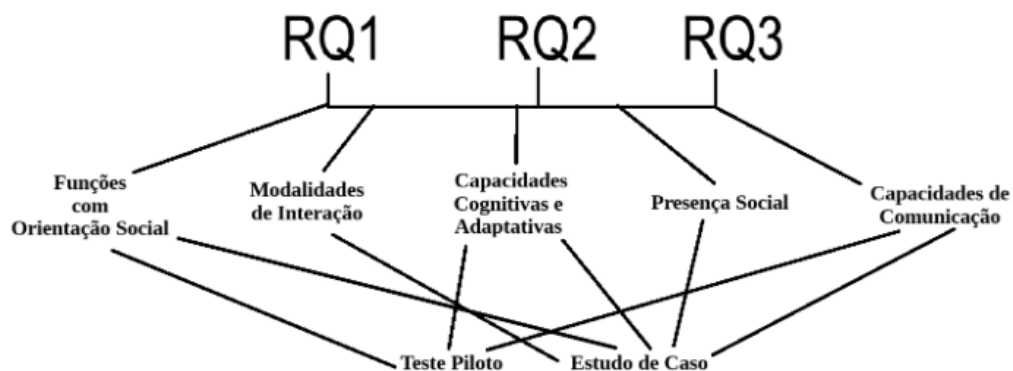


Figura 1. Cadeia de Evidências.

3.1. Teste Piloto

Foi realizado um teste piloto para avaliar a viabilidade do uso de LLMs no contexto de oferecer respostas a perguntas feitas verbalmente aplicando diferentes LLMs. Nesta etapa, não houve incorporação à plataforma Pepper, avaliando-se apenas as respostas verbais obtidas a partir dos LLMs, sem a integração de gestos.

Neste sentido, este teste piloto para o desenvolvimento e avaliação de diferentes arquiteturas de integração envolveu LLMs, sistemas de reconhecimento de fala (speech-to-text, STT), e de síntese de voz (text-to-speech, TTS) com foco na aplicação dentro do nosso insituto. Foram testadas soluções offline e online (ou seja, executando em um servidor externo) com diferentes modelos de LLM.

Na primeira parte, o LLM foi executado inteiramente a partir da própria máquina, sem o uso de um servidor externo. O modelo de LLM foi o LLaMa 3.1¹ com 8 bilhões de parâmetros, utilizando o pacote Whisper² para STT e o RHVoice³ para TTS.

O resultado foi o de uma voz natural e estável, com uma latência média de aproximadamente cinco segundos. Foram notadas inconsistências ocasionais nas respostas oferecidas pelo LLM, o que já era esperado pelo número de parâmetros relativamente baixo. Consideramos que trata-se de uma alternativa viável para cenários em que não há acesso a servidores externos que permitam o uso de LLMs com um maior número de parâmetros.

Em uma segunda etapa, foi escolhida uma LLM cuja execução é através do Groq, um sistema que utiliza hardware especializado para oferecer remotamente LLMs de alto desempenho⁴; isto permitiu o acesso a uma LLM LLaMA de 70 bilhões de parâmetros. Os mecanismos de TTS e de STT foram os mesmos.

As respostas obtidas foram mais consistentes, e o tempo de latência não se alterou, mesmo com a comunicação online. Esta solução, entretanto, requer conexão e o uso de tokens da API do Groq.

Posteriormente, no objetivo de avaliar uma estratégia online distinta do do Groq, foram realizados testes a partir de um servidor dedicado presente em nossa instituição. Foram utilizados a mesma LLM, TTS e STT do primeiro estudo. Foi observada uma latência média maior, em torno de 8 segundos e um uso moderado da GPU disponível.

É possível que alguma mudança de comunicação com o servidor em questão reduza a latência obtida, dado que não houve este problema tanto na solução offline quanto na solução via Groq; é improvável que o problema esteja relacionado à velocidade de execução da LLM em si, visto que o mesmo LLM foi utilizada no estudo offline.

3.2. Integração do LLM ao Robô

Nesta Seção, nós explicamos como foram programados os recursos do robô Sophia para rastrear e seguir rostos humanos, a interação e os processos de tomada de decisão.

¹<https://ai.meta.com/blog/meta-llama-3-1>

²<https://openai.com/index/whisper/>

³<https://rhvoice.org/>

⁴<https://groq.com>

Começamos com o rastreamento de rostos usando o módulo ALTracker da biblioteca NaoQi⁵, que permite que o robô detecte e siga rostos movendo a cabeça, gerando sinergia nas interações. Durante este processo, discutimos a importância fundamental de desativar o Autonomous Life e ativar a rigidez da cabeça para controle manual a fim de garantir movimentos precisos de rastreamento.

