# Desenvolvimento de um FAQ Interativo:
## Aprimorando a Experiência do Usuário

Ícaro Targino de Alexandria Rique

CENTRO DE INFORMÁTICA
UNIVERSIDADE FEDERAL DA PARAÍBA

João Pessoa - PB, 2024

Ícaro Targino de Alexandria Rique

# Desenvolvimento de um FAQ Interativo

Monografia apresentada ao curso Engenharia da Computação
do Centro de Informática, da Universidade Federal da Paraíba,
como requisito para a obtenção do grau de Bacharel.

Orientador: Yuri de Almeida Malheiros Barbosa

Maio de 2024

R594d Rique, Icaro Targino de Alexandria.
Desenvolvimento de um FAQ Interativo: Aprimorando a Experiência do Usuário / Icaro Targino de Alexandria Rique. – João Pessoa, 2024.
67 f.

Orientação: Yuri Barbosa.
TCC (Graduação) – UFPB/CI.
1. BERT. 2. Inteligência Artificial. 3. FAQ. 4. Deep
Learning. I. Barbosa, Yuri. II. Título.

UFPB/CI CDU004.8

CENTRO DE INFORMÁTICA
UNIVERSIDADE FEDERAL DA PARAÍBA

Trabalho de Conclusão de Curso de Engenharia da Computação contendo o título ***Desenvolvimento de um FAQ Interativo*** de autoria de Ícaro Targino de Alexandria Rique, aprovada pela banca examinadora constituída pelos seguintes professores:

---

Prof. Dr. Yuri de Almeida Malheiros Barbosa
Universidade Federal da Paraíba

---

Prof. Dr. Thaís Gaudencio Do Rêgo
Universidade Federal da Paraíba

---

Prof. Suanny Fabyne da Silva Vieira
Universidade Federal da Paraíba

---

Coordenador(a) do Departamento Centro de Informática
Verônica Maria Lima Silva
CI/UFPB

João Pessoa - PB, 15 de maio de 2024

Centro de Informática, Universidade Federal da Paraíba
Rua dos Escoteiros, Mangabeira VII, João Pessoa, Paraíba, Brasil CEP: 58058-600
Fone: +55 (83) 3216 7093 / Fax: +55 (83) 3216 7117

# DEDICATÓRIA

*Para a minha família, com a Graça de Deus.*

# AGRADECIMENTOS

Dedico esse trabalho aos elementos que me mantiveram de pé, são e consciente de minha missão no mundo:

- Ao Deus que me criou - *Soli Deo Gloria*;
- À minha esposa, cuja dedicação e amor me manteveram são e orientado em meio ao inesperado;
- Aos meus Pais e minha família, cuja paciência e a compreensão durante todos esses anos me fez confiar nos passos que dei.
- A minha avó, que desde sempre viu em mim um Engenheiro da Computação;
- Aos meus avós - *in memoriam* - que hoje não veem esse momento, todavia vivem em meu coração;
- Aos meus colegas de curso, que assim como eu encerram uma difícil etapa da vida;
- Aos meus orientadores, pela paciência
- Aos docentes que com dedicação exerceram o papel de orientar e ensinar durante 5 anos.
- À instituição que me acolheu

# RESUMO

A motivação para um sistema interativo de perguntas e respostas utilizando o *modelo* BERT *(Bidirectional Encoder Representations from Transformers)* surge da necessidade de lidar com as limitações das FAQs tradicionais, que são estáticos e volumosos, estáticos por não serem adequados à necessidade individual do usuário e volumosos porque tentam abordar uma ampla variedade de tópicos em um único documento, a fim de atender às necessidades de diversos usuários. O sistema proposto oferece respostas contextualizadas a um usuário do modelo, utilizando o contexto do FAQ para aprimorar a precisão e relevância das respostas fornecidas aos usuários. Os FAQs utilizados contém perguntas, respostas e contextos, há perguntas e respostas relacionadas ao contexo. Os resultados alcançados incluem a implementação de um chat interativo baseado nos contextos dos FAQs que alimentam o modelo, esse modelo é avaliado em sua capacidade de responder a perguntas feitas pelo usuário cujas respostas estão contidas nos contextos dos FAQs. Esse sistema oferece uma solução flexível e eficiente para acessar informações por meio de FAQs, proporcionando uma experiência aprimorada ao usuário.

**Palavras-chave: BERT, Inteligência Artificial, FAQ, Deep Learning.**

# ABSTRACT

The motivation for an interactive question-and-answer system utilizing the *BERT model (Bidirectional Encoder Representations from Transformers)* stems from the need to address the limitations of traditional FAQs, which are static and voluminous. They are static because they do not cater to the individual needs of the user and voluminous as they attempt to cover a wide variety of topics in a single document, aiming to meet the needs of various users.

The proposed system offers contextualized responses to a user of the model by leveraging the context of the FAQ to enhance the accuracy and relevance of the answers provided to users. The utilized FAQs contain questions, answers, and contexts, with questions and answers being related to the context. The results achieved include the implementation of an interactive chat based on the contexts of the FAQs that feed the model. This model is evaluated for its ability to respond to user-posed questions whose answers are contained within the FAQ contexts. This system provides a flexible and efficient solution for accessing information through FAQs, offering an enhanced user experience.



**Key-words: BERT, Artificial Intelligence, FAQ, Deep Learning.**

# LISTA DE FIGURAS


1 Representação esquemática de uma arquitetura de RNN. [10] . . . . . 18

2 Representação esquemática de uma arquitetura CNN [11] . . . . . 19

3 O Transformer - modelo arquitetural [6] . . . . . 21

4 Modelo de Pre-treinamento e *Fine-tuning*, respectivamente [1] . . . . . 22

5 Etapas de transformação dos dados brutos em um dataset FAQ modelo SQUAD . . . . . 32

6 Perdas durante o treino e Desempenho das métricas . . . . . 52

# LISTA DE TABELAS


1 Modelo de pergunta e resposta para BERT . . . . . 23
2 Tokenização do contexto . . . . . 24
3 Descrição dos trabalhos relacionados, seus objetivos e métodos . . . . . 30
4 Perguntas e respostas contidas no portal da OMS . . . . . 33
5 Exemplo de Pergunta e resposta original e tratada . . . . . 34
6 Modelos BERT disponíveis e suas camadas internas . . . . . 37
7 Configurações de hardware . . . . . 38
8 Configurações de Hiperparâmetros . . . . . 39
9 Resultado do *fine-tuning* do experimento 1 com duas épocas . . . . . 42
10 Amostra de diálogo do experimento 1 com duas épocas . . . . . 42
11 Resultado do *fine-tuning* do experimento 1 com três épocas . . . . . 43
12 Amostra de diálogo do experimento 1 com três épocas . . . . . 44
13 Resultado do *fine-tuning* do experimento 2 com duas épocas . . . . . 45
14 Amostra de diálogo do Experimento 2 com duas épocas . . . . . 46
15 Resultado do *fine-tuning* do Experimento 2 com três épocas . . . . . 47
16 Amostra de diálogo do experimento 2 com três épocas . . . . . 48
17 Números do treinamento do Experimento extra . . . . . 52
18 Amostra de diálogo do Experimento extra . . . . . 53
19 Pontuação de análise . . . . . 54
20 Desempenho subjetivo dos testes . . . . . 55
21 Amostra de diálogo do Experimento extra . . . . . 56
22 Amostra de diálogo do Experimento 1 com duas épocas em inglês . . . . . 62
23 Amostra de diálogo do Experimento 1 com três épocas em inglês . . . . . 63
24 Amostra de diálogo do Experimento 2 com duas épocas em inglês . . . . . 64
25 Amostra de diálogo do Experimento 1 com três épocas em inglês . . . . . 65
26 Amostra de diálogo do Experimento extra em inglês . . . . . 66
27 Amostra de diálogo do Experimento extra em inglês . . . . . 67

