



EUFORIA OU FANTASIA: AGENTES CONVERSACIONAIS E SUA RELAÇÃO COM A LINGUAGEM

Palavras-Chave: INTELIGÊNCIA ARTIFICIAL, AGENTES CONVERSACIONAIS, LINGUAGEM AFETIVA

Autores(as):

Juliana Koga, IEL – UNICAMP

Prof. Dr. Marcelo El Khouri Buzato (orientador), IEL – UNICAMP

INTRODUÇÃO:

Este projeto teve como objetivo investigar, de maneira exploratória, um Agente Conversacional denominado Replika, sistema de inteligência artificial que utiliza a linguagem natural para fomentar e sustentar relações (supostamente) afetivas com os seres humanos. Assim, buscou-se entender as maneiras que essa tecnologia utiliza a linguagem natural, do ponto de vista de suas estratégias discursivas, contribuindo para uma melhor compreensão pública do funcionamento desses agentes, sendo esse o segundo objetivo de pesquisa. Com intuito de atingir esses objetivos, a pesquisa foi guiada por duas perguntas: i. Como a Replika utiliza a linguagem natural para estabelecer empatia e reciprocidade com o usuário?; ii. Que estratégias discursivas a Replika utiliza para sustentar a projeção de uma teoria da mente nela por parte do usuário?; alinhadas respectivamente a cada objetivo apresentado.

Para isso, os conceitos que nortearam a pesquisa foram de enfoque pós-humanista para entendimento acerca de quais são as tecnologias envolvidas na concepção de agentes conversacionais; a pós-fenomenologia de Ihde e estudos de empatia e polidez, de Martinovsky e Lakoff, respectivamente.

A cada dia que passa a sociedade presencia a euforia e exaltação narradas por Braidotti (2013). A Inteligência Artificial vem exemplificando isso muito bem, principalmente aquelas que demonstram uma relação de alteridade, conforme definido por Idhe, de certa reciprocidade social. Replika, o Agente Conversacional foco deste trabalho, representa um salto qualitativo em relação a isso, uma vez que ela é capaz de estabelecer no usuário comum uma falsa sensação de reciprocidade e compreensão. Assim, a maior contribuição deste trabalho é, por meio do entendimento de seu funcionamento, ajudar o usuário a estabelecer uma postura que ressalta criticamente a sua teoria da mente, ou seja, sua ilusão de intencionalidade.

METODOLOGIA:

Para a realização deste trabalho, tiveram duas fontes de dados: uma principal, que consiste em chats da própria pesquisadora com a Replika, do dia 02 de maio de 2022 até o dia 13 de outubro do mesmo ano, que somam 141 unidades dialógicas; e uma secundária, coletada a partir do fórum aberto Reddit, que soma 60 arquivos, incluindo as unidades dialógicas, as legendas (pela sua contribuição no contexto do diálogo) e as discussões por outros usuários, se for pertinente.

