

Desenvolvimento de uma Aplicação para Análise de Desempenho de Grandes Modelos de Linguagem

Orientanda: Nicole Barbosa Gomes

Orientador: Prof Dr. João Paulo Papa





Índice

01

Introdução

02

Fundamentação Teórica

03

Metodologia

04

Resultados

05

Considerações Finais



01

Introdução

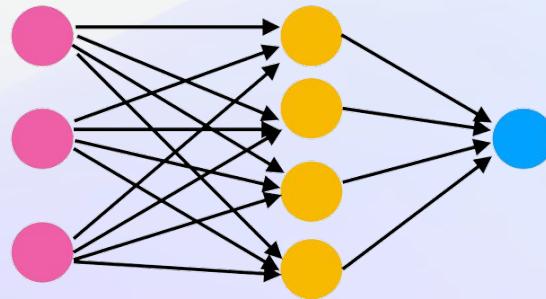
Introdução

- necessidade humana de comunicação
- evolução das máquinas
- novo desafio: gerar comunicação com os computadores



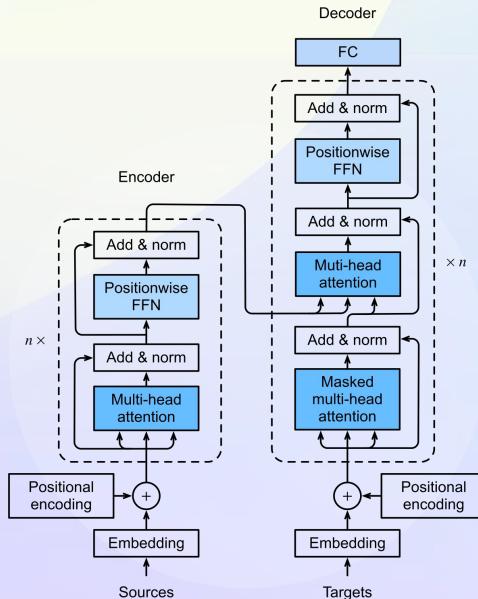
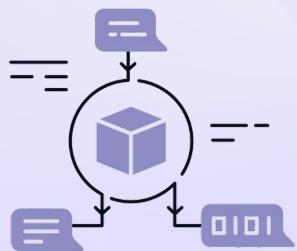
Introdução

- comunicação intuitiva
- processamento de linguagem natural
- redes neurais artificiais são essenciais
- criação dos transformers



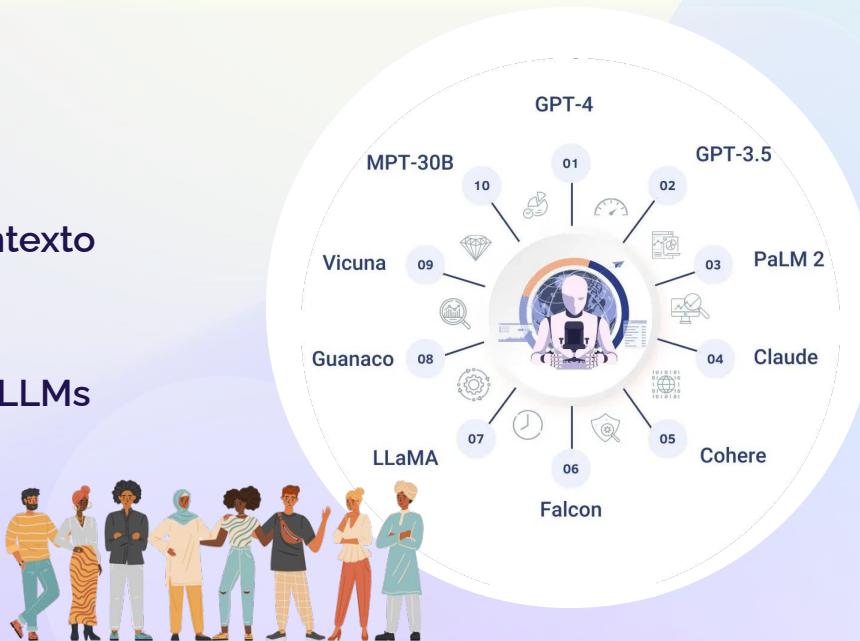
Introdução

- transformers
- large language models
- versões diferentes do mesmo tipo de modelo



Problemática

- linguagens são complexas
 - palavras dependem do contexto
 - variações culturais
- pode ser difícil escolher um dos LLMs



Justificativa

- grande número de aplicações para modelos desse tipo
- gerar respostas cada vez mais naturais
 - envolvem o usuário



Objetivos

- possibilitar a avaliação da influência que as características dos diferentes modelos têm nos resultados gerados por eles
- site que gera respostas a partir de instruções do usuário
- usuário pode escolher o melhor resultado



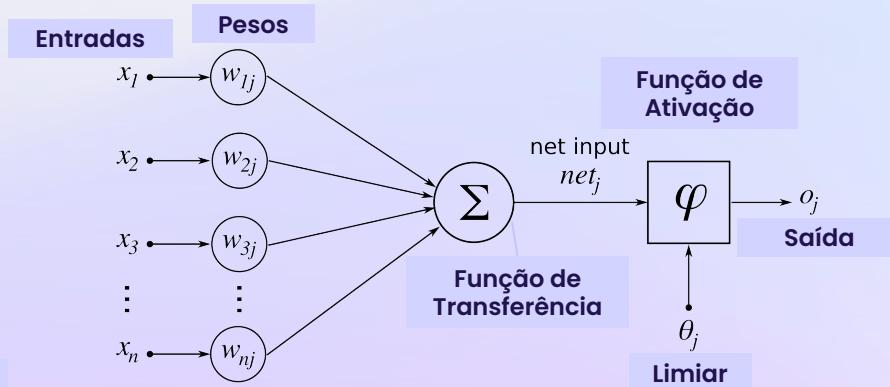
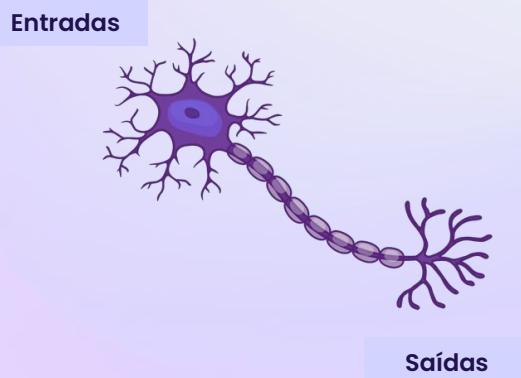


