

## **Flexibilização da criação de chatbots**

Leonardo Paiva da Costa Pereira<sup>1</sup>, Thiago Raposo Milhomem de Carvalho<sup>2</sup>

**RESUMO:** Os chatbots têm sido cada vez mais utilizados na sociedade. Esse aumento de interesse deve-se, principalmente, à sua capacidade de processar dados e, com isso, se comunicar em linguagem natural sobre determinados temas. Este trabalho de conclusão de curso visa desenvolver uma ferramenta que integra dados provenientes de diferentes fontes (arquivos TXT, arquivos CSV e arquivos PDF). A principal funcionalidade da ferramenta é permitir que o usuário insira dados de seu interesse para, em seguida, gerar um chatbot baseado em Modelos de Linguagem Natural (LLMs), que será capaz de interagir com os usuários finais e responder a perguntas específicas relacionadas aos dados fornecidos. O objetivo final é criar um sistema que facilite a análise e a interpretação de dados complexos por meio de uma interface conversacional intuitiva, melhorando a acessibilidade e a compreensão das informações.

**Palavras-chave:** Chatbot. Inteligência Artificial. LLM. Processamento de Linguagem Natural. RAG

## **1 INTRODUÇÃO**

Bot é a abreviação de "robot" e refere-se a um software desenvolvido para imitar ações humanas repetidamente e simular uma interação humano-computador. Os primeiros bots eram simples, com pouca capacidade de interação. Hoje, no entanto, a tecnologia evoluiu, permitindo que eles sejam programados para interagir

de forma mais natural. Chatbots são assistentes virtuais que simulam uma conversa para responder perguntas de maneira que o usuário tenha a impressão de estar conversando com uma pessoa real.

Os chatbots têm despertado grande interesse tanto no meio acadêmico quanto no mercado, devido às suas interfaces amigáveis e à naturalidade que oferecem nas interações. Eles utilizam uma vasta gama de ferramentas e conceitos da computação, sendo os principais a Inteligência Artificial (IA), com ênfase em Aprendizado de Máquina (Machine Learning) e Processamento de Linguagem Natural (PLN). Para que a comunicação seja eficaz, as sentenças processadas pelos chatbots são analisadas em termos morfológicos, sintáticos, semânticos e pragmáticos (TEIXEIRA; MENEZES, 2003).

De acordo com Mowbray (2012, p. 1), bots são "pessoas que atuam de acordo com um programa de software, em vez de serem controladas diretamente por um usuário humano". Ou seja, são programas desenvolvidos para interagir com seres humanos de forma transparente e que podem se adaptar ao contexto, respondendo a perguntas, fornecendo informações ou fazendo sugestões.

A IA surgiu com a aplicação das redes neurais em 1943 e, desde então, tem sido utilizada com sucesso em diversos setores da sociedade, como saúde, turismo, pesquisa e negócios. O interesse comercial por agentes conversacionais tem aumentado expressivamente, pois a capaci-

<sup>1</sup>Graduando em Ciência da Computação. E-mail: leonardo.paiva@iesb.edu.br

<sup>2</sup>Doutor em Física. E-mail: thiago.raposo@iesb.edu.br

dade desses agentes de manter diálogos em linguagem natural os torna muito atrativos como facilitadores em vários contextos.

Diante desse cenário, o objetivo deste trabalho é propor uma abordagem que permita aumentar a rapidez e o suporte na utilização de grandes volumes de dados, demonstrando que a importância dessa solução reside em sua capacidade de democratizar o acesso a dados complexos, eliminando a necessidade de intermediários ou especialistas em TI. A interface conversacional facilita a interação com os dados, permitindo que qualquer usuário, independentemente de seu nível técnico, possa realizar consultas e obter insights relevantes com facilidade. Além disso, o uso de chatbots baseados em LLMs oferece respostas ágeis e precisas, aprimorando a eficiência e a qualidade da análise de dados.

## 1.1 JUSTIFICATIVA

Atualmente, o volume de dados gerados por empresas e organizações cresce exponencialmente, exigindo a criação de ferramentas eficientes que permitam a análise, organização e interpretação desses dados de forma acessível e rápida. Dessa forma, é possível agregar valor aos dados, transformando-os em informações valiosas. Tradicionalmente, a consulta e o processamento de dados demandam conhecimento técnico especializado, o que pode se tornar uma barreira significativa para muitos usuários, especialmente em ambientes corporativos e de atendimento ao cliente.