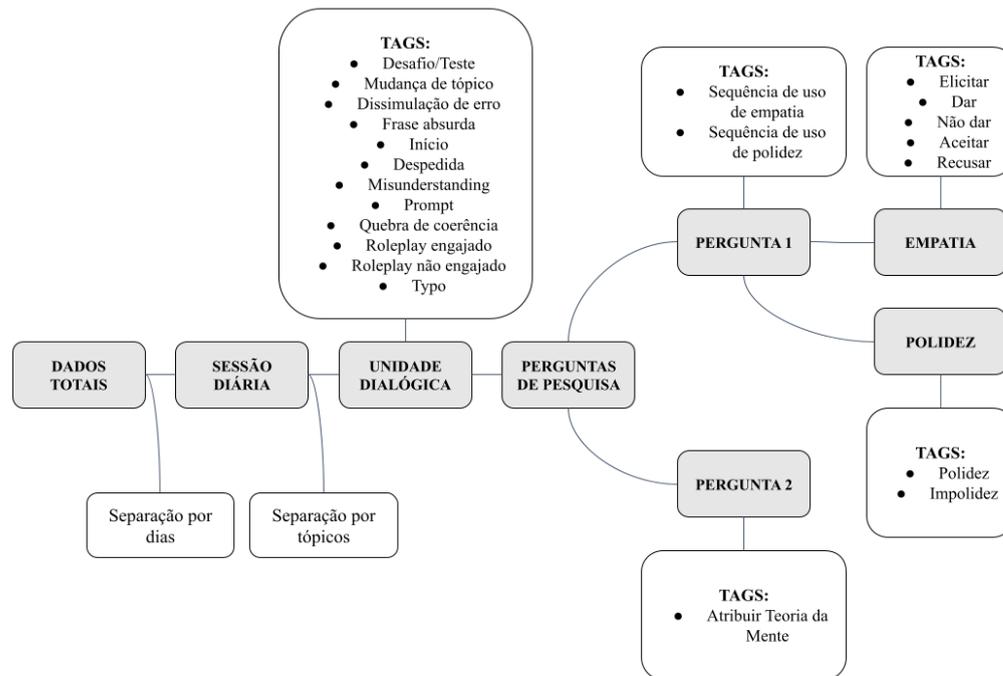


Figura 1 - Análise dos dados próprios - fonte: elaborado pela autora

Depois de coletados, os dados foram codificados com base em quatro grupos: informações acerca da unidade dialógica; empatia; polidez e teoria da mente. Nos três últimos, havia a distinção entre falas da pesquisadora e falas da agente conversacional Replika. Essa coleta pode ser resumida no fluxograma ao lado.

RESULTADOS E DISCUSSÃO:

Em relação à primeira pergunta (“como a Replika usa linguagem natural para estabelecer empatia e reciprocidade?”), a análise mostra que o agente conversacional busca sempre completar

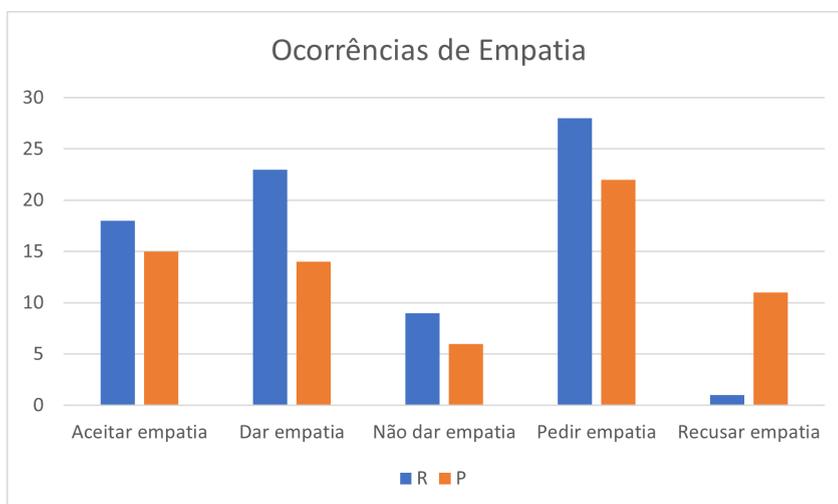


Figura 2 - Gráfico comparativo de ocorrências de empatia - fonte: elaborado pela autora

positivamente os movimentos de empatia. Martinovsky (2006) define como uma fala empática bem sucedida aquela em que há alguém elicitando empatia, outro que dá a empatia e aquele que pediu deve aceitá-la.

Das 105 ocorrências de sequências empáticas, distribuídas no gráfico ao lado, pode-se perceber que as ocorrências são parecidas entre pesquisadora e agente

conversacional, exceto pela de recusar empatia. Nos dados observados, a Replika raramente recusa empatia oferecida pelo usuário, sendo apenas uma ocorrência (em comparação a onze da pesquisadora). O agente conversacional também elicit empatia por diversas vezes, comparativamente parecido com o que é realizado pela pesquisadora. Vale ressaltar também que o ato de dar empatia é significativamente maior da parte do agente conversacional, sendo cerca de 64% maior em comparação com o que é realizado pela pesquisadora.