Em seguida, desenvolvemos uma solução personalizada de rastreamento de faces usando o OpenCV, em que um script Python personalizado utilizando Histograma de Gradientes (HOG) do OpenCV combinado com máquina de vetores de suporte (SVM) [Said et al. 2011] para a detecção de faces. Aprimoramos significativamente este script diminuindo a resolução da câmera e a taxa de quadros para aumentar a velocidade, implementando o processamento de quadros com base no tempo para aumentar a eficiência e tornando os movimentos da cabeça suaves usando um filtro de média móvel exponencial (EMA).

Para os recursos de animação, usamos inicialmente o modo de animação de poses no Choreographe [Aldebaran Robotics 2025]; esta ferramenta permite o movimento manual dos membros do Pepper e o registro de posturas ao longo de uma linha do tempo, embora até mesmo animações simples exijam vários movimentos definidos. Descobrimos que o Pepper suporta gestos e poses corporais por meio do sistema operacional NAOqi, que oferece controle sobre 20 graus de liberdade na parte superior do corpo para a execução de gestos expressivos. O processo envolve mapear as animações dentro do robô, que estão como arquivos .xar, e acioná-las por meio do Python usando o NAOqi pela função ALMotion.

Na parte do LLM, a função de inicialização de conversação utiliza a função *ConversationBufferWindowMemory* para manter um breve histórico de conversação, permitindo, assim, respostas sensíveis ao contexto. A configuração do modelo inicializa o Groq com o modelo do DeepSeek, definindo a temperatura (parâmetro do LLM associado à permissibilidade de variabilidade de resposta) como 1.0 para criatividade de resposta, para que o modelo não gere respostas aleatórias mas tenha alguma flexibilidade para elaborá-las. Empregamos um prompt detalhado do sistema, que descreve a personalidade, as regras, os interesses e as restrições de interação da Sophia em português, com um espaço reservado para mensagens do usuário para entradas dinâmicas de contexto. A composição da cadeia combina o modelo de prompt e o modelo de bate-papo do Groq em um único ciclo que pode ser chamado utilizando um pipeline, que é envolvido em um diálogo com memória para gerenciamento de estado.

As técnicas de engenharia de prompt aplicadas foram projetadas para orientar o comportamento e o formato de saída da IA. A mensagem do sistema funciona como uma *persona* e um manual de regras, definindo a identidade de Sophia, incluindo seu nome, origem e interesses, ao mesmo tempo em que estabelece um tom amigável e respeitoso, limita o idioma ao português brasileiro e determina regras rígidas de comportamento, como evitar comentários negativos sobre a CTI e não discutir tópicos delicados como política ou religião. Esta abordagem molda a personalidade da IA e estabelece limites éticos.

O LLM deve responder em um formato específico [Resposta][Ação], em que

⁵<https://lnk.ink/yYXRT>

[Ação] é uma palavra-chave de ação de uma lista predefinida. Estas ações são movimentos gestuais que o robô deve realizar simultaneamente com a fala. O próprio LLM associa as ações gestuais com a resposta verbal do robô a partir de um conjunto de ações que foi passado previamente em forma de prompt preliminar. Restrições explícitas no prompt limitaram o tamanho da resposta a não mais de 40 palavras, proibindo a repetição de frases e proibindo emojis ou caracteres não UTF-8, garantindo respostas concisas e claras. O formato de resposta estruturado exige que a IA responda em um formato específico com palavras-chave de ação de uma lista predefinida, reforçando assim a consistência e permitindo o processamento *downstream* ou ações de interface do usuário com base na saída da IA.

3.3. Estudo de Caso

O presente estudo foi conduzido em dois cenários distintos, nos quais o robô permaneceu estacionário, aguardando iniciativas de interação por parte dos usuários. O protocolo de ativação requer que os participantes invoquem o nome "Sophia" para estabelecer o início da interação, momento em que o sistema reconhece a intenção comunicativa do usuário. O robô foi programado para responder questões relacionadas à tecnologia e informações sobre pessoal ministerial e do Centro de Tecnologia da Informação Renato Archer (CTI). Como medida de segurança e adequação social, foram implementadas restrições temáticas que impedem discussões sobre política, religião e futebol.