Considerando a dificuldade de manipular grandes volumes de dados com alta complexidade e provenientes de fontes heterogêneas, justifica-se o desenvolvimento de uma ferramenta que integre diferentes fontes de dados, como arquivos CSV, PDF e TXT, associada ao suporte de tecnologias tais quais Inteligência Artificial, Aprendizado de Máquina, Processa-

mento de Linguagem Natural e Modelos de Linguagem Natural (LLMs). A capacidade de gerar chatbots que interajam diretamente com essas bases de dados e ofereçam respostas imediatas e precisas pode transformar a maneira como empresas e indivíduos consultam e interpretam informações, oferecendo um conjunto de recursos eficazes para o sucesso no processo de análise de dados.

Ao facilitar a interação com dados complexos, a ferramenta proposta pode contribuir significativamente para a tomada de decisões mais rápidas e informadas, especialmente em ambientes onde a precisão e a agilidade na interpretação das informações são cruciais.

## 1.2 PROBLEMA

Atualmente, um dos grandes desafios enfrentados por organizações que lidam com grandes volumes de dados é a dificuldade em acessar, interpretar e utilizar essas informações de maneira eficiente. Com a crescente disponibilidade de dados armazenados em diferentes formatos, a falta de integração entre essas fontes gera obstáculos no processo de consulta e análise. Para muitos usuários, a complexidade envolvida na manipulação de dados estruturados ou semiestruturados é uma barreira, pois exige conhecimentos técnicos avançados, como habilidades em SQL, programação e análise de dados.

Adicionalmente, as ferramentas tradicionais de consulta de dados, como sistemas de Business Intelligence (BI) e relatórios automatizados, ainda requerem configurações manuais complexas e tempo considerável para extrair insights relevantes. Esse cenário se torna ainda mais desafiador em organizações que precisam responder rapidamente a questões estratégicas e operacionais. A ausência de uma interface acessível e intuitiva para consulta e interpretação de dados impede que usuários sem especialização

técnica obtenham as informações de que precisam, reduzindo a eficácia e a produtividade nas tomadas de decisão.

Outro aspecto crítico do problema é a falta de ferramentas que consigam proporcionar uma interação dinâmica com os dados. A maioria das soluções disponíveis não são capazes de oferecer uma experiência conversacional natural, em que os usuários possam simplesmente fazer perguntas em linguagem natural e obter respostas claras, precisas e contextualizadas.

Portanto, o problema central que este trabalho busca resolver é a inexistência de uma ferramenta que permita a integração de diferentes fontes de dados e a consulta eficiente dessas informações por meio de uma interface conversacional simples e acessível. A criação de um chatbot que utilize Modelos de Linguagem Natural (LLMs) e seja capaz de responder à perguntas específicas sobre os dados fornecidos visa justamente preencher essa lacuna, democratizando o acesso à informações complexas e tornando a análise de dados mais acessível a todos os tipos de usuários.

### 3 OBJETIVOS

O intuito deste trabalho é desenvolver uma ferramenta que facilite a análise e interpretação de dados, integrando diferentes fontes (arquivos .CSV, .PDF e .TXT) e permitindo a criação de um chatbot baseado em Modelos de Linguagem Natural (LLMs) para interagir com usuários finais. Para atingir esse propósito, foram definidos objetivos gerais e específicos.

#### 3.1 OBJETIVO GERAL

Espera-se desenvolver uma plataforma interativa que integre dados complexos provenientes de diferentes fontes e, com base em Modelos de Linguagem Natural, automatize o processo

de análise de dados, oferecendo uma interface conversacional intuitiva. Centralizando a necessidade de criar uma ferramenta que simplifique o acesso e o uso de grandes volumes de dados, oferecendo aos usuários finais uma forma acessível de interagir com as informações sem a necessidade de conhecimentos técnicos aprofundados.

#### 3.2 OBJETIVOS ESPECÍFICOS

- Desenvolver uma interface de usuário que permita a importação e configuração de dados provenientes de arquivos .Txt, .Pdf e .Csv.
- Implementar um mecanismo de processamento de dados que transforme e adapte os dados importados para serem utilizados por um chatbot.
- Criar um chatbot que utilize LLMs para responder às perguntas dos clientes finais sobre os dados inseridos, proporcionando respostas relevantes e contextualmente precisas.
- Avaliar a eficácia do chatbot em termos de precisão das respostas e facilidade de uso, realizando testes com usuários finais para assegurar que a ferramenta atenda às necessidades de interpretação e consulta de dados.