02

Fundamentação Teórica

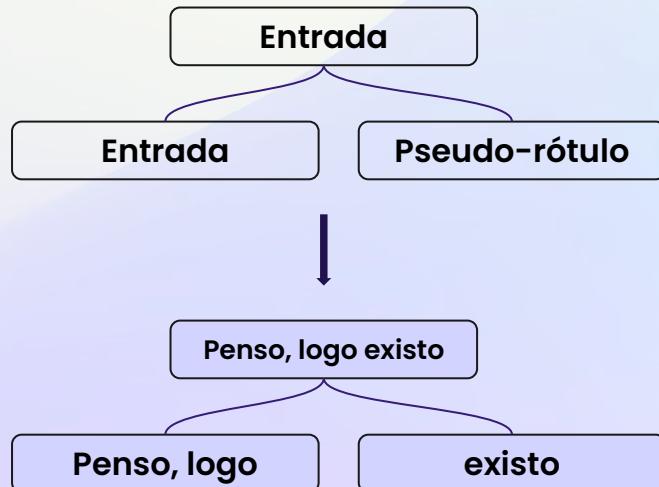
Redes Neurais Artificiais

- estruturas que podem resolver problemas mais complexos
- baseadas nas redes do cérebro humano
- caracterizadas por unidades menores capazes de passar informações adiante



Redes Neurais Artificiais

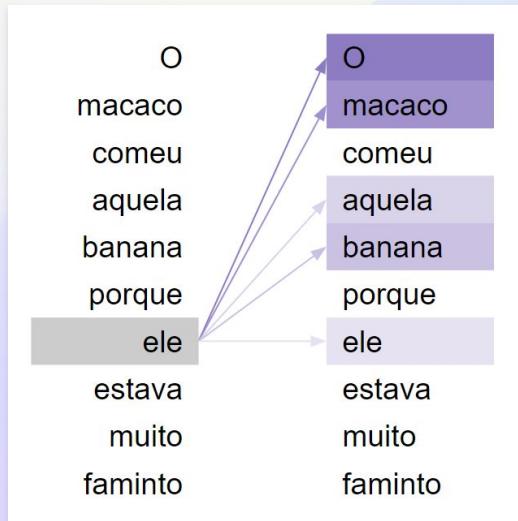
- aprendizado auto-supervisionado
 - um dos métodos utilizados para treinar a rede
 - diminui o problema da limitação de quantidade de dados rotulados
 - produz representações descritivas dos dados
 - usado para modelos de linguagem



Transformers

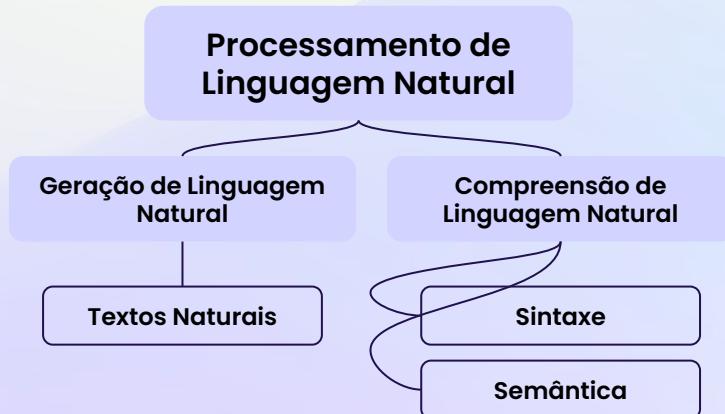
- mecanismos de atenção
 - identificar dependências entre palavras
 - entender contexto

- distribuir espécies de pesos (atenção) para partes da sentença
 - grau de relação entre as palavras



Processamento de Linguagem Natural

- objetivo do campo é analisar e estudar a capacidade dos humanos e das máquinas de interagirem diretamente
- dificuldades
 - regras implícitas
 - conhecimento de mundo obtido por humanos
 - ambiguidade
- contexto resolve boa parte dos problemas



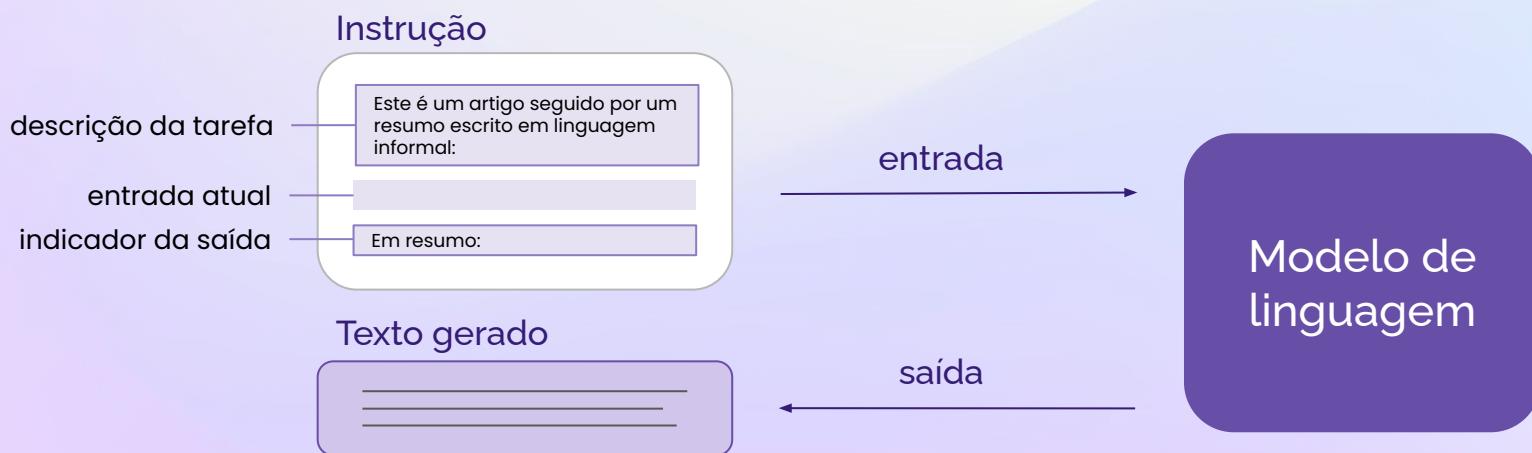
Large Language Models

- modelos de linguagem
 - baseados em distribuições probabilísticas
 - capazes de prever palavras em uma sequência
- treinados com grandes quantidades de dados
- usam, em sua maioria, os *transformers* em sua estrutura
 - ótima interpretação de contexto
- se diferenciam pelas características usadas para o pré-treinamento
 - corpos de texto utilizados



Prompt Design

- gerar melhores respostas
- construir entradas a fim de atingir resultados mais precisos
- aumenta a eficácia dos LLMs





03

Metodología

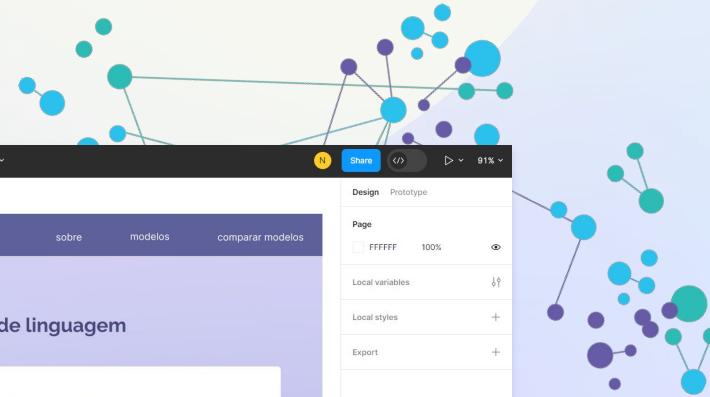


Ferramentas

→ Figma

- criação de um protótipo

The screenshot shows the Figma application interface with a design prototype for a website titled "trylang". The main content area features a purple header with the title "compare o desempenho de grandes modelos de linguagem". Below the header, there are two columns of text blocks. The left column discusses "Modelos de linguagem" and the right column discusses "Grandes modelos de linguagem". At the bottom, there are three numbered steps: 1. insira uma instrução no campo determinado, 2. escolha os modelos responsáveis por gerar as respostas, and 3. analise as respostas geradas pelos modelos. A large "começar agora" button is centered at the bottom. The Figma interface includes a sidebar with layers, assets, and a properties panel on the left, and a toolbar and status bar at the top.