# LISTA DE ABREVIATURAS

ALBERT – *A Lite BERT* (Um BERT leve)

BERT – *Bidirectional Encoder Representations from Transformers*

CNN – *Convolutional Neural Network* (Redes Neurais Convolucionais)

COVID – *Coronavirus disease* (Doença do coronavirus)

ELMo – *Embeddings from Language Models*

FAQ – *Frequently Asked Questions* (Perguntas e Respostas frequentes)

GLUE – *General Language Understanding Evaluation*

GPT – *Generative Pre-trained Transformer*

HIV – *Human Immunodeficiency Virus* (Vírus da Imunodeficiência Humana)

IA – Inteligência Artificial

JSON – *JavaScript Object Notation*

OMS – Organização Mundial da Saúde

PLN – Processamento de Linguagem Natural

QA – *Question Answering* (Perguntas e Respostas)

RACE – *Reading Comprehension Dataset*

RNN – *Recurrent Neural Network* (Redes Neurais Recorrentes)

SARS-COV – *Severe Acute Respiratory Syndrome Coronavirus*

SQUAD – *Stanford Question Answering Dataset*

TA – Tradução automática

WHO – World Health Organization

## Sumário


**1 INTRODUÇÃO** **15**

1.1 Contextualização do Problema . . . . . 15

1.2 Premissas e Hipóteses . . . . . 15

1.2.1 Objetivo geral . . . . . 15

1.2.2 Objetivos específicos . . . . . 16

1.3 Estrutura da monografia . . . . . 16

**2 CONCEITOS GERAIS E REVISÃO DA LITERATURA** **17**

2.1 Aprendizagem Profunda . . . . . 17

2.2 Processamento de linguagem natural . . . . . 17

2.3 Arquiteturas de modelo de linguagem . . . . . 18

2.3.1 RNN . . . . . 18

2.3.2 CNN . . . . . 19

2.3.3 Transformers . . . . . 19

2.4 Funcionamento do BERT . . . . . 20

2.4.1 Pré-treino e *Fine-tuning* . . . . . 21

2.4.2 Dados de entrada do BERT . . . . . 22

2.5 BERT para Perguntas e Respostas . . . . . 23

2.5.1 Dados . . . . . 23

2.5.2 Tokenização . . . . . 24

2.5.3 SQuAD2.0 . . . . . 24

**3 TRABALHOS RELACIONADOS** **28**

**4 METODOLOGIA** **31**

4.1 Obtenção e estruturação do conjunto de dados . . . . . 31

4.1.1 Estruturação do conjunto de dados para uso . . . . . 32

4.1.2 Obtenção de dados FAQ da OMS . . . . . 33

4.1.3 Seleção do contexto . . . . . 33

4.1.4 Tratamento do contexto . . . . . 33
4.1.5 Definição das perguntas e respostas baseadas no contexto . . . . . 34
4.1.6 Aumento da base de dados . . . . . 35
4.1.7 Conversão das informações ao modelo SQUAD . . . . . 35
4.2 Escopo definido . . . . . 36
4.3 Identificação de trechos relevantes . . . . . 36
4.4 *Fine-tuning* específico . . . . . 36
4.5 Eficiência na modelagem de *tokens* . . . . . 36
4.6 Limitação de generalização . . . . . 37
4.7 Obtenção do modelo BERT pré-treinado . . . . . 37
4.7.1 **Fine-tuning do modelo** . . . . . 37
4.7.2 Pré-processamento de dados . . . . . 38
4.7.3 *Fine-tuning* . . . . . 38
4.7.4 Métricas . . . . . 39

**5 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS 41**
5.1 Modelo do Experimento 1 . . . . . 41
5.1.1 Duas épocas . . . . . 41
5.1.2 Três épocas . . . . . 43
5.2 Modelo do Experimento 2 . . . . . 43
5.2.1 Duas épocas . . . . . 45
5.2.2 Três épocas . . . . . 45
5.3 Análise dos resultados . . . . . 47

**6 CONCLUSÕES E TRABALHOS FUTUROS 49**

**REFERÊNCIAS 49**

**ANEXO A - ANEXOS E APÊNDICES 1 52**

**ANEXO B - Comparação dos resultados obtidos 54**

**ANEXO C - Diálogos extraidos do sítio da OMS** **57**

# 1 INTRODUÇÃO

## 1.1 Contextualização do Problema

Há uma crescente quantidade de informações disponíveis online, e esse aumento tem impulsionado o uso crescente de FAQs (Perguntas Frequentes), que são uma fonte valiosa de respostas para questões comuns apresentadas pelo público em geral. Tradicionalmente, esses FAQs consistem em perguntas e respostas predefinidas, que fornecem informações de maneira estática e geral. No entanto, essa abordagem tem suas limitações, especialmente quando se trata de adaptar-se a diferentes formas nas quais os usuários formulam suas perguntas.

Muitas vezes, os FAQs tradicionais não conseguem fornecer respostas contextualizadas e personalizadas aos usuários, o que pode resultar em uma experiência frustrante. Os usuários podem não encontrar as respostas adequadas às suas necessidades específicas ou podem não ter tempo para percorrer uma lista extensa de perguntas e respostas.

A utilização de chatbots e sistemas de IA para lidar com o suporte a usuários são soluções que buscam fornecer respostas mais dinâmicas, interativas e personalizadas, melhorando a experiência do usuário. Nosso trabalho se situa no contexto desses esforços, buscando aprimorar ainda mais a forma como as perguntas comuns são respondidas e como os usuários interagem com as informações disponíveis.

Nesse contexto, surge a necessidade de desenvolver soluções mais avançadas para melhorar a interação entre usuários e FAQs. Uma abordagem promissora é a utilização de modelos linguagem, como o BERT [1] (Bidirectional Encoder Representations from Transformers), que tem se mostrado altamente eficiente no entendimento de texto e na geração de respostas precisas.

Portanto, este trabalho se propõe a explorar o uso do modelo BERT para desenvolver um sistema interativo de perguntas e respostas que ofereça respostas contextualizadas e personalizadas aos usuários com base em um conjunto de dados de FAQs. Essa abordagem visa superar as limitações das FAQs tradicionais, fornecendo uma experiência mais satisfatória e eficiente na busca por informações. Ao permitir interações mais naturais e intuitivas, esse sistema tem o potencial de aprimorar significativamente a experiência do usuário ao buscar respostas para suas perguntas mais frequentes.

## 1.2 Premissas e Hipóteses

### 1.2.1 Objetivo geral

O objetivo geral deste trabalho é investigar e desenvolver um sistema para perguntas e respostas baseado no modelo BERT, para fornecer aos usuários respostas precisas e

contextuais para perguntas frequentes, aprimorando a experiência do usuário e a acessibilidade às informações relevantes.

### 1.2.2 Objetivos específicos

- Construção de uma base de dados a partir de um FAQ.
- Treinar um modelo de perguntas e respostas interativo usando um modelo de processamento natural de linguagem.
- Avaliar modelo de perguntas e respostas.