### 4 FUNDAMENTAÇÃO TEÓRICA

O, presente, trabalho se fundamenta em diversas áreas do conhecimento, sendo elas, Ciência de Dados, Inteligência Artificial (IA), Processamento de Linguagem Natural (PLN) e Integração de Dados, abordando teorias e metodologias que dão suporte à construção da ferramenta proposta.

Uma das bases centrais do estudo está na Inteligência Artificial, mais especificamente no desenvolvimento de Modelos de Linguagem Natural (LLMs), que são capazes de compreender e gerar linguagem humana de forma contextualizada. Autores como Christopher D. Manning, Hinrich Schütze e Dan Jurafsky são referências na área de Processamento de Linguagem Natural, especialmente com a obra "Foundations of Statistical Natural Language Processing"(1999), que oferece uma sólida base para entender como os modelos de linguagem podem ser aplicados à interação conversacional com dados.

Além disso, o desenvolvimento de Large Language Models, como o GPT (Generative Pretrained Transformer), é um dos principais avanços recentes em IA. Vaswani et al. (2017), com o artigo "Attention is All You Need", introduziu o modelo de transformador, que revolucionou o campo de PLN ao melhorar significativamente a capacidade de modelos de linguagem em gerar texto coerente e realizar tarefas de compreensão. Esse modelo forma a base para diversos LLMs que podem ser usados em chatbots, sendo essencial para a tecnologia proposta neste trabalho.

HALL e LINDZEY (1984) destacam a importância da persona (ou personalidade) no desenvolvimento de chatbots. A criação de personas específicas para um chatbot pode resultar em interações mais produtivas, pois, dependendo do objetivo do usuário, uma persona mais formal ou informal, por exemplo, pode influenciar significativamente a coerência e a eficácia das respostas fornecidas pelo programa. Além disso, uma persona bem elaborada pode aprofundar a experiência do usuário, proporcionando a sensação de estar se comunicando com um ser humano, tornando a interação mais natural e envolvente.

Os trabalhos citados destacam a relevância

de cada área mencionada para o desenvolvimento de chatbots, especialmente no que se refere à aplicação de modelos de linguagem à interação conversacional com dados. Esses modelos não só facilitam a interpretação e análise dos dados, mas também possibilitam a criação de personas personalizadas, o que torna as interações mais naturais e intuitivas. O presente trabalho se diferencia ao propor uma experiência conversacional inovadora, na qual o usuário pode integrar suas próprias bases de dados e personalizar as interações por meio da escolha de diferentes personas. Essa abordagem oferece um nível de flexibilidade e personalização superior, permitindo que o usuário selecione a persona mais adequada para suas necessidades específicas, tornando o processo de comunicação mais dinâmico e eficiente.

## **5 DESENVOLVIMENTO**

### **5.1 REQUISITOS**

Para viabilizar este trabalho, foi necessário trabalhar com diferentes bibliotecas em Python, sendo as principais: 'express', empregada para viabilizar a conexão do front-end com o back-end; 'langchain', *extensive plitters', utilizada para segmentar o texto*

A linguagem escolhida para o desenvolvimento foi Python, principalmente devido ao seu rico ecossistema de bibliotecas e ferramentas, que são amplamente utilizados em Inteligência Artificial e Machine Learning. A facilidade de uso do Python também permite que os desenvolvedores foquem na lógica do projeto em si, sem a necessidade de lidar com complexidades técnicas excessivas. Além disso, Python é amplamente compatível com APIs de IA, como a da OpenAI, facilitando a criação de aplicações que utilizam modelos avançados de linguagem

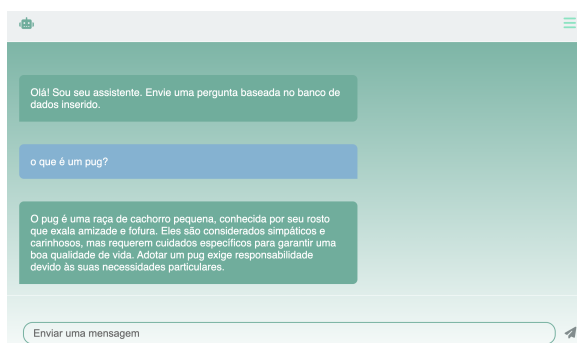
para tarefas como geração de texto e tradução.