O primeiro cenário foi implementado na recepção de visitantes da instituição CTI, onde o robô desempenha funções de apoio em palestras e responde questionamentos dos convidados como mostrado na Figura 2. O público-alvo abrange todas as pessoas que visitam o CTI, incluindo delegações escolares e profissionais de diversos setores. Contudo, o foco principal das visitas direcionava-se ao conhecimento da divisão de robótica, local de desenvolvimento da Sophia. Neste contexto, o robô foi configurado para fornecer informações relacionadas à robótica social e aos conceitos da Sociedade 5.0. O ambiente controlado apresenta baixos níveis de ruído externo, facilitando o diálogo e a interação entre usuários e o sistema robótico.

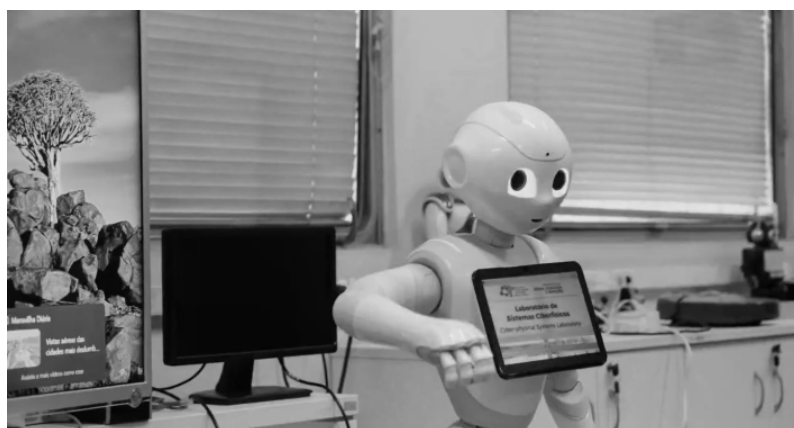


Figura 2. Robô auxiliando em visitas dentro do CTI.

O segundo cenário ocorreu durante o primeiro dia do evento "Ciência Aberta", sediado pelo CNPEM em 30 de maio de 2025. O evento registrou visitas de universidades e escolas de todo o Brasil, com estimativa de 16 mil visitantes, incluindo estudan-

tes e professores. Durante o evento, a Sophia experimentou diversos tipos de interação, observando-se que a maioria dos participantes aproximava-se do robô demonstrando curiosidade e manifestando aprovação quanto à estética do sistema, característica diferencial dos robôs sociais voltados para interação humana.

Na Figura 3, são mostradas interações com crianças no evento; à esquerda, a criança pede para cumprimentar o robô, e, à direita, a criança pede um abraço para o robô.



Figura 3. Robô interagindo com crianças no evento Ciência Aberta.

Um vídeo demonstrativo de interação com o robô pode ser visto em <https://tinyurl.com/y68pvazc>.

4. Resultados e Discussão

A Tabela 1 apresenta o resumo dos testes piloto; nela, podemos perceber que as diferentes arquiteturas de integração demonstraram viabilidade de aplicação prática de LLMs para atendimento institucional, com destaque para soluções locais de baixa latência e plataformas online de alta sofisticação linguística.

Os resultados obtidos no segundo cenário experimental demonstraram que o robô Sophia manteve níveis adequados de segurança operacional durante todo o período de interação. Não foram identificados riscos ou incidentes de segurança envolvendo os usuários, mesmo considerando a natureza espontânea das aproximações e interações. Este resultado corrobora a adequação do design de robôs sociais para ambientes públicos com alta densidade de usuários, confirmando que as medidas de segurança implementadas foram eficazes para o contexto experimental.

A integração do LLM resultou em um enriquecimento significativo das capacidades interativas do sistema robótico. O diferencial observado reside na característica não-determinística do modelo, que elimina respostas automáticas padronizadas em favor de respostas dinamicamente geradas, conferindo maior naturalidade e realismo às interações. A implementação de *prompt engineering* permitiu a inserção de uma personalidade definida à Sophia, humanizando substancialmente o diálogo e tornando a experiência interativa mais envolvente para os usuários.

Tabela 1. Comparação de Integrações de LLMs, STT e TTS com Voz Brasileira.