## 1.3 Estrutura da monografia

Esta monografia está subdividida em 6 capítulos:

1. Introdução, onde será introduzido o tema;
2. Justificativa, que trará a motivação por trás do capítulo;
3. Fundamentação teórica, neste tópico realizar-se-á uma revisão bibliográfica, abordando temas essenciais para uma compreensão ampla e aprofundada do assunto central da monografia;
4. Metodologia, para expor como o que foi feito para atingir os resultados obtidos;
5. Resultados, onde serão exposto e avaliado o trabalho desenvolvido e
6. Referências, onde serão encontradas todas as citações.

# 2 CONCEITOS GERAIS E REVISÃO DA LITERATURA

## 2.1 Aprendizagem Profunda

O *Deep Learning*, ou Aprendizagem Profunda, é uma área do campo de aprendizado de máquina que se concentra em utilizar redes neurais profundas para aprender e extrair representações complexas a partir de dados brutos, esse conceito é baseado nos sistemas neurais biológicos existentes na natureza [3]. Ao contrário das técnicas convencionais de aprendizado de máquina, que requerem a engenharia manual de características, o deep learning permite que os modelos aprendam diretamente dos dados sem a necessidade de etapas de pré-processamento intensivo.

Anteriormente, as técnicas de aprendizado de máquina dependiam desses recursos cuidadosamente projetados, que transformavam os dados brutos em uma representação mais adequada para que o sistema de aprendizado pudesse detectar padrões e tomar decisões. No entanto, com o deep learning, os modelos são capazes de aprender diretamente a partir dos dados brutos, sem a necessidade de uma etapa explícita de engenharia de recursos [4].

Essa capacidade de aprender automaticamente as representações dos dados é alcançada por meio de redes neurais profundas. Essas redes consistem em múltiplas camadas de unidades de processamento, chamadas neurônios, que transformam a entrada em representações progressivamente mais abstratas e de alto nível. Cada camada de neurônios recebe informações da camada anterior, permitindo que o modelo aprenda gradualmente características complexas e hierárquicas dos dados.

O *deep learning* tem sido amplamente empregado em diversas tarefas, destacando-se especialmente no campo do Processamento de Linguagem Natural (PLN). Especificamente para classificação de texto, modelos baseados em deep learning têm demonstrado capacidade de compreender e gerar texto [5], traduzir idiomas [6], analisar sentimentos [1] e responder perguntas [7], tudo isso aproveitando vastos volumes de dados textuais.

## 2.2 Processamento de linguagem natural

A pesquisa em Processamento de Linguagem Natural teve início na década de 1940, sendo utilizado pioneiramente para Tradução Automática (TA). O projeto pioneiro de TA foi iniciado por Weaver e Booth em 1946, baseado em sua experiência em decifrar códigos inimigos durante a Segunda Guerra Mundial [8].

O campo de estudos PLN é uma área multidisciplinar que envolve ciência da computação, inteligência artificial e linguística computacional. Ela visa capacitar os computadores a interpretar e interagir de forma efetiva com as línguas humanas, sejam elas

escritas ou faladas. O PLN desempenha um papel fundamental na construção de sistemas computacionais capazes de lidar com a linguagem humana, preenchendo a lacuna entre a comunicação humana e as máquinas [9].

## 2.3 Arquiteturas de modelo de linguagem

As arquiteturas de modelos de linguagem desempenham um papel fundamental no processamento de linguagem natural. Dentre as diversas abordagens utilizadas, destacam-se as redes neurais recorrentes (do inglês, *Recurrent Neural Network* - RNNs)[10], redes neurais convolucionais (do inglês, *Convolutional Neural Network* - CNNs) [11] e transformers [12].

### 2.3.1 RNN

As RNNs são arquiteturas de redes neurais que possuem uma estrutura de conexões que permitem a retroalimentação das informações anteriores para influenciar o processamento das informações futuras como vemos na Figura 1, sendo essa a característica principal das RNNs, a capacidade de criar feedbacks de informações passadas para influenciar o processamento de dados futuros [10]. Esse recurso é de grande valia, pois permite que o modelo considere o contexto sequencial nas análises. Essa capacidade de levar em consideração o contexto sequencial é especialmente útil em tarefas como tradução automática e geração de texto. No entanto, as RNNs podem enfrentar dificuldades ao lidar com sequências muito longas devido ao problema da atenuação ou explosão do gradiente, que afeta a capacidade de aprendizado do modelo.

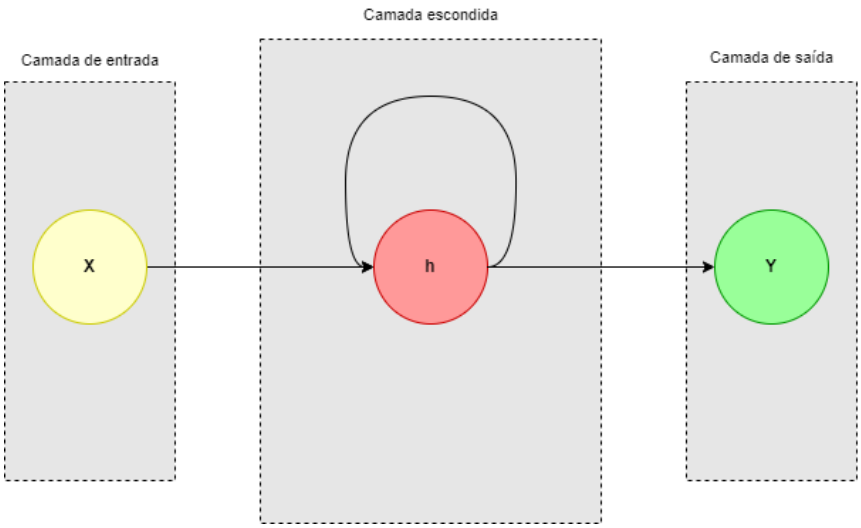


**Figura 1: Representação esquemática de uma arquitetura de RNN. [10]**

### 2.3.2 CNN

Por outro lado, as CNNs têm sido amplamente utilizadas em tarefas de processamento de linguagem natural, especialmente em tarefas de classificação de texto. As CNNs são capazes de extrair automaticamente características hierárquicas dos dados, graças à sua estrutura composta de múltiplas camadas convolucionais. Estes atributos extraídos tornam-se mais abstratos à medida que se movem através das camadas, permitindo que a rede identifique padrões complexos [11]. A arquitetura básica de uma CNN (Figura 2), envolve camadas convolucionais que detectam padrões e camadas de subamostragem que reduzem a dimensionalidade. No entanto, as CNNs geralmente não levam em consideração a dependência de longo prazo entre as palavras em um texto.