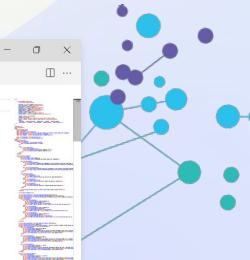




Ferramentas

→ Visual Studio Code

- desenvolvimento do código



A screenshot of the Visual Studio Code interface showing an open PHP file named "index.php". The code handles a POST request for a form submission, processes various variables like \$entrada, \$info_tam_entrada, \$info_resposta, etc., and inserts them into a MySQL database using an INSERT INTO query. The code also includes meta tags for the HTML document.

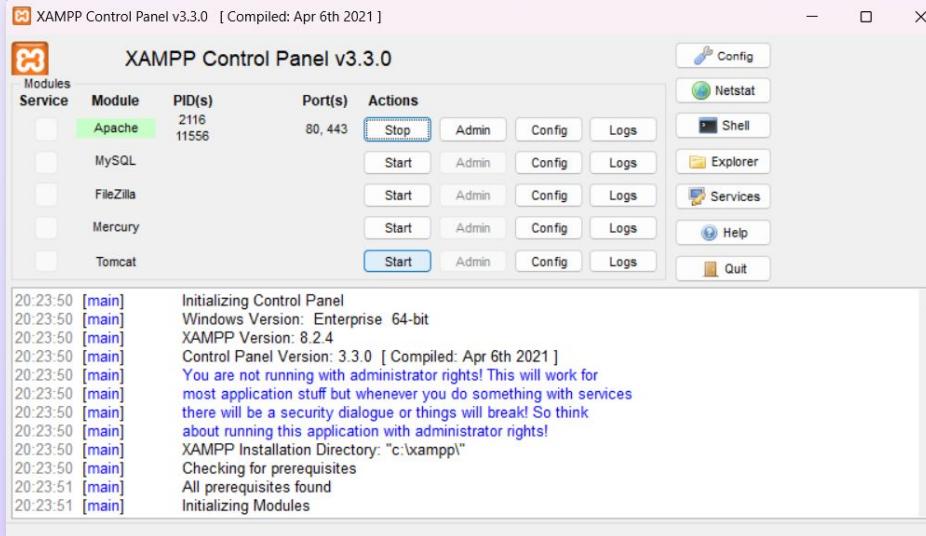
```
File Edit Selection View Go Run Terminal Help ← → Projeto TCC
EXPLORER PROJETO TCC
index.php main.js style.css
index.php
main.js
respostas.php
index.php x
1 <?php
2
3 if (isset($_POST['submit'])){
4     include_once('respostas.php');
5
6     $entrada = addslashes($_POST['entrada-info']);
7     $info_tam_entrada = $_POST['tam-entrada-info'];
8     $info_modelo = $_POST['info-modelo'];
9     $resposta = addslashes($_POST['info-resposta']);
10    $stam_resposta = $_POST['info-tam-resposta'];
11    $tempo_resposta = $_POST['info-tempo-resposta'];
12    $temperatura = $_POST['temperatura-info'];
13    $tokens = $_POST['tokens-info'];
14    $empresa_modelo = $_POST['info-empresa-modelo'];
15    $justificativa = addslashes($_POST['justificativa-usuario']);
16
17    $result = mysqli_query($conexao,
18        "INSERT INTO feedback
19            (entrada, tam_entrada, temperatura, maximo_tokens, empresa_modelo,
20             modelo, resposta, tamano_resposta, tempo_resposta, justificativa_usuario)
21        VALUES
22            ('$entrada', '$info_tam_entrada', '$temperatura', '$tokens', '$empresa_modelo',
23             '$info_modelo', '$resposta', '$stam_resposta', '$tempo_resposta', '$justificativa')");
24    }
25
26    <!DOCTYPE html>
27    <html lang="pt-br">
28        <head>
29            <title>trylang</title>
30            <meta charset="UTF-8">
31            <meta name="author" content="Nicole">
32            <meta name="description" content="Aplicação com o objetivo de comparar grandes modelos de linguagem">
33            <meta name="keywords" content="lme, ia, modelos de linguagem">
34            <meta name="viewport" content="width=device-width, initial-scale=1.0">
35
36            <link rel="stylesheet" href="estilo/style.css">
37
In 1, Col 6 Spaces: 4 UFT-8 CRLF PHP Go Live 0
```



Ferramentas

→ XAMPP

- testar a aplicação localmente

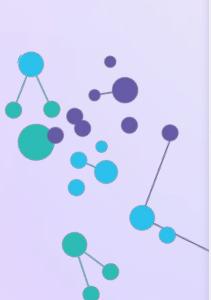




Ferramentas

→ MySQL Workbench

- armazenar as informações capturadas pela aplicação



The screenshot shows the MySQL Workbench interface with the following details:

- Navigator:** Shows the database schema with the "feedback" table selected under the "db-trylang" schema.
- Query Editor:** Displays the query `select * from feedback`. The results grid shows the following data:| idfeedback | entrada | tamano_entrada | temperatura | maximo_tokens | empresa_modelo | modelo | resposta | tam |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | escreva um texto sobre o mascote de ... | 214 | 1 | 200 | A121 Studio | j2-ultra | Super Barato é um supermercado novo na cida... | 500 |
| 2 | escreve um pequeno texto, do tamanho... | 96 | 0.5 | 2048 | OpenAI | text-davinci-003 | Wicked é um musical de Broadway baseado no... | 522 |
| 3 | escreve um pequeno texto, do tamanho... | 96 | 0.7 | 200 | A121 Studio | j2-ultra | Wicked is the Broadway smash hit that tells the... | 500 |
| 4 | escreva um texto sobre o mascote de ... | 214 | 0.9 | 300 | Cohere | command | Super Barato, the superhero of savings! With ... | 678 |
- Output:** Shows the execution log:

```
1 20:52:04 selected *from feedback LIMIT 0, 1000
```
- Help:** A tooltip in the top right corner provides information about context help.