Quanto ao Front-end da aplicação, a qual é a parte responsável pela interação com o usuário, foram utilizadas as linguagens HTML e CSS, para estruturar e organizar o conteúdo da página web, e estilizar a apresentação visual dos elementos HTML.

## 5.2 INTERFACE DO USUÁRIO

Para proporcionar uma interação eficaz entre o usuário e a aplicação, o projeto tem como objetivo desenvolver um site onde o usuário possa fazer login ou se cadastrar, caso ainda não tenha uma conta. Com os dados carregados, o usuário poderá visualizar seus chats previamente criados com os agentes específicos, criar novos chats utilizando o mesmo arquivo, mas com diferentes agentes, ou até mesmo iniciar novos chats vinculados à diferentes arquivos.

A interação entre o usuário e o robô de conversação ocorrerá por meio de uma interface gráfica acessível em qualquer dispositivo com um navegador web. A comunicação será realizada por meio de perguntas, fornecidas pelo usuário e enviados à IA por meio de APIs. A IA processará as informações recebidas através de processos como tokenização, gerando uma resposta que será devolvida à aplicação por meio das APIs, permitindo que o robô responda ao questionamento do usuário.



## 5.3 INTEGRAÇÃO COM A OPEN AI

Programadores, experientes ou iniciantes nos campos teórico e prático da inteligência artificial, têm a oportunidade de desenvolver projetos que utilizam seus recursos. Neste trabalho, está sendo integrada a API da OpenAI, cuja tecnologia ChatGPT traz soluções e oferece os recursos necessários para o desenvolvimento de chatterbots.

A OpenAI utiliza técnicas de Modelos de Linguagem Natural (LLMs), como o deep learning, treinadas com grandes quantidades de texto. Esses modelos são capazes de gerar respostas coerentes e contextualmente relevantes para diversas solicitações, sendo particularmente eficazes para responder perguntas, o principal objetivo deste projeto. Eles compreendem as relações estatísticas entre palavras e frases, identificando quais têm maior probabilidade de aparecer juntas e como a ordem das palavras afeta o significado de uma frase. O LLM é composto por uma rede neural – um modelo matemático complexo que aprende a reconhecer padrões nos dados. A precisão das previsões é constantemente avaliada e a rede neural é ajustada para melhorar sua exatidão, repetindo o processo até que a rede seja capaz de prever corretamente a próxima palavra na sequência.

O treinamento do ChatGPT envolve a maximização da probabilidade condicional de cada resposta com base em uma sequência de interações anteriores. O modelo é alimentado com uma sequência de tokens de entrada e treinado para prever o próximo token correto, processo este conhecido como tokenização. Esse processo é repetido várias vezes, utilizando técnicas como amostragem aleatória e treinamento por reforço, para melhorar a qualidade das respostas geradas.

A comunicação com a API ocorre por meio de requisições HTTP, enviadas e retornadas como respostas aos links utilizados.

Conforme descrito: "O ChatGPT é treinado



numa tarefa de ‘conclusão de texto’, em que pega numa parte incompleta do texto e gera uma continuação plausível. É capaz de compreender e gerar texto numa variedade de estilos e domínios, desde conversas casuais a escrita técnica, informação factual, descrições, respostas a perguntas e muito mais”(SOUZA; BARBOSA; ALMEIDA, 2023, p. 66).

#### **5.4 ACESSO AOS DADOS**

Uma das etapas mais importantes durante o desenvolvimento do projeto é a aquisição dos dados sobre os quais o chatbot será especialista. Esses dados podem ser extraídos das seguintes fontes externas: arquivos em formato .csv, .txt ou .pdf.

Para cada tipo de fonte, há um mecanismo específico de extração de dados. A linguagem Python facilita essa etapa, oferecendo diversas bibliotecas e módulos integrados para a leitura de arquivos em variados formatos - sendo pandas utilizada para arquivos .csv e pypdf para arquivos .pdf -.

Após a extração, os dados são carregados em uma base de dados local. Dessa forma, quando o usuário solicita uma informação ao chatbot, esta já se encontra previamente armazenada na base de dados, permitindo que o retorno seja feito no tempo de resposta do banco.