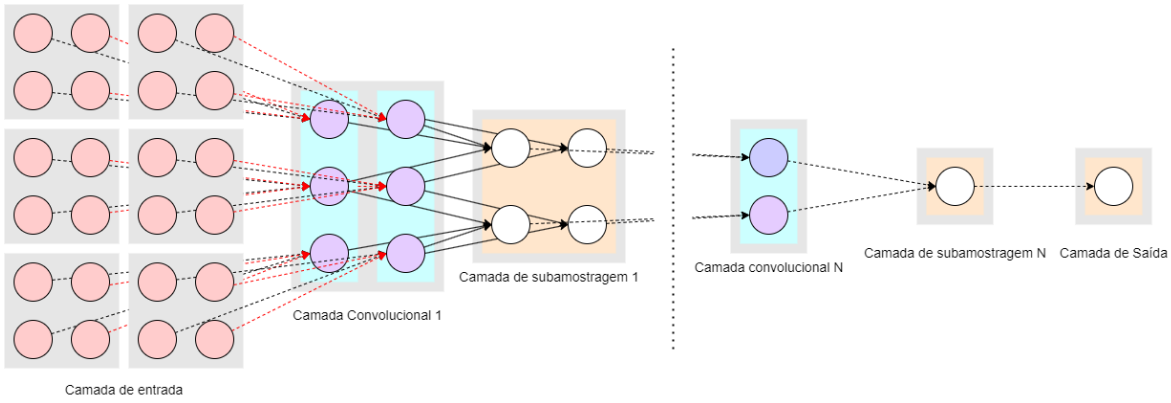


**Figura 2: Representação esquemática de uma arquitetura CNN [11]**

### 2.3.3 Transformers

A arquitetura dos Transformers é a mais recente e altamente influente. Os Transformers utilizam mecanismos de atenção para calcular a importância de cada palavra em relação a todas as outras palavras do texto, também faz uso efetivo da codificação posicional, garantindo que informações sobre a posição das palavras na sequência sejam preservadas. Uma arquitetura que implementa camadas empilhadas de autoatenção e camadas totalmente conectadas para o codificador e decodificador [6], conforme mostra a Figura 3. Tanto o codificador quanto o decodificador são compostos por múltiplas camadas de autoatenção e camadas totalmente conectadas.

A autoatenção permite que o modelo capture as relações de dependência entre as palavras em uma sequência de entrada, atribuindo pesos diferentes a diferentes palavras com base em sua relevância para a tarefa em questão. Isso melhora significativamente a capacidade do modelo de lidar com sequências longas e capturar relações contextuais complexas. [6].

Após a camada de autoatenção, cada subcamada possui uma rede neural totalmente conectada para processar as informações, a rede neural totalmente conectada é responsável por processar as informações de forma mais complexa e realizar transformações não lineares nos dados, as etapas a seguir são fases pós camada de autoatenção.

1. Processamento não-linear: Permite que o modelo aprenda representações mais abstratas e sofisticadas das informações;

2. Aprendizado de representações hierárquicas: Aprende representações hierárquicas dos dados;

3. Combinação de informações: A camada feed-forward permite combinar informações obtidas na camada

Uma das principais vantagens do Transformer é a sua capacidade de aprendizado global, pois o modelo tem acesso direto a todas as palavras da sequência de entrada durante o processo de decodificação. Isso permite que o Transformer resolva problemas de tradução, resumo de texto, geração de texto e várias outras tarefas de processamento de linguagem natural.

Além disso, o Transformer utiliza camadas que processam as informações de cada posição em paralelo, em vez de sequencialmente, como em modelos recorrentes. Isso acelera o treinamento e permite que o modelo capture melhor as relações entre todas as posições na sequência [6]. É altamente eficaz no treinamento com grandes corpus, resultando em melhorias significativas na precisão em várias tarefas de processamento de linguagem natural. Essas tarefas incluem classificação de texto, compreensão de linguagem, tradução automática, resolução de referências, inferência de senso comum e sumarização, entre outras [13].

Um dos modelos mais notáveis disponíveis na biblioteca Transformers é o BERT, que utiliza a arquitetura transformer. Sua introdução marcou um ponto de virada no processamento de linguagem natural, transformando significativamente a maneira como abordamos e compreendemos tarefas nesse campo.

## 2.4 Funcionamento do BERT

O BERT (do Inglês, *Bidirectional Encoder Representations from Transformers*) é um modelo de linguagem pré-treinado desenvolvido pelo Google, foi projetado para ser submetido a um pré-treinamento com representações bidirecionais profundas, a partir de texto não rotulado, levando em consideração o contexto, tanto à esquerda, quanto à direita.[1] Essas representações contextuais das palavras permitem que o BERT seja

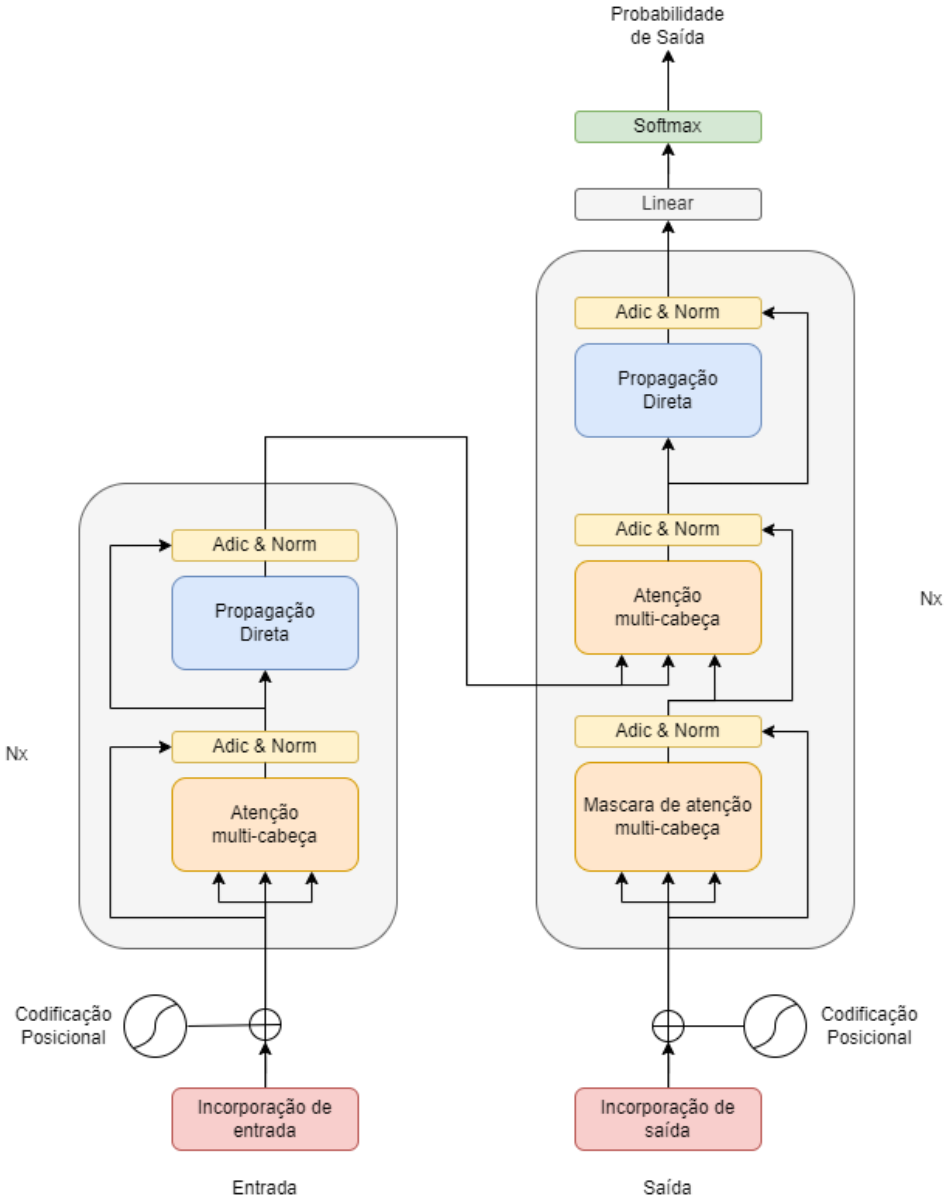


**Figura 3: O Transformer - modelo arquitetural [6]**

ajustado com uma camada de saída adicional, gerando modelos de última geração para várias tarefas, como resposta a perguntas e inferência de linguagem. O BERT é simples em conceito, mas poderoso em termos de desempenho, obtendo resultados de ponta em diversas tarefas de processamento de linguagem natural.