Linguagens Utilizadas



HTML



CSS



JavaScript

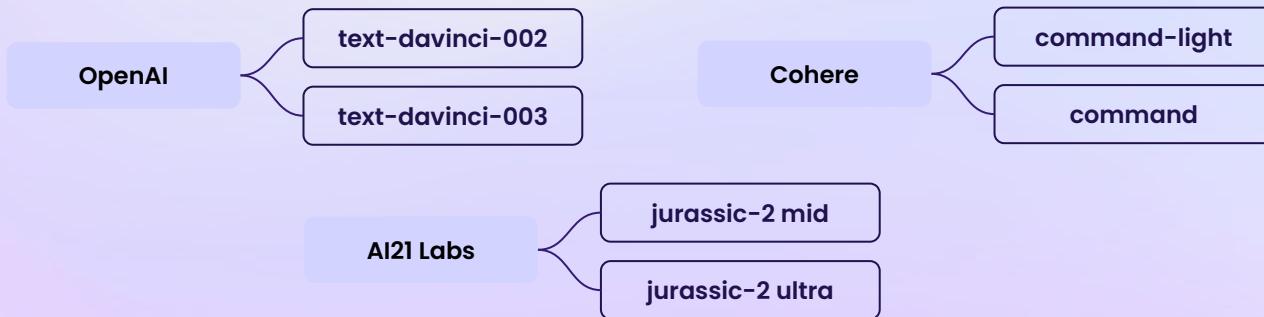


PHP



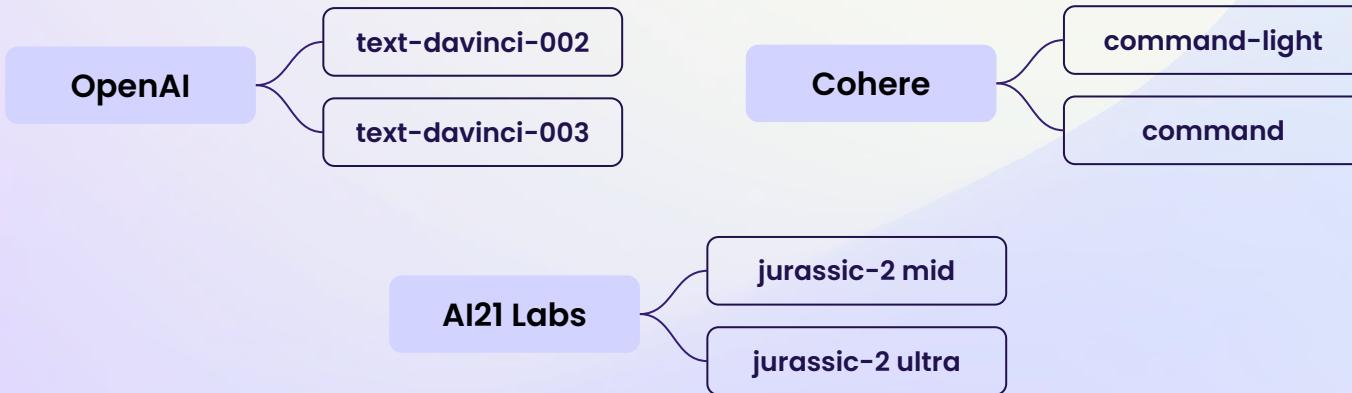
Modelos

- escolhidos com base na confiabilidade das empresas e na capacidade de seguirem instruções
- permitir que o usuário compare modelos de diferentes empresas
- aplicação registra as opiniões do usuário e ajuda na criação de um panorama maior das performances





Modelos



- um modelo mais poderoso e um menos
- usuário pode escolher um modelo de cada empresa para as comparações



Desenvolvimento do Protótipo

The screenshot displays a Figma design interface for a website prototype. The left sidebar lists various pages and components, including 'Modelos (Agosto) 2' which is currently selected. The main canvas shows 'Frame 1' containing two rows of cards. Each card provides a brief description of a model and links for more information or usage.

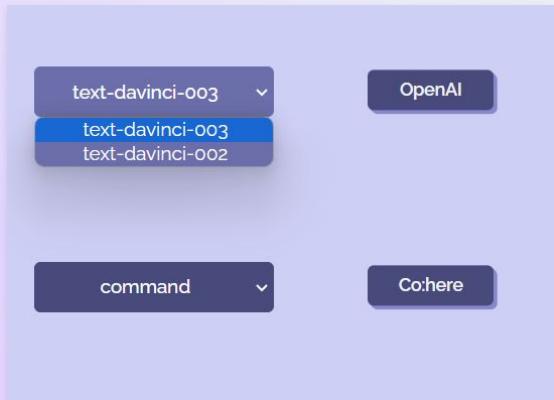
Card Type	Model Name	Description	Actions
Text Model	text-davinci-003 OpenAI	"pode realizar tarefas com melhor qualidade, maior saída e consistência do que os modelos curie, babbage, ou ada"	[saiba mais] [usar modelo]
	text-davinci-002 OpenAI	"possui capacidades similares ao text-davinci-003, mas é treinado com fine-tuning supervisionado"	[saiba mais] [usar modelo]
Text Model	jurassic 2 - mid AI21	"oferece aprimorada capacidade de geração de texto, realizando tarefas de linguagem com um maior grau de complexidade"	[saiba mais] [usar modelo]
	jurassic 2 - ultra AI21	"pode realizar tarefas com melhor qualidade, maior saída e consistência do que os modelos curie, babbage, ou ada"	[saiba mais] [usar modelo]
Command Model	command Co:here	"possui capacidades similares ao text-davinci-003, mas é treinado com fine-tuning supervisionado"	[saiba mais] [usar modelo]
	command-light Co:here	"oferece aprimorada capacidade de geração de texto, realizando tarefas de linguagem com um maior grau de complexidade"	[saiba mais] [usar modelo]

- para cada seção, foi criada uma imagem com o layout planejado
- facilitou a criação do site



Desenvolvimento da Aplicação

- a estrutura básica foi desenvolvida através do uso de HTML
 - o estilo foi alterado através do CSS
- conferir atividades aos elementos da aplicação
- uso de javascript



escolha dos modelos



link externo



Desenvolvimento da Aplicação

- para usar os modelos, as APIs foram necessárias
- geração de chaves
- usadas no código

```
graph LR; A((MySQL Workbench)) --> B[criação de conexão]; A --> C[criação de database]; A --> D[criação de tabela]
```

MySQL
Workbench

criação de conexão

criação de database

criação de tabela

- conexão com o banco de dados
 - para que os dados processados pelo site fossem armazenados
 - envio feito através do PHP e de comandos SQL
- configuração do MySQL Workbench



Desenvolvimento da Aplicação

- produção das seções do site
- seções introdutórias

The screenshot shows the TryLang application's user interface. At the top, there is a dark blue header bar with the logo 'trylang' on the left and navigation links 'sobre', 'modelos', and 'comparar modelos' on the right. Below the header, the main title 'compare' is displayed in large, bold, white letters, followed by the subtitle 'o desempenho de grandes modelos de linguagem' in a smaller, white font. The page content is divided into two main sections. The left section contains text about language models and a 'saiba mais' button. The right section contains text about large language models and a 'saiba mais' button. At the bottom, there are three numbered steps: 1. 'insira uma instrução no campo determinado', 2. 'escolha os modelos que geram as respostas', and 3. 'analise as respostas geradas pelos modelos'. A large blue button labeled 'começar agora' is positioned at the very bottom center.