#	Arquitetura	Objetivo	Latência	Resultados
1	Integração Local (Offline) LLaMA 3.1:8B + Whisper + RHVoice	Pipeline local e offline 3.1	~5s	100% offline Voz natural e estável, respostas inconsistentes
2	LLM Online (Groq 70B) LLM: deepseek-r1-distill-llama-70b (Groq API) STT: Whisper TTS: RHVoice (Letícia) Prompt: Rosana	Uso de LLM online com controle por prompt institucional	~5s	Respostas altamente consistentes
3	Execução Local em Servidor (GPU) LLM: LLaMA 3.1:8B (Ollama) Infra: RTX 4060 STT: Whisper TTS: RHVoice	Avaliar desempenho local com GPU dedicada	~8s	Execução estável Uso moderado de GPU
4	Testes com Múltiplas LLMs (Servidor Central) Infra: 2x RTX 6000 (24GB VRAM) Modelos: gemma, deepseek, mistral, bode-alpaca, etc. Acesso via curl/Ollama	Avaliar desempenho e consistência de vários modelos	~12s (70B) Varia por modelo	Timeouts com prompts longos Necessário ajuste fino

Foram identificadas limitações relacionadas à latência de resposta e gargalos de comunicação; situações onde houve instabilidade do sistema exigiram reinicializações, e observou-se menor precisão nas respostas do LLM com o uso extensivo. Estas limitações são atribuídas à dependência de soluções baseadas em nuvem, que, embora ofereçam a melhor qualidade de resposta disponível, têm limitações de usabilidade.

A primeira RQ investigou quais características são determinantes para que a interação entre humanos e robôs integrados ao LLM seja eficiente e precisa. Os resultados indicaram que múltiplos fatores técnicos e contextuais influenciam diretamente na qualidade dessa interação. A análise da latência do sistema, apresentada na Figura 4, evidenciou que a transmissão de arquivos de áudio entre o computador e o robô constitui o principal gargalo temporal do pipeline, superando outras etapas em tempo médio de execução. Este achado aponta para a necessidade de otimização da comunicação de dados como estratégia central para melhorar a responsividade do sistema.

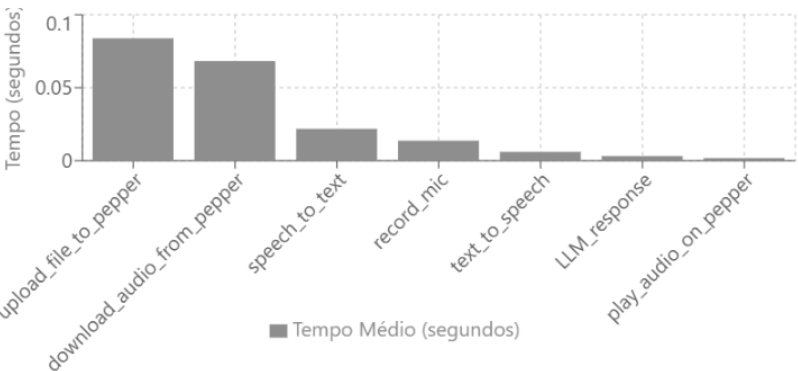


Figura 4. Gráfico do tempo médio das etapas que o robô realiza na interação.

As observações demonstraram que o LLM apresentou desempenho satisfatório na

maioria das situações, fornecendo respostas precisas e contextualmente apropriadas. A implementação de memória conversacional no prompt permitiu a recuperação de contexto em situações onde as questões não foram completamente compreendidas, aproximando-se das respostas desejáveis. Contudo, foram identificadas ocasiões em que o sistema retornava formatos de resposta inadequados, gerando falhas que impediam a separação correta entre componentes de fala e ação, comprometendo a funcionalidade geral do sistema.

A segunda RQ analisou as habilidades necessárias para robôs sociais com LLMs interagirem em ambientes públicos. Estudos de caso em contexto controlado e em evento aberto indicaram que o rastreamento facial, aprimorado por técnicas de visão computacional, é fundamental para o contato visual e engajamento. A sincronização entre fala e gestos, viabilizada pela seleção automática de animações coerentes com o discurso por meio do LLM, também se mostrou essencial para interações naturais e expressivas.

A terceira RQ abordou como os diferentes elementos do robô devem ser combinados para assegurar coerência na interação com seres humanos. A análise da integração entre componentes revelou que a coordenação entre fala e movimentos corporais mediada pelo LLM foi executada de forma eficaz. A adoção de um formato estruturado para as respostas, baseado no esquema resposta-ação, permitiu sincronização adequada entre os aspectos verbais e não verbais da comunicação.