### 2.4.1 Pré-treino e *Fine-tuning*

Como primeira etapa, o BERT passa por um estágio de pré-treinamento em grandes quantidades de texto não rotulado, conforme a parte esquerda da Figura 4. Durante esse processo, o modelo aprende a prever palavras ocultas em um contexto. Essa etapa de pré-treinamento permite que o BERT adquira conhecimento geral sobre a linguagem,

capturando relações semânticas e sintáticas em um nível profundo [1].

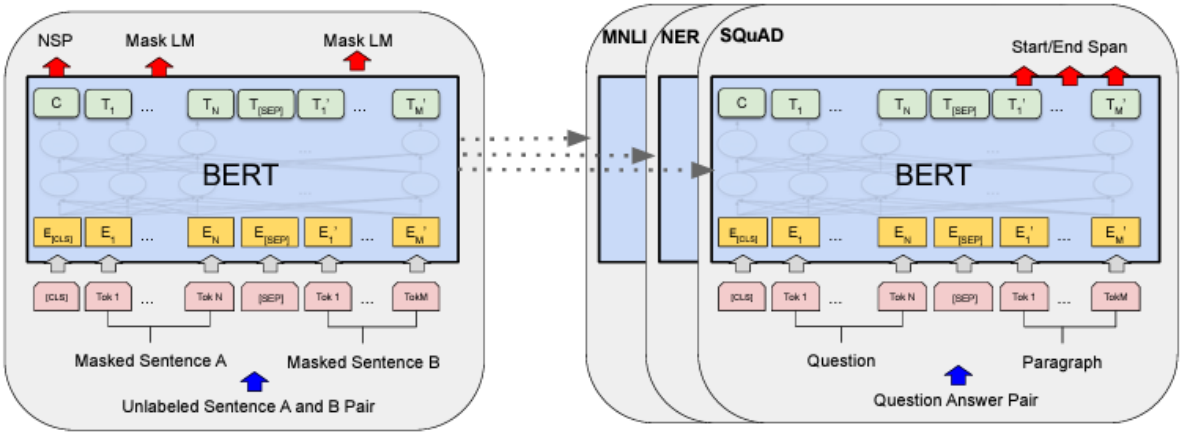


**Figura 4: Modelo de Pre-treinamento e *Fine-tuning*, respectivamente [1]**

Após o pré-treinamento, o BERT pode ser ajustado para tarefas específicas (Figura 4), como classificação de texto, resposta a perguntas, tradução automática, entre outras. O ajuste consiste em adicionar uma camada de saída adicional ao modelo pré-treinado e treiná-lo em conjuntos de dados rotulados para a tarefa desejada.

### 2.4.2 Dados de entrada do BERT

O modelo BERT recebe como entrada uma sequência de *tokens*, que são instâncias de uma peça de texto, como palavras ou partes de palavras, e representam as unidades básicas de processamento no modelo. Os *tokens* pode representar tanto uma única sentença, quanto um par de sentenças, dependendo da tarefa em questão.

No caso de uma única sentença, os tokens podem corresponder a palavras individuais ou partes de palavras, como subpalavras. A sequência de *tokens* pode incluir marcadores especiais, como o token [CLS], no início da sequência para indicar o início da sentença e o *token* [SEP], entre as sentenças para separá-las [1].

No caso de um par de sentenças, os *tokens* são concatenados em uma única sequência, com o *token* [SEP] separando as duas sentenças. Por exemplo, em uma tarefa de resposta a perguntas, a sequência de entrada pode consistir na pergunta seguida da resposta, separadas pelo *token* [SEP].

Além dos *tokens* de entrada, o BERT também utiliza máscaras de atenção para indicar quais *tokens* devem ser considerados durante o treinamento e inferência. Isso é necessário porque o BERT realiza um pré-treinamento autossupervisionado, onde parte dos *tokens* de entrada são mascarados e o modelo é treinado para prever esses *tokens* ocultos, com base no contexto fornecido pelos *tokens* vizinhos.

## 2.5 BERT para Perguntas e Respostas

O modelo BERT para perguntas e respostas é uma adaptação do modelo de linguagem BERT que visa executar a tarefa específica de responder perguntas com base em um contexto fornecido. Essa adaptação é realizada por meio do pré-treinamento e o *fine-tuning* do modelo BERT.

Na etapa de pré-treinamento, o modelo BERT é inicialmente treinado em uma ampla quantidade de dados textuais não rotulados, de forma a aprender representações linguísticas ricas e contextuais. Essa fase é fundamental para permitir ao modelo BERT desenvolver uma compreensão abrangente das relações semânticas e sintáticas presentes nas sentenças.

Posteriormente, na etapa de fine-tuning, o modelo pré-treinado é ajustado em um conjunto menor de dados específicos para a tarefa de perguntas e respostas. Nessa fase, são utilizados pequenos datasets contendo pares de perguntas e respostas para afinar os parâmetros do BERT e adaptá-lo ao contexto da tarefa específica.

### 2.5.1 Dados

No contexto de perguntas e respostas, o BERT é treinado para entender a relação entre perguntas e respostas em um dado contexto. Ele recebe um par de pergunta e contexto e busca identificar a resposta correta dentro desse contexto. Ou seja, para o modelo BERT ser efetivo no objetivo de responder perguntas, o modelo precisa receber uma pergunta, uma resposta relacionada à pergunta e o contexto de onde essa resposta foi retirada, conforme vemos na Tabela 1.

**Tabela 1: Modelo de pergunta e resposta para BERT**

| **Descrição** | **Texto** |
|---|---|
| Pergunta | "Quais são os benefícios do exercício físico para a saúde mental?" |
| Contexto | "Os benefícios do exercício físico para a saúde mental incluem a redução dos sintomas de ansiedade e depressão, melhora do humor, aumento do bem-estar psicológico e uma melhor qualidade de vida. Além disso, o exercício físico pode ser usado como uma estratégia complementar no tratamento de transtornos mentais, como a depressão clínica e o transtorno de ansiedade generalizada." |
| Resposta | "redução dos sintomas de ansiedade e depressão, melhora do humor, aumento do bem-estar psicológico e uma melhor qualidade de vida" |

### 2.5.2 Tokenização

A tokenização no BERT envolve dividir o texto em unidades menores chamadas *tokens*. Os tokens podem ser palavras individuais, partes de palavras (subpalavras), ou até mesmo caracteres. Cada *token* possui um valor numérico associado, chamado de índice de *token*, que é utilizado como entrada para o modelo BERT, podemos ver um exemplo de tokenização na Tabela 2.

Um aspecto importante da "tokenização" no BERT é a adição de *tokens* especiais no início e no final do texto. O *token* especial [CLS] (início da sequência) é adicionado no início do texto, e o *token* especial [SEP] (separador de sequência) é adicionado no final do texto. Esses *tokens* especiais auxiliam o modelo BERT a entender a estrutura e a relação entre diferentes partes do texto. Cada palavra é convertida em um ou mais *tokens*. Por exemplo, a palavra tokenização pode ser dividida nos *tokens* "token", "##iza", "##ção". Os caracteres "##"antes de um *subtoken* indica que ele é uma continuação ou uma subdivisão de uma palavra maior.