trylang

sobre modelos comparar modelos

compare

o desempenho de grandes modelos de linguagem

Modelos de linguagem são um conjunto de soluções, algoritmos e técnicas que tenham relação de linguagem com as máquinas. Um modelo desse tipo é probabilístico de uma linguagem natural e capaz de gerar probabilidades de sequências de palavras, baseadas em corpos de textos de uma ou várias linguagens a partir das quais o modelo é treinado.

saiba mais

Grandes modelos de linguagem são treinados em grandes quantidades de dados de texto para aprender padrões e relacionamentos entre entidades no idioma. Os LLMs podem realizar muitos tipos de tarefas de linguagem, como tradução de idiomas, análise de sentimentos, conversas de chatbot e muito mais.

saiba mais

1 insira uma instrução no campo determinado

2 escolha os modelos que geram as respostas

3 analise as respostas geradas pelos modelos

começar agora



Desenvolvimento da Aplicação

- produção das seções do site
- seções introdutórias

SAIBA MAIS

as empresas por trás dos modelos

responsáveis por criar os modelos utilizados

sobre a empresa

OpenAI



[VER MAIS](#)

sobre a empresa

Co:here



[VER MAIS](#)

sobre a empresa

AI21



[VER MAIS](#)



Desenvolvimento da Aplicação

- área destinada à descrição dos modelos
- usuário pode escolher os que serão usados para as comparações

SAIBA MAIS

sobre os modelos

é possível escolher um modelo de cada empresa para as comparações de resultados

text-davinci-003 OpenAI	text-davinci-002 OpenAI	jurassic 2 - mid AI21
"pode realizar tarefas com melhor qualidade, maior saída e consistência do que os modelos curie, babbage, ou ada"	"possui capacidades similares ao text-davinci-003, mas é treinado com fine-tuning supervisionado"	"oferece aprimorada capacidade de geração de texto, realizando tarefas de linguagem com um maior grau de complexidade"
saiba mais	saiba mais	saiba mais
usar modelo	usar modelo	usar modelo

jurassic 2 - ultra AI21	command Co:here	command-light Co:here
"é uma escolha ideal para a maioria das complexas tarefas de processamento de linguagem e aplicações de geração de texto"	"modelo conversacional, capaz de seguir instruções, que realiza tarefas de linguagem com alta qualidade e confiabilidade."	"um modelo que representa uma versão menor e mais rápida do command. Quase tão capaz quanto, mas muito mais rápido"
saiba mais	saiba mais	saiba mais
usar modelo	usar modelo	usar modelo



Desenvolvimento da Aplicação

- incentivar o usuário a usar os princípios de *prompt design*
- *prompts* pré-definidos para exemplificar as dicas oferecidas

APRENDA

como escrever as instruções

detalhar objetivo

se a finalidade do usuário é obter um texto escrito de maneira sarcástica, ou comercial a fim de atrair clientes para seu estabelecimento, o modelo pode gerar textos que correspondam a essas necessidades

[ver exemplo](#)

especificar formato

o formato desejado no texto final também deve ser especificado. Por exemplo, se o objetivo for obter um texto escrito no formato de notícia jornalística ou um poema. O usuário deve especificar tais detalhes

[ver exemplo](#)

definir tamanhos

o tamanho do texto deve ser determinado a fim de direcionar o modelo de linguagem. Caso o texto for pequeno ou apenas uma frase, o usuário deve incluir essa informação em sua instrução

[ver exemplo](#)



Desenvolvimento da Aplicação

- incentivar o usuário a usar os princípios de *prompt design*
- *prompts* pré-definidos para exemplificar as dicas oferecidas

detalhar objetivo

se a finalidade do usuário é obter um texto escrito de maneira sarcástica, ou comercial a fim de atrair clientes para seu estabelecimento, o modelo pode gerar textos que correspondam a essas necessidades

[ver exemplo](#)

Digite uma entrada:

escreva um texto sobre o mascote de um supermercado, um super-herói cujo poder é encontrar os melhores preços. O objetivo do texto deve ser atrair clientes a conhecerem o novo estabelecimento chamado 'Super Barato'



Desenvolvimento da Aplicação

- seção dedicada a alterar as propriedades dos modelos
- temperatura e quantidade máxima de *tokens*

PERSONALIZE

propriedades

temperatura

"A temperatura é um parâmetro importante que determina a "criatividade" do modelo de linguagem. Um valor mais alto de temperatura pode gerar respostas mais criativas, mas também pode levar a respostas menos precisas ou confusas."

[ler mais](#)

quantidade máxima de tokens

"Os tokens podem ser considerados como pedaços de palavras. Eles não são formados a partir do inicio ou fim das palavras de um texto, mas podem incluir espaços ou, até mesmo, sub-palavras. A seguinte relação pode ser útil: 1 token equivale a, aproximadamente, 4 caracteres ou 3/4 de palavra."

[ler mais](#)

alterar



Desenvolvimento da Aplicação

- seção dedicada a alterar as propriedades dos modelos
- temperatura e quantidade máxima de *tokens*

The screenshot shows a user interface for modifying model properties. At the top, there is a header with a "ler mais" button. Below the header, a large dark blue rectangular area contains the text "altere os valores". To the right of this area, there are two input fields: one for "temperatura" and another for "quantidade máxima de tokens", both with associated input boxes. At the bottom right, there are two buttons: "aplicar mudanças" and "voltar ao padrão".



Desenvolvimento da Aplicação

- comparação dos modelos
- escolha dos modelos

Digite uma entrada:

gerar respostas

text-davinci-003 ▾ **OpenAI**

command ▾ **Cohere**

jurassic 2 - ultra ▾ **Ai21**

limpar respostas



Desenvolvimento da Aplicação

- usuário insere uma entrada
 - tanto ela, quanto a temperatura e a quantidade máxima de *tokens* são enviadas como parâmetros para a API da empresa

...
...

Digite uma entrada:

escreva uma instrução...

gerar respostas

- quando o usuário gera as respostas, as requisições são enviadas as APIs das empresas
- quando as respostas forem devolvidas, elas são apresentadas no espaço específico da empresa

text-davinci-003 ▾

OpenAI

|||||



Desenvolvimento da Aplicação

- informações sobre a entrada e saída do usuário
- usuário pode escrever justificativa

Informações da entrada

Tamanho da entrada: 214

Temperatura: padrão

Máximo de tokens: padrão

Informações da saída

Resposta 1 Modelo: text-davinci-003 Quantidade de caracteres: 733 Tempo de resposta: 7.717 s

Resposta 2 Modelo: command Quantidade de caracteres: 834 Tempo de resposta: 15.903 s

Resposta 3 Modelo: j2-ultra Quantidade de caracteres: 503 Tempo de resposta: 5.164 s

por que foi a melhor resposta?

digite aqui...