#### **5.5 LLM RAG**

Para o processamento dos dados, este projeto adota o conceito de Retrieval-Augmented Generation (RAG), aliado a Modelos de Linguagem de Grande Escala (LLMs). Essa abordagem melhora significativamente a precisão e a qualidade das respostas, ao utilizar apenas informações relevantes extraídas de uma base de dados anexada, em vez de depender exclusivamente dos dados utilizados no treinamento do

modelo.

No funcionamento do sistema, o texto da pergunta é convertido em embeddings — representações vetoriais do conteúdo. O mesmo processo é aplicado à base de dados fornecida pelo usuário, que é previamente fragmentada em trechos menores, os quais também são codificados em embeddings. Em seguida, realiza-se uma comparação de similaridade entre os embeddings da pergunta e os da base de dados. Quando são encontradas correspondências relevantes, os trechos associados aos embeddings mais semelhantes são utilizados como base para a geração da resposta. Dessa forma, o modelo fundamenta suas respostas em informações verificáveis e atualizadas, extraídas diretamente dos documentos anexados.

Essa técnica se diferencia substancialmente do processamento tradicional baseado apenas em prompt engineering. Muitas vezes, os LLMs não possuem conhecimento sobre informações específicas ou recentes, devido à limitação de seu treinamento. Isso pode levar à geração de respostas imprecisas ou até incorretas, baseadas em fontes desatualizadas ou não confiáveis.

A principal vantagem do RAG está em sua capacidade de fornecer respostas mais confiáveis e contextualizadas, além de reduzir custos com re-treinamento de modelos para domínios específicos. Como os dados utilizados na resposta podem ser verificados diretamente pelo usuário, também há um aumento na transparência e na confiança nas informações fornecidas pelo sistema.

### **6 RESULTADOS PARCIAIS**

Até o presente momento, diversas etapas do desenvolvimento da ferramenta proposta foram concluídas com sucesso. Inicialmente, foram implementados os mecanismos de importação de arquivos, possibilitando ao usuário carregar

suas próprias fontes de dados.

Em seguida, foi realizada a integração com a API da OpenAI, permitindo a criação de chatbots baseados em Modelos de Linguagem Natural (LLMs). A funcionalidade foi testada com sucesso: após a inserção dos dados, o chatbot gerado é capaz de responder perguntas com base no conteúdo dos arquivos anexados, utilizando a técnica de *Retrieval-Augmented Generation* (RAG). Esse processo garante maior precisão nas respostas, uma vez que o modelo busca, no próprio conteúdo fornecido, os trechos mais relevantes para formular a resposta ao usuário.

A interface web, desenvolvida com HTML, CSS e JavaScript, está operacional e permite que o usuário realize login, cadastre-se na plataforma, visualize chats anteriores e crie novos chats com diferentes agentes. Essa estrutura foi projetada com foco na usabilidade, visando facilitar a navegação e o gerenciamento das interações.

Além disso, já foi implementado o armazenamento local das bases de dados processadas, permitindo rápida recuperação da informação no momento da consulta, sem a necessidade de reprocessamento.

Ainda que os resultados sejam preliminares, os testes realizados até o momento demonstram a robustez da abordagem adotada e indicam um cenário promissor para os próximos ciclos de desenvolvimento e avaliação do sistema. Isso reforça o potencial da aplicação para ser utilizada por usuários sem conhecimento técnico avançado, democratizando o acesso a informações complexas.

## 7 PRÓXIMOS PASSOS

Como próximos passos do projeto, pretende-se aprimorar a eficiência do chatbot, ampliando o contexto das respostas com base não apenas nos arquivos anexados, mas também em infor-

mações fornecidas diretamente pelos usuários. Além disso, está em planejamento a migração das interações com os bots para o WhatsApp, de modo que a aplicação web assuma uma função administrativa, voltada à gestão dos chatbots, enquanto as conversas acontecerão diretamente pela plataforma de mensagens, facilitando o acesso e a usabilidade. Outro avanço previsto é a ampliação das fontes de dados compatíveis com o sistema, incluindo a possibilidade de integração com bancos de dados relacionais, como o PostgreSQL.

## 8 CONCLUSÃO

Conclui-se que a aplicação proposta facilitará a integração da tecnologia no cotidiano das pessoas, além de potencialmente aumentar a eficiência na análise de grandes volumes de dados e fornecer chatbots especializados para diversos temas e finalidades.