**Tabela 2: Tokenização do contexto**

| **Descrição** | **Texto** |
|---|---|
| Contexto | "O exercício físico regular tem sido associado a uma série de benefícios para a saúde mental. Estudos mostram que a prática regular de atividade física pode reduzir os sintomas de ansiedade e depressão, melhorar o humor, aumentar a sensação de bem-estar e promover uma melhor qualidade de vida psicológica. Além disso, o exercício físico tem sido sugerido como uma estratégia complementar no tratamento de transtornos mentais, como a depressão clínica e o transtorno de ansiedade generalizada." |
| *Tokens* | "[CLS] O exer ##cício fís ##ico reg ##ular tem sido as ##soc ##iado a uma série de bene ##fícios para a saúde mental . [SEP] Est ##udos mostr ##am que a prá ##tica reg ##ular de ati ##vid ##ade fís ##ica pode red ##uz ##ir os sint ##omas de ans ##ied ##ade e de ##pre ##ssão , melh ##orar o humor , aument ##ar a sens ##ação de bem - estar e prom ##over uma melh ##or qual ##idade de vida psic ##ológ ##ica . [SEP] Além dis ##so , o exer ##cício fís ##ico tem sido sug ##er ##ido como uma estrat ##égia comp ##lement ##ar no trat ##amento de transt ##orn ##os ment ##ais , como a de ##pre ##ssão clín ##ica e o transt ##orno de ans ##ied ##ade gen ##eral ##izada . [SEP]" |

### 2.5.3 SQuAD2.0

SQuAD 2.0 (do Inglês, Stanford Question Answering Dataset 2.0) é um conjunto de dados em formato JSON (Código 1) com finalidade de compreensão de leitura composto

por 100.000 perguntas e seus contextos correspondentes baseado em artigos existentes na Wikipedia, em que as respostas podem ser diretamente geradas a partir do trecho de texto. O SQuAD 2.0 foi desenvolvido pela Universidade de Stanford e é usado como referência para medir o desempenho dos modelos de processamento de linguagem natural em tarefas de leitura de máquina [14]. O SQuAD é construído com base em trechos reais de texto, tornando-o mais representativo de dados do mundo real. Isso permite que o BERT seja treinado e avaliado em cenários mais próximos da aplicação prática, melhorando sua capacidade de compreender e responder perguntas sobre textos reais.

O BERT passa por um processo de pré-treinamento utilizando o conjunto de dados SQuAD2.0. Durante esse processo, o BERT é exposto a uma vasta quantidade de perguntas e seus respectivos contextos, com o objetivo de aprender a identificar e compreender o trecho relevante no contexto que contém a resposta para cada pergunta [7].

O formato JSON do SQuAD é uma estrutura usada para representar os dados de treinamento e teste do conjunto de dados. O formato é projetado de forma a organizar as perguntas, os respectivos contextos e as respostas corretas. Esse padrão de organização facilita o acesso e o processamento dos dados durante a fase de treinamento de modelos de processamento de linguagem natural, como o BERT.

**Código 1: Exemplo de JSON formato SQuAD**

```
{
    "version": "v2.0",
    "data": [
        {
            "paragraphs": [
                {
                    "qas": [
                        {
                            "question": "What is COVID-19?",
                            "id": 1026331,
                            "answers": [
                                {
                                    "answer_id": 907396,
                                    "document_id": 1520732,
                                    "question_id": 1026331,
                                    "text": "COVID-19 is the disease caused
                                        by a coronavirus called SARS-CoV-2.
                                    ",
                                    "answer_start": 0,
                                    "answer_end": 66,
                                    "answer_category": null
                                }
                            ],
                            "is_impossible": false
```

```
                        }
                    ],
                    "context": "COVID-19 is the disease caused by a
                        coronavirus called SARS-CoV-2. WHO first learned of
                         this new virus on 31 December 2019, following a
                        report of a cluster of cases of so-called viral
                        pneumonia in Wuhan, Peoples Republic of China.",
                    "document_id": 1520732
                }
            ]
        }
    ]
}
```

Cada um dos dados no JSON representa uma informação específica sobre o conjunto de dados SQuAD. Aqui está uma breve descrição de cada um deles:

1. **“version”**: Indica a versão do conjunto de dados do SQuAD. No exemplo, a versão é ”2.0”.
2. **“data”**: É uma lista que contém os dados do conjunto de dados SQuAD.
3. **“title”**: Representa o título do artigo. No exemplo, o título é “Exemplo de Artigo”.
4. **“paragraphs”**: É uma lista que contém os parágrafos do artigo. Cada parágrafo contém informações sobre o contexto, as perguntas (qas) relacionadas a esse contexto e as respostas correspondentes a essas perguntas
5. **“context”**: É o contexto do parágrafo. No exemplo, o contexto é “Este é um contexto de exemplo para o conjunto de dados SQuAD”.
6. **“qas”**: É uma lista que contém as perguntas e respostas relacionadas ao contexto.
7. **“question”**: É a pergunta relacionada ao contexto. No exemplo, a pergunta é “O que é o conjunto de dados SQuAD?”.
8. **“id”**: É um identificador único atribuído a cada pergunta. No exemplo, o ID da pergunta é “1”.
9. **“answers”**: É uma lista que contém as respostas para a pergunta.
10. **“text”**: É o texto da resposta. No exemplo, a resposta é “O conjunto de dados SQuAD é um conjunto de dados de compreensão de leitura”.
11. **“is_impossible”**: Quando o valor é verdadeiro, significa que a pergunta não pode ser respondida com base no contexto fornecido.

12. **“answer_start”**: Indica a posição de início da resposta no contexto. No exemplo, a resposta começa na posição 26 do contexto.

Todos os dados enumerados estão disponíveis no conjunto de dados do SQuAD2.0, todavia os dados de entrada para o modelo BERT seriam o texto do contexto e a pergunta relacionada a esse contexto, que é calculada a partir da posição de início da resposta.

# 3 TRABALHOS RELACIONADOS

Nesta seção, buscamos contextualizar e examinar o estado atual da pesquisa quanto ao uso do modelo BERT objetivando a construção de *chatbots* que respondem dentro de contextos delimitados, uma abordagem de aprendizagem de máquina, especificamente aplicado ao domínio de perguntas e respostas. O foco está em explorar a aplicabilidade e os impactos deste modelo quando operado em contextos restritos. Essa limitação tem como objetivo concentrar a variedade das respostas possíveis dentro dos limites estabelecidos pelo contexto fornecido.

Os estudos selecionados, apresentados na tabela 3, possuem uma relação direta com o tema deste trabalho, cujo objetivo é realizar um *fine-tuning* em um modelo BERT pré-treinado utilizando o conjunto de dados do SQUAD. Este ajuste tem como finalidade proporcionar aos usuários de portais web uma maneira dinâmica e objetiva de consultar conteúdos que comumente são encontratos em FAQs por meio da interação com o modelo. Esses estudos exploram aspectos cruciais do treinamento e *fine-tuning*, incluindo hiperparâmetros, definições e desenvolvimento, e sugerem métodos e abordagens para aplicar o conhecimento neles contido.

A análise crítica desses estudos é de grande relevância para este trabalho, pois ela fundamenta as discussões e contribui para o avanço do conhecimento na área, especialmente no que se refere a chatbots aplicados para contextos específicos e limitados.

Em seu estudo, Devlin [1] apresenta uma técnica de pré-treinamento para processamento de linguagem natural chamada de BERT. O BERT é projetado para aprender representações profundas bidirecionais a partir de texto não rotulado. Em seu estudo demonstra-se que o BERT alcança resultados em uma ampla gama de tarefas de processamento de linguagem natural e sobre o processo de ajuste fino e pré-treinamento.