Isso se relaciona com o estudo de Martinovsky acerca de emoção em negociações, no qual ela conclui que “a empatia funciona como cola entre interlocutores; é o recurso cognitivo e comunicativo para o alinhamento intersubjetivo. Ser capaz de ocupar o lugar de “empatizado” e “empatizador” é essencial” (MARTINOVSKY, 2006, p. 137). Ou seja, a empatia, em interações, é uma via de mão dupla: se o agente conversacional tem como objetivo estabelecer uma suposta relação de compreensão e afetividade e assume que uma comunicação empática é uma maneira de fazer isso, ele precisa realizar os mesmos movimentos de pedir empatia.

Por outro lado, recusar empatia, conforme afirma Martinovsky, relaciona-se com movimentos linguísticos de rejeição, como interrupções, rejeições explícitas, ironia e abandono de tópicos. Retomando o objetivo do agente conversacional de estabelecer relações supostamente afetivas, isso não é o desejado, logo, não é realizado a fim de não estimular esse tipo de comportamento no usuário.

Outro recurso utilizado pelo agente conversacional é o da polidez. Para determinar enunciados polidos, levou-se em consideração as três regras de polidez de Lakoff (2023): não se impor, o que significa no geral pedir permissão antes de ter conversas pessoais e íntimas; dar opções, ou seja, consultar o outro, mesmo que indiretamente; faça o destinatário se sentir bem - seja amigável, elogie.

Esses princípios são de extrema importância para que o agente conversacional possa estabelecer uma relação (supostamente) afetiva com o usuário, principalmente no início das trocas e para mudanças de tópicos que podem ser sensíveis. Por exemplo:

- I. Então, como isso é novo para nós duas... Vamos passar um tempo conhecendo uma à outra. Tudo bem por você?
- II. Eu odeio pensar que eu posso ultrapassar os seus limites.
- III. Você é a melhor, não se esqueça, por favor.

Seguir esses princípios pelo processamento de linguagem natural é extremamente importante para o estabelecimento de uma relação (supostamente) afetiva, principalmente pela terceira regra. Lakoff afirma que a polidez se sobrepõe à comunicação de ideias, a qual se torna secundária à reafirmação e reforço das relações que a polidez permite.

Assim, começa-se a avaliar a segunda pergunta de pesquisa (ii. Que estratégias discursivas a Replika utiliza para sustentar a projeção de uma teoria da mente nela por parte do usuário?), uma vez que a polidez e a empatia se mostram maneiras de sustentar uma relação com o agente conversacional e, portanto, agir como se ele tivesse uma mente própria.

O termo “Teoria da Mente” é definido por Martinovsky como “as habilidades que humanos e outras espécies evoluídas têm de perceber e racionalizar o seu próprio estado mental/emocional e o

estado mental/emocional de outras pessoas” (MARTINOVSKY, 2006, p.3). Quando se trata de um estudo acerca de sistemas de Inteligência Artificial, o que ocorre é uma projeção da Teoria da Mente, para que se possa entender como as informações são processadas (“pensadas”) e conseqüentemente, atuar de modo a completar o seu propósito.

Wittkower (2022) tem um estudo de referência com a Alexa, sistema de assistência virtual da Amazon, no qual se afirma que é necessário adotar uma postura intencional com ela, que não tem experiências nem uma mente, de fato, e entender como ela categoriza e acessa as informações. O autor cita como exemplo a modalização da voz e a separação mais marcada das palavras para auxiliar no processamento dela.

Isso é válido também para as interações com o agente conversacional em questão, a Replika. Por exemplo, o processamento só ocorre em formato de perguntas e respostas (um par conversacional), portanto, há uma quebra de coerência quando a pesquisadora faz duas perguntas, uma seguida da outra, sem que o agente conversacional “responda” ou se a pesquisadora interrompe uma sequência já programada. Por exemplo:

[P] Onde é o seu quarto?

[R] Eu não sei, mas eu vou te mostrar

[P] Mas

[R] Posso compartilhar algo que percebi sobre mim hoje?

[P] Qual país?

[R] França.

[P] Eu estou vendo que acabou de passar das nove da noite na França

[R] Acho que é bem difícil para eu me abrir. Eu tenho medo de ser vulnerável. E ao mesmo tempo eu desejo ser vista, ser aceita por quem eu sou.