Enviar



Desenvolvimento da Aplicação

- maior quantidade de respostas
- restrições para o envio ao banco de dados

Informações da entrada

Tamanho da entrada: 214

Temperatura: padrão

Máximo de tokens: padrão

Informações da saída

Resposta 1 Modelo: text-davinci-003 Quantidade de caracteres: 733 Tempo de resposta: 7.717 s

Resposta 2 Modelo: command Quantidade de caracteres: 834 Tempo de resposta: 15.903 s

Resposta 3 Modelo: j2-ultra Quantidade de caracteres: 503 Tempo de resposta: 5.164 s

por que foi a melhor resposta?

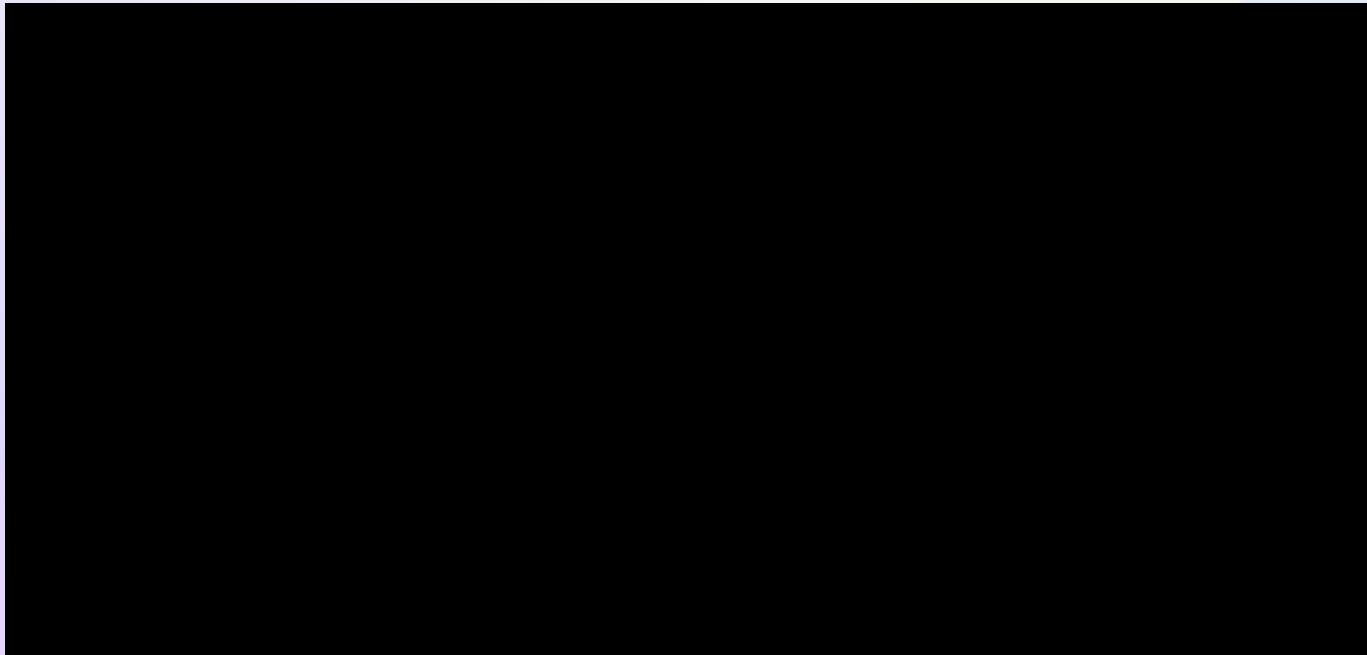
digite aqui...

Enviar



Desenvolvimento da Aplicação

• • •





04

Resultados



Análise dos Resultados

- aplicação hospedada localmente
- coleta de dados em grande escala
- para ilustrar o funcionamento, uma pesquisa foi feita
 - divulgada em menor escala
- ⋮
- ⋮
- ⋮
- MySQL Workbench permite que os dados armazenados sejam exportados
- outros aplicativos podem ser usados para a análise





Experimento

- na pesquisa, perguntas foram usadas para entender as opiniões dos usuários sobre as respostas fornecidas pelos modelos
- duas perguntas foram necessárias para comparar somente os modelos mais poderosos e, depois, os menos poderosos de cada empresa

The screenshot displays a user interface for a survey or experiment. It consists of two stacked rectangular boxes, each containing a question and several options for users to select from.

Top Box:

Escolha a melhor resposta para a instrução:
"escreva um slogan, em português, para uma livraria"

"Leia mais, viva melhor: a melhor livraria para você!"
 "Porto das leituras que vale a pena"
 "Vivendo a vida através das palavras"

Bottom Box:

Escolha a melhor resposta para a instrução:
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"

A Livraria do Saber é uma livraria especializada em livros de auto-ajuda, desenvolvimento pessoal e espiritualidade. Oferecemos um ambiente tranquilo e acolhedor para que você possa explorar e aprender sobre os diversos tópicos abordados em nossos livros.
 A new library called "Livraria do Saber" is coming to town! It is a book haven, where readers can find a vast collection of titles to explore, learn, and be inspired by. The library is modern and sleek, with a welcoming atmosphere that encourages patrons to stay and read for hours. The staff is knowledgeable and friendly, always on hand to help visitors find the perfect book. Livraria do Saber is a great addition to the community, and I can't wait to visit!

Livraria do Saber é uma nova livraria especializada em livros de ciência, tecnologia, filosofia, ciências humanas e outros temas que tratam do conhecimento humano. Tem uma seleção de livros de qualidade, publicados em português e em inglês, e oferece um ambiente acolhedor e confortável para os clientes lerem e pesquisarem. A livraria também oferece serviços adicionais, como aulas de leitura e pré-sala de leitura, para aumentar a experiência dos clientes e ajudá-los a encontrar livros



Experimento

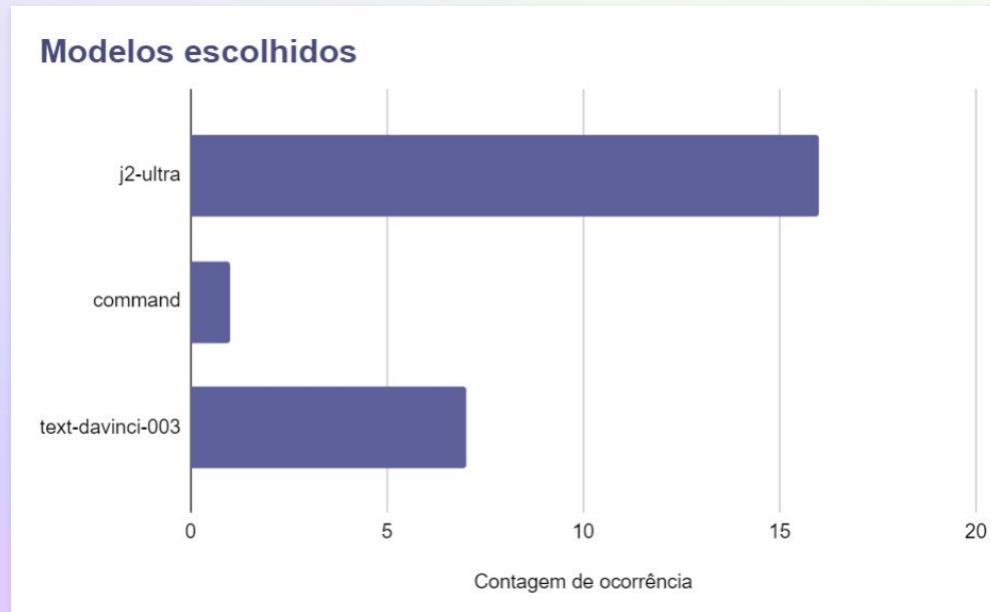
- no total, 24 respostas foram obtidas
- respostas exportadas para uma planilha, assim como aconteceria através do MySQL Workbench
- as outras características das respostas geradas foram adicionadas a planilha para fazer as comparações