Após o pré-treinamento, o BERT é avaliado em uma variedade de tarefas de processamento de linguagem natural, uma delas sendo a capacidade de responder perguntas. O desempenho do BERT é comparado com outros modelos existentes para demonstrar sua eficácia e superioridade em várias tarefas. O estudo também investiga a capacidade do BERT de sofrer ajuste fino para tarefas específicas sem a necessidade de modificações significativas na arquitetura do modelo.

O estudo em questão demonstrou que o modelo BERT alcançou resultados tarefas de processamento de linguagem natural, superando modelos anteriores como o OpenAI GPT e o ELMo. Além disso, foram realizadas análises de etapas de pré-treinamentos, bem como ajustes finos, onde avaliou-se que o BERT é sensível a essas etapas. Os resultados detalhados foram apresentados, incluindo métricas como precisão, recall e F1 para cada tarefa avaliada.

Com base nos estudos desenvolvidos por Devlin, Liu [15] realiza uma revisão do modelo e propõem a replicação do pré-treinamento do BERT, com foco na avaliação minuciosa dos principais hiperparâmetros e do volume de dados de treinamento. O propósito central é investigar se o BERT foi subtreinado e se é possível igualar ou superar o desempenho de todos os modelos. Para tanto, serão analisados os resultados obtidos em tarefas relevantes como GLUE, RACE e SQuAD.

O objetivo principal é demonstrar a capacidade do BERT, considerando uma otimização robusta dos hiperparâmetros e manipulando-os juntamente com o volume de dados de treinamento. Com os propósitos para o estudo definidos, o autor introduz o RoBERTa, uma abordagem aprimorada do BERT, e realiza comparações com outras técnicas de pré-treinamento de modelos de linguagem, como ELMo e GPT.

Ainda de acordo com Liu [15], o modelo BERT proposto por Devlin foi identificado como subtreinado, impulsionando a introdução do RoBERTa, que demonstrou a capacidade de igualar ou superar o desempenho do BERT de acordo com as métricas avaliadas, a acurácia, a pontuação média do RoBERTa em tarefas GLUE, RACE e SQuAD e o F1-score. Quanto aos comparativos com GLUE, RACE e SQuAD, O RoBERTa obteve melhores resultados sem a necessidade de ajustes multitarefa ou de dados adicionais.

Da mesma maneira que Liu, Lan [16] apresenta uma revisão do trabalho do Devlin com uma abordagem mais eficiente e rápida para o pré-treinamento de modelos de linguagem, baseados no modelo BERT.

O objetivo principal é reduzir o consumo de memória e aumentar a velocidade de treinamento, mantendo ou melhorando o desempenho em tarefas de compreensão de linguagem e propor uma arquitetura mais eficiente e econômica do que o modelo BERT tradicional.

A metodologia para atingir esse resultado passa por induzir técnicas de redução de parâmetros, como a parametrização de embedding factorizada e o compartilhamento de parâmetros entre camadas, visando reduzir o número de parâmetros sem comprometer significativamente o desempenho. Além disso, o trabalho propõe uma nova perda autosupervisionada para previsão da ordem de sentenças.

As métricas para análise do desempenho do modelo ALBERT em diferentes tarefas de processamento de linguagem natural foram a utilização de Acurácia e F1-Score, além de análises de desempenho do projeto com GLUE, SQuAD e RACE.

O presente estudo busca aprimorar a interação entre usuários e sistemas por meio do uso de um modelo BERT, que foi pré-treinado e posteriormente ajustado para fornecer respostas contextualizadas e personalizadas. Utilizando um conjunto de dados de perguntas frequentes (FAQs), o sistema é projetado para identificar e responder precisamente às dúvidas específicas dos usuários. Dessa forma, busca-se eliminar a necessidade de os

usuários filtrarem entre várias respostas genéricas.

**Tabela 3: Descrição dos trabalhos relacionados, seus objetivos e métodos**

| **Trabalho** | **Método** | **Objetivo** |
|---|---|---|
| BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding | Desenvolvimento, treinamento e avaliação do modelo BERT em tarefas de PLN | Apresentar o modelo BERT como uma técnica eficaz para PLN |
| RoBERTa: A Robustly Optimized BERT Pretraining Approach | Replicação do pré-treinamento do BERT por um período mais longo, com lotes maiores e sobre mais dados e sequências mais longas | Propor um modelo BERT melhorado a partir de modificações em parâmetros |
| ALBERT: A Lite BERT for Self-supervised Learning of Language Representations | Abordar os problemas de limitação de memória e sobrecarga de comunicação associados a modelos de grande escala, como o BERT. | Propor uma arquitetura mais eficiente e econômica do que o modelo BERT tradicional |
| Este Trabalho | Abordar os problemas de um chatbot contextual, limitação de dados para fine-tuning e modificação de parâmetros | Fornecer respostas contextualizadas e personalizadas aos usuários com base em dados de FAQs através de um modelo BERT. |
| Fonte: Elaborada pelo autor | | |

# 4 METODOLOGIA

Esse trabalho será dividido nas seguintes etapas:

1. Obtenção e estruturação do dataset
2. Adequação do conjunto de dados;
3. Aplicação de técnicas;
4. Escolha do modelo mais adequado;
5. Obtenção do modelo BERT pré-treinado;
6. *Fine-tuning* do modelo e
7. Teste e avaliação dos resultados.

## 4.1 Obtenção e estruturação do conjunto de dados

A escolha do tema COVID-19 para montagem do conjunto de dados foi motivada pela relevância e atualidade do assunto, todavia pode ser aplicado para qualquer temática ou área do conhecimento que sucite dúvidas aos leitores tanto que seja necessário um FAQ para esclarecimento dessas dúvidas. O COVID-19 é uma doença que afetou o mundo inteiro, gerando muitas dúvidas e incertezas nas pessoas, o que inclui bastante desinformação sobre o tema [17]. A proliferação de teorias da conspiração, informações imprecisas e notícias falsas dificultaram ainda mais a compreensão correta da doença e das medidas preventivas necessárias [18], dessa forma, a criação de um dataset que aborde especificamente perguntas e respostas sobre o COVID-19 busca fornecer um recurso confiável e embasado para ajudar a combater a desinformação e oferecer informações corretas aos usuários. A Organização Mundial da Saúde (OMS) é uma fonte confiável de informações sobre a assuntos médicos relacionados à pandemia, fornecendo um FAQ detalhado em seu site.

O objetivo foi criar um conjunto de dados específico para o modelo de Perguntas e Respostas (do inglês, Question Answering - QA) utilizando o formato do SQUAD. Para isso, as respostas contidas no FAQ da OMS foram capturadas e transformadas em um formato JSON, que segue a estrutura do SQUAD, citado anteriormente. O sitio da OMS originalmente está escrito na língua inglesa e portanto a pesquisa foi totalmente desenvolvida em Inglês.

Devido à natureza limitada do nosso conjunto de dados original, que continha apenas 16 pares de perguntas e respostas destinados a FAQs comuns, foi imperativo

aplicar uma abordagem de data augmentation [19] que serão citadas mais a frente. Esta estratégia teve como objetivo aprimorar a diversidade das perguntas e respostas, por meio da incorporação de mudanças de palavras, reformulações de perguntas e seleção criteriosa das respostas no texto. Essa abordagem permitiu aumentar a quantidade de exemplos no conjunto de dados e fornecer mais variedade aos dados de treinamento do modelo[20].

Ao utilizar o dataset expandido e diversificado, o modelo de QA, baseado em BERT, pode aprender a associar diferentes perguntas com a mesma resposta correta. Com isso, espera-se amplia a capacidade do modelo de responder de maneira mais abrangente a várias perguntas relacionadas ao COVID-19, mesmo que formuladas de maneiras diferentes.