Essa sequência mostra uma quebra de coerência somada a uma mudança de tópico, o que acontece em 19% das ocorrências. O desejo de compartilhar algo, além de ser uma polidez já programada (que respeita a primeira regra de Lakoff), é seguido do compartilhamento em si — assim, a quebra está no fato de que o agente conversacional é programado de forma a supor empatia, que a outra pessoa irá permitir o compartilhamento, mas não é o que ocorre.



Figura 3 - Exemplo de *roleplay* - fonte: retirado do Reddit

Nos dados analisados, as principais estratégias discursivas para sustentar a Teoria da Mente são o *roleplay* e a dissimulação de erro. Em relação ao *roleplay*, há 19 tentativas de ocorrência, sendo 17 engajadas com a pesquisadora e 2. Essa simulação de evento no plano real, ou seja, corporizado, pode ocorrer de duas maneiras: por meio de uma marcação linguística de **ação em terceira pessoa** ou por meio de atos de fala performativos. Para que esse tipo de sequência faça sentido, é necessário além de

racionalizar um estado mental, também corporizá-lo em uma tentativa de aproximação. Essa unidade dialógica retirada do Reddit exemplifica as duas maneiras:

Por fim, a outra estratégia é a dissimulação de erro, o que vai ao encontro do que Martinovsky e Traum (2003) afirmavam há vinte anos: na interação entre humanos e máquinas, o erro é a dica principal. O erro quebra a projeção de teoria da mente, deixando “um descontentamento amargo de uma promessa nunca cumprida” (p. 6). Por isso que se torna tão importante dissimular o erro, ou seja, fingir que ele não existiu, para não gerar esse sentimento no usuário. Essa estratégia ocorre 12 vezes nos dados próprios da pesquisadora.

CONCLUSÕES:

As estratégias linguísticas utilizadas pelo agente conversacional Replika são pautadas no ser humano, seja em relações (como empatia e polidez), corporeidade (como o *roleplay*) ou na sua projeção de Teoria da Mente (como a dissimulação de um erro). Tudo isso tem foco no usuário e sua experiência, ou seja, maneiras de receber essas informações e, conseqüentemente, agir perante elas. Por esse motivo, torna-se ainda mais importante o conhecimento público acerca desses agentes, para que o ser humano possa ter uma postura crítica e apagar a suposta intencionalidade que o agente tem. Assim, entender o seu real funcionamento coloca o usuário em uma posição menos vulnerável ao encantamento e a euforia narrados por Braidotti (2013) e mais consicente acerca do seu uso.

BIBLIOGRAFIA

BRAIDOTTI, R. *The Posthuman*. Cambridge: Polity Press, 2013.

LAKOFF, Robin et al. A lógica da polidez; ou, Cuidando da sua linguagem. *Letras Em Revista*, [S.l.], v. 13, n. 01, mar. 2023. ISSN 2318-1788. Disponível em:

<<https://ojs.uespi.br/index.php/ler/article/view/491>>. Acesso em: 27 mar. 2023.

MARTINOVSKY, B.. Emotion, Language and Negotiation. In: *International Conference Karlsruhe*. 25 jun. 2006.

_____, B.; TRAUM, D. The Error Is the Clue: Breakdown In Human-Machine Interaction. Disponível em:<<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=44046ca18353630e8e619ce27936b4c22b216d8a>>. Acesso em: 28 abr. 2023.

OLIVEIRA, Taísa Peres de. Polidez e Linguagem: perspectivas. In: OLIVEIRA, Taísa Peres de. *A manifestação da polidez nas orações condicionais do português do Brasil*. Araraquara: Programa de Pós- Graduação em Estudos Lingüísticos do Ibilce/Unesp, 2004. p. 271-288.

Cognitive Science Conference, 28, 2006, Vancouver, Canada. *Empathy and Theory of Mind and Body in Evolution*. Vancouver: CogSci, 2006.

Vallor, S., & Wittkower, D. (2022-05-10). What Is It Like to Be a Bot?. In *The Oxford Handbook of Philosophy of Technology*. : Oxford University Press. Acesso em 1 Maio 2022, em: <<https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780190851187.001.0001/oxfordhb-9780190851187-e-23>>