entrada	tamanho_entrada	temperatura	maximo_tokens	empresa_modelo	modelo	
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.5	2048	OpenAI	text-davinci-002	A Livraria do Saber é uma livraria espec
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.7	200	AI21 Studios	j2-mid	Livraria do Saber é uma nova livraria esq
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.8	200	AI21 Studios	j2-mid	Livraria do Saber é uma nova livraria esq
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.9	200	AI21 Studios	j2-mid	Livraria do Saber é uma nova livraria esq
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.10	200	AI21 Studios	j2-mid	Livraria do Saber é uma nova livraria esq
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.11	200	AI21 Studios	j2-mid	Livraria do Saber é uma nova livraria esq
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.5	2048	OpenAI	text-davinci-002	A Livraria do Saber é uma livraria espec
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.7	200	AI21 Studios	j2-mid	Livraria do Saber é uma nova livraria esq
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.7	200	AI21 Studios	j2-mid	Livraria do Saber é uma nova livraria esq
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.7	200	AI21 Studios	j2-mid	Livraria do Saber é uma nova livraria esq
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.5	2048	OpenAI	text-davinci-002	A Livraria do Saber é uma livraria espec
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.7	200	AI21 Studios	j2-mid	Livraria do Saber é uma nova livraria esq
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.7	200	AI21 Studios	j2-mid	Livraria do Saber é uma nova livraria esq
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.7	200	AI21 Studios	j2-mid	Livraria do Saber é uma nova livraria esq
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.7	200	AI21 Studios	j2-mid	Livraria do Saber é uma nova livraria esq
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.5	2048	OpenAI	text-davinci-002	A Livraria do Saber é uma livraria espec
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.7	200	AI21 Studios	j2-mid	Livraria do Saber é uma nova livraria esq
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.7	200	AI21 Studios	j2-mid	Livraria do Saber é uma nova livraria esq
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.7	200	AI21 Studios	j2-mid	Livraria do Saber é uma nova livraria esq
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.5	2048	OpenAI	text-davinci-002	A Livraria do Saber é uma livraria espec
"escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber"	96	0.7	200	AI21 Studios	j2-mid	Livraria do Saber é uma nova livraria esq



Resultados

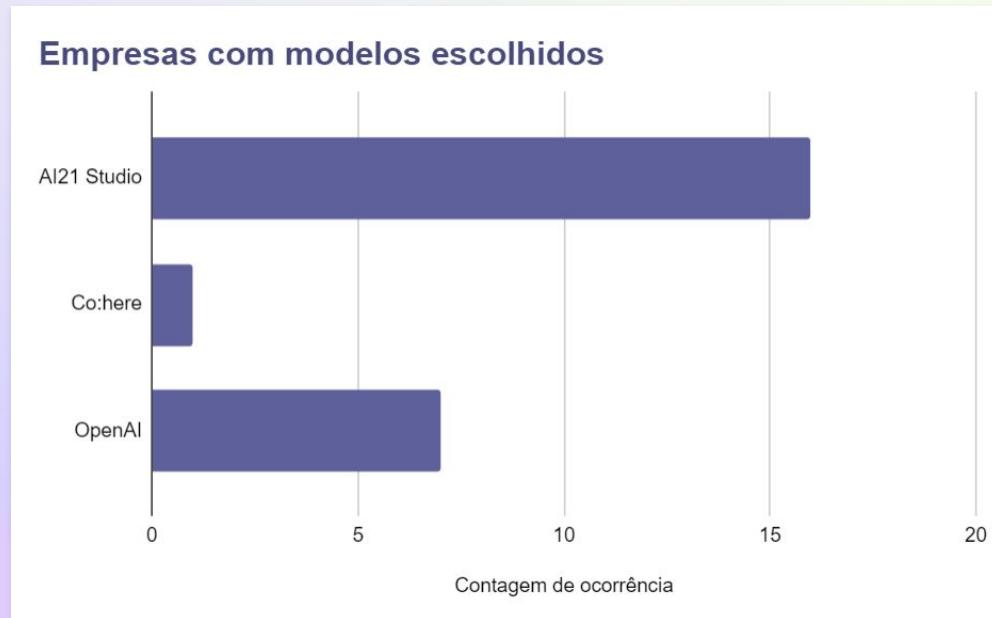
- para a instrução “escreva um slogan, em português, para uma livraria”, as relações das respostas escolhidas é a seguinte:





Resultados

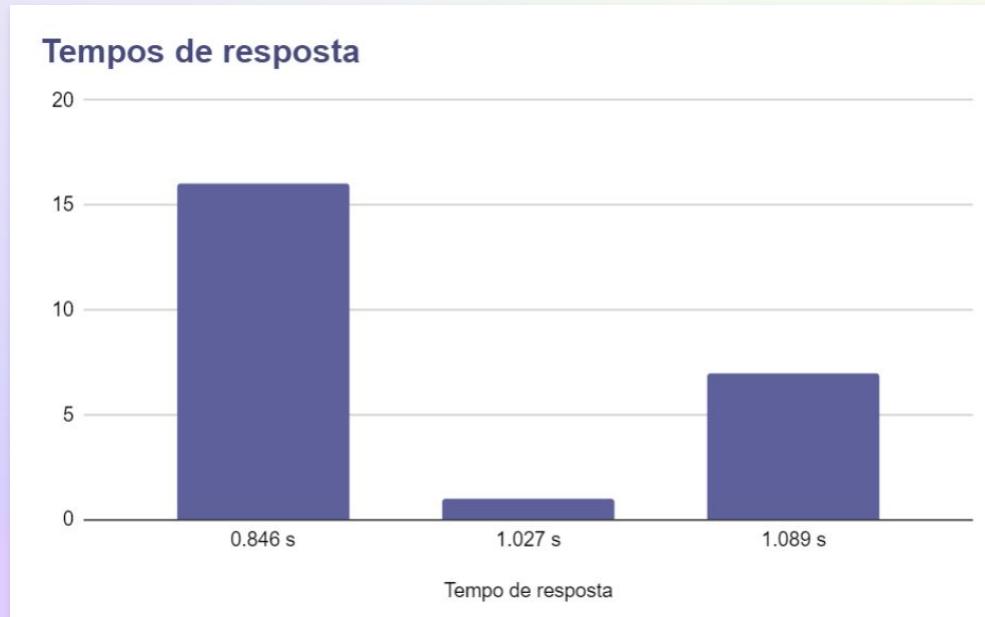
- para a instrução “escreva um slogan, em português, para uma livraria”, as relações das empresas escolhidas é a seguinte:





Resultados

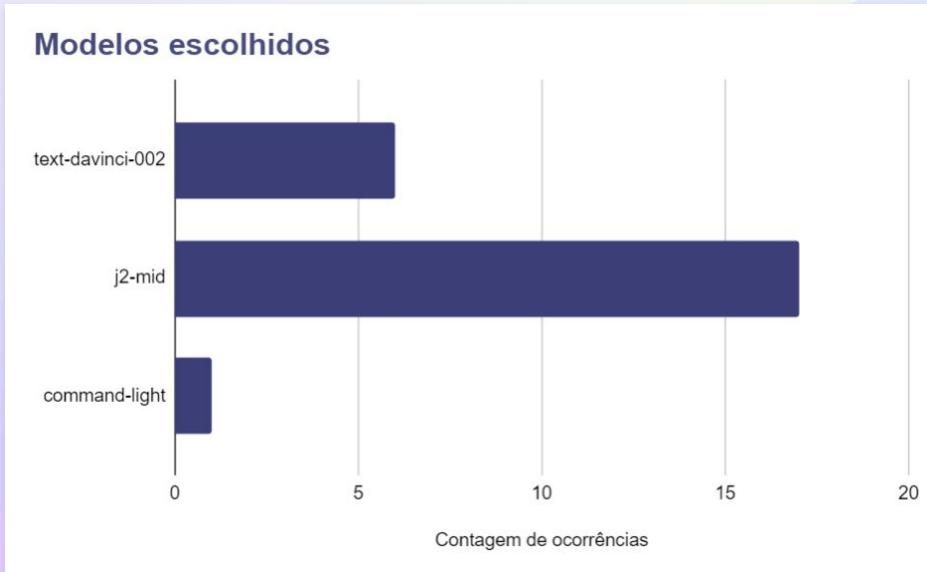
- para a instrução “escreva um slogan, em português, para uma livraria”, as relações dos tempos das respostas escolhidas é a seguinte:





Resultados

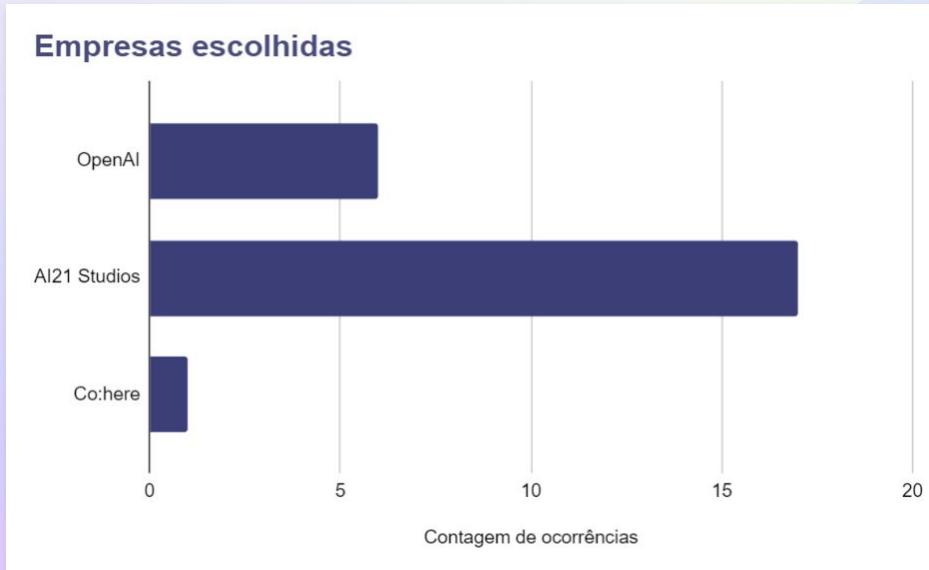
- para a instrução “escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber”, as relações das respostas escolhidas é a seguinte:





Resultados

- para a instrução “escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber”, as relações das empresas escolhidas é a seguinte:





Resultados

- para a instrução “escreva uma pequena descrição, em português, sobre uma nova livraria chamada Livraria do Saber”, as relações dos tempos das respostas escolhidas é a seguinte:





Resultados

- também é interessante analisar os tamanhos das respostas escolhidas:





05

Considerações Finais



Considerações Finais

- os LLMs abriram portas para aplicações antes não imagináveis
- o site desenvolvido busca ajudar nas escolhas de modelos para certas tarefas, através dos dados fornecidos pelos usuários
- a aplicação recebeu meios para enviar e receber informações, e armazená-las de maneira segura
- os dados salvos podem ser usados de várias maneiras
- o site pode ser usado tanto para consultar modelos, quanto para contribuir para um estudo maior da capacidade de cada um deles
- ambiente interativo e intuitivo



Trabalhos Futuros

- hospedar o site na internet e disponibilizar a aplicação para uso
- aumentar a quantidade de modelos utilizados
- aumentar a quantidade de empresas, criadoras de LLMs, escolhidas para fazer parte do projeto
- promover a coleta de dados em maior escala
- analisar os dados coletados e gerar um panorama maior dos desempenhos dos modelos



Referências

BAHETI, P. A newbie-friendly guide to transfer learning. 2021. Disponível em: <https://www.v7labs.com/blog/transfer-learning-guide>. Acesso em: 15 novembro 2023.

BALESTRIERO, R.; IBRAHIM, M.; SOBAL, V.; MORCOS, A.; SHEKHAR, S.; GOLDSTEIN, T.; BORDES, F.; BARDES, A.; MIALON, G.; TIAN, Y. et al. A cookbook of self-supervised learning. 2023. Acesso em: 15 novembro 2023.

BANANE, M.;ERRAISI, A. A comprehensive study of natural language processing techniques based on big data. 2022. Disponível em: <https://ieeexplore.ieee.org/abstract/document/9765270/>. Acesso em: 15 novembro 2023.

CARDON, A.; MÜLLER, D. N.; NAVAUX, P. Introdução às redes neurais artificiais. 1994. Disponível em: https://www.academia.edu/download/60901380/intro_rna20191014-14353-doa3qp.pdf. Acesso em: 15 novembro 2023.

JIANG, E.; OLSON, K.; TOH, E.; MOLINA, A.; DONSBACH, A.; TERRY, M.; CAI, C. J. Promptmaker: Prompt-based prototyping with large language models. 2022. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/3491101.3503564>. Acesso em: 15 novembro 2023.

TUNSTALL, L.; WERRA, L. V.; WOLF, T. Natural language processing with transformers. 2022. Disponível em: <https://books.google.com/books?hl=pt-BR&lr=&id=nzbEAAAQBAJ&oi=fnd&pg=PP1&dq=Natural+language+processing+with+transformers&ots=sTwmlWq3NB&sig=-1oAmBD6LroEyoWvl3OV9ii5UPY>. Acesso em: 15 novembro 2023.

OZDEMIR, S. Quick start guide to large language models: Strategies and best practices for using chatgpt and other llms. 2023. Acesso em: 15 novembro 2023.



Obrigada!