**ABSTRACT:** Chatbots have been increasingly used in society. This growing interest is primarily due to their ability to process data and thus communicate in natural language about specific topics. This end-of-course project aims to develop an innovative tool that integrates data from different sources (Excel files and a PostgreSQL database). The tool's main functionality is to allow users to input data from their chosen sources and then generate a chatbot based on Large Language Models (LLMs). This chatbot will interact with end-users and answer specific questions related to the provided data. The ultimate goal is to create a system that facilitates the analysis and interpretation of complex data through an intuitive conversational interface, improving both accessibility and comprehension of the available information.

**Keywords:** Chatbot. Artificial intelligence. NLM. Natural language processing. RAG.

## REFERÊNCIAS

TEIXEIRA, J.; MENEZES, J. Estudo de Linguagem Natural Aplicada a Interação Humano-Computador. Revista Ciência da Computação, 2003. MOWBRAY, M. Bots in the Digital Age. IEEE Spectrum, 2012.

Universidade Aberta de Portugal. Flexibilidade na Criação de Chatbots a partir de Bases de Dados. Repositório Aberto, Universidade Aberta, 2019. Disponível em: . Acesso em: 10 out. 2024.

FERREIRA, Leandro; UCHOA, Joaquim. Implementação de Chatbots no Ambiente Educacional: Desafios e Oportunidades. Revista Científica UNIFIL, Londrina: Centro Universitário Filadélfia, 2016. Disponível em: . Acesso em: 15 out. 2024.

FERREIRA, Leandro; UCHOA, Joaquim. A Influência da Tecnologia na Educação: Um Estudo sobre Chatbots. Educadores Dia a Dia, Paraná: Secretaria de Estado da Educação, 2012. Disponível em: . Acesso em: 15 out. 2024.

MACHADO, Ronaldo; SOUZA, Guilherme. Aplicações de Chatbots no Ensino a Distância: Um Estudo de Caso. RENOTE - Revista Novas Tecnologias na Educação, Porto Alegre: Universidade Federal do Rio Grande do Sul (UFRGS), 2018. Disponível em: . Acesso em: 16 out. 2024.

CHAKRABORTY, Parth; LUGER, George. Chatterbots: A Conversational Interface for the Web. University of New Mexico, 2015. Disponível em: . Acesso em: 23 out. 2024.

HALL, C. S.; LINDZEY, G. Teorias da personalidade. 1. ed. São Paulo: E.P.U, 1984.

MARIN, Ana Beatriz de F. MARINABOT – Construção de um simulador de diálogo. 2021. Trabalho de Conclusão de Curso (Tecnologia em Análise e Desenvolvimento de Sistemas) – Faculdade de Tecnologia de Americana, Centro Paula Souza, 2021. Disponível em: . Acesso em: 23 out. 2024.

SOUZA, Ana Luísa Santos de. Desenvolvimento de um Assistente Virtual para Atendimento ao Público com Foco na Interação e Personalização. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) – Universidade Federal da Bahia, Salvador, 2022. Disponível em: . Acesso em: 31 out. 2024.

SOUZA, João Paulo Silva; BARBOSA, Luciano Rodrigues; ALMEIDA, Bruno José de. Assistentes Virtuais Inteligentes para Atendimento ao Cliente: Revisão e Perspectivas. Brazilian Journal of Systems, [S. l.], v. 5, n. 2, p. 85-97, 2023. Disponível em: . Acesso em: 31 out. 2024.

MELO, Wendel Henrique. Tutorial NLTK. Uberlândia: Universidade Federal de Uberlândia, Faculdade de Computação, [s.d.]. Disponível em: . Acesso em: 01 nov. 2024.

PEREIRA, M. S.; SILVA, J. A. C. Uso de Inteligência Artificial para Processamento de Linguagem Natural em Chatbots. Revista INJEI, São Paulo, v. 2, n. 1, p. 1-14, 2023. Disponível em: . Acesso em: 04 nov. 2024.

LIU, Linjun; ZHANG, Tianle; CAO, Ruibo; JIANG, Yifan; KIM, Jiho; LEVINE, Alexander. Assessing the Reliability of Out-of-Distribution Sensitivity Estimation Methods for Neural Networks. In: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS (NEURIPS), 2023, Proceedings... Disponível em: . Acesso em: 11 nov. 2024