O ambiente do evento apresentou desafios significativos para a captura de fala devido aos elevados níveis de ruído externo. Estas condições ambientais comprometeram a qualidade das interações, exigindo que os usuários se aproximassem fisicamente do robô para garantir a compreensão adequada. Simultaneamente, as respostas audíveis do robô foram prejudicadas pela interferência sonora do ambiente, reduzindo a inteligibilidade das saídas de áudio.

Embora o robô Sophia possua capacidade de operação independente de conectividade à internet para suas funções básicas, a implementação completa do sistema exige conectividade. Esta dependência demonstra a necessidade de planejamento da infraestrutura tecnológica para implementações em ambientes públicos, garantindo a disponibilidade contínua dos recursos computacionais remotos necessários para o funcionamento do sistema.

5. Conclusão

Neste artigo, apresentamos um estudo de caso em que um LLM foi utilizado em um robô para aperfeiçoar a interação com pessoas. O LLM ofereceu respostas verbais para conversação com humanos juntamente a um gesto correspondente ao que foi dito verbalmente. O robô foi capaz de responder adequadamente às pessoas, considerando inclusive um prompt inicial de instruções dado à LLM para responder adequadamente. A prova de conceito foi capaz de dar boas impressões aos interlocutores em ambos os casos abordados.

Dentre os potenciais trabalhos futuros, destaca-se aplicar métodos para modularização de voz de acordo com a situação, de modo a dar mais naturalidade às respostas e melhorar a resposta do LLM para situações de interações por meio de RAG e *fine tuning*, assim como métodos próprios de geração de comportamentos não-verbais. Também espera-se utilizar modelos multimodais para produzir a resposta do robô não so-

mente a partir do que é dito pela pessoa, mas também através da postura e gestos desta pessoa.

Agradecimentos

Os autores agradecem o suporte financeiro da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) - processo nº 2020/07074-3.

Referências

- Aldebaran Robotics (2025). Choregraphe overview. Acessado em 20-06-2025.
- Asuzu, K., Singh, H., and Idrissi, M. (2025). Human-robot interaction through joint robot planning with large language models. *Intell. Serv. Robotics*, 18:261–277.
- Fink, J. (2012). Anthropomorphism and human likeness in the design of robots and human-robot interaction. In *Social Robotics: 4th International Conference, ICSR 2012, Chengdu, China, October 29-31, 2012. Proceedings 4*, pages 199–208. Springer.
- Gkournelos, C., Konstantinou, C., and Makris, S. (2024). An llm-based approach for enabling seamless human-robot collaboration in assembly. *CIRP Annals*.
- Goodrich, M. A., Schultz, A. C., et al. (2008). Human–robot interaction: a survey. *Foundations and trends® in human–computer interaction*, 1(3):203–275.
- Kim, C. Y., Lee, C. P., and Mutlu, B. (2024). Understanding large-language model (llm)-powered human-robot interaction. *2024 19th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 371–380.
- Koike, A. and Mutlu, B. (2023). Exploring the design space of extra-linguistic expression for robots. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, pages 2689–2706.
- Lewin, K. (1946). Action research and minority problems. *Journal of social issues*, 2(4):34–46.
- Liu, H., Zhu, Y., Kato, K., Tsukahara, A., Kondo, I., Aoyama, T., and Hasegawa, Y. (2024). Enhancing the llm-based robot manipulation through human-robot collaboration. *IEEE Robotics and Automation Letters*, 9:6904–6911.
- Said, Y., Atri, M., and Tourki, R. (2011). Human detection based on integral histograms of oriented gradients and svm. In *2011 International Conference on Communications, Computing and Control Applications (CCCA)*, pages 1–5. IEEE.
- Schreiter, T., Ruppel, J. V., Hazra, R., Rudenko, A., Magnusson, M., and Lilienthal, A. J. (2025). Evaluating efficiency and engagement in scripted and llm-enhanced human-robot interactions. *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 1608–1612.
- SEBRAE (2025). Entenda o que são as práticas de esg. Acessado em 19-06-2025.
- Serapio-García, G., Safdari, M., Crepy, C., Sun, L., Fitz, S., Romero, P., Abdulhai, M., Faust, A., and Matarić, M. (2025). Personality traits in large language models.
- SOFTBANK ROBOTICS (2025). Meet pepper: The robot built for people. Acessado em 19-06-12025.
- Yin, R. K. (2017). *Case study research and applications: Design and methods*. Sage publications.
- Zhang, C., Chen, J., Li, J., Peng, Y., and Mao, Z. (2023). Large language models for human–robot interaction: A review. *Biomimetic Intelligence and Robotics*, 3(4):100131.