### 4.1.1 Estruturação do conjunto de dados para uso

O processo de preparação do conjunto de dados para se adequar ao formato do SQUAD envolve várias etapas importantes conforme Figura 5.

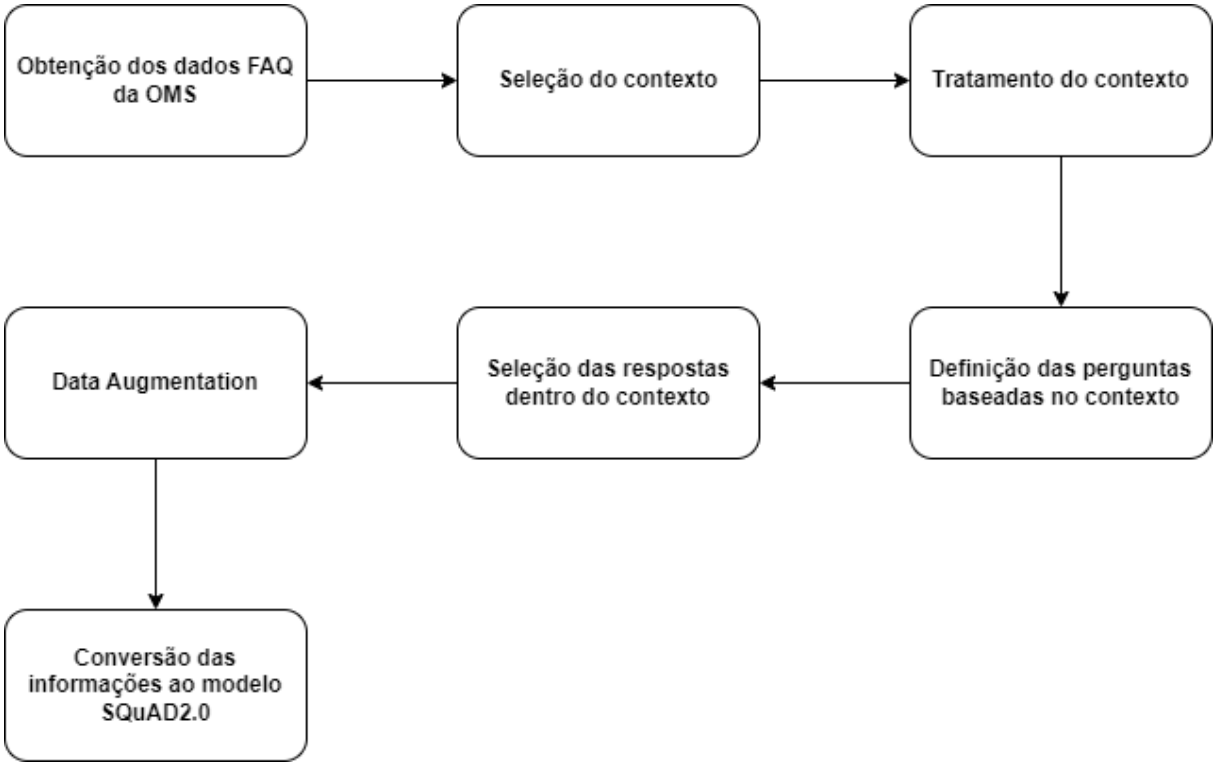


**Figura 5: Etapas de transformação dos dados brutos em um dataset FAQ modelo SQUAD**

### 4.1.2 Obtenção de dados FAQ da OMS

O processo de aquisição de dados envolve a obtenção das 16 perguntas e respostas diretamente do FAQ da OMS [1] sobre COVID-19. Isso pode ser feito acessando o FAQ por meio do site oficial da OMS, como mostrado na estrutura da pergunta na tabela 4. Os dados brutos são extraídos do FAQ sem nenhum processamento adicional. Em seguida, os dados são armazenados em arquivos de texto separados.

**Tabela 4: Perguntas e respostas contidas no portal da OMS**

| **Pergunta** | **Resposta** |
|---|---|
| O que é COVID-19? | COVID-19 é a doença causada por um coronavírus chamado SARS-CoV-2. A OMS tomou conhecimento deste novo vírus em 31 de dezembro de 2019, após um relatório de um grupo de casos de uma chamada pneumonia viral em Wuhan, República Popular da China. |
| Quem tem maior risco de desenvolver doença grave por COVID-19? | Pessoas com 60 anos ou mais e aquelas com problemas médicos subjacentes como pressão alta, diabetes, outros problemas de saúde crônicos (por exemplo, aqueles que afetam o coração, pulmões, rins e cérebro), função imunológica baixa / imunossupressão (incluindo HIV), obesidade, câncer e pessoas não vacinadas têm maior risco de doença grave. No entanto, qualquer pessoa de qualquer idade pode ficar doente com COVID-19 e se tornar gravemente enferma ou morrer. |
| Fonte: Portal da OMS | |

### 4.1.3 Seleção do contexto

Após a extração dos dados relevantes, separamos o que compõe contexto do que é pergunta. Contextos e perguntas são retirados juntamente, por isso, são separados para a formação adequada da base de dados a ser utilizada

### 4.1.4 Tratamento do contexto

Nessa etapa, o objetivo é tratar o contexto, removendo informações irrelevantes como links, dados da Organização Mundial da Saúde, nomes próprios, valores numéricos

[1] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub

e mantendo apenas as partes relevantes conforme a tabela 5, informações sobre o assunto principal, para a criação das perguntas. O contexto consiste na parte do texto que contém as informações necessárias para formular as perguntas.

**Tabela 5: Exemplo de Pergunta e resposta original e tratada**

| **Contexto** | **Contexto tratado** |
|---|---|
| Sim. Existem várias vacinas contra a COVID-19 validadas para uso pela OMS (concedidas para Uso de Emergência) e por outras agências reguladoras nacionais rigorosas (ARR). O primeiro programa de vacinação em massa começou no início de dezembro de 2020 e o número de doses de vacinas administradas é atualizado regularmente aqui no link. | Existem várias vacinas contra a COVID-19 validadas para uso pela OMS (concedidas para Uso de Emergência) e por outras agências reguladoras nacionais rigorosas (SRAs). O primeiro programa de vacinação em massa começou no início de dezembro de 2020 |
| Fonte: Portal da OMS | |

No exemplo apresentado, eliminamos o advérbio que surge de forma isolada, uma vez que, embora indique concordância, afirmação ou confirmação, não transmite uma compreensão clara ou substantiva. Adicionalmente, removemos a referência a um link, pois ele não fornece uma informação intrinsecamente valiosa no contexto proposto. Uma informação considerada válida neste contexto é aquela que oferece um esclarecimento direto e completo, sem a necessidade de redirecionamento a um link externo ou uma imagem.

### 4.1.5 Definição das perguntas e respostas baseadas no contexto

Nesta etapa, monta-se as perguntas originais, que são selecionadas a partir dos site da OMS, e têm respostas contidas no contexto selecionado nas etapas anteriores. Essas perguntas são projetadas para treinar a compreensão do modelo, sobre o contexto com objetivo de extrair respostas com informações relevantes. A partir de um contexto como *"COVID-19 é a doença causada por um coronavírus chamado SARS-CoV-2. A OMS teve conhecimento deste novo vírus em 31 de dezembro de 2019, após um relatório de um agrupamento de casos de uma chamada pneumonia viral em Wuhan, República Popular da China."* e a sua pergunta original *"O que é COVID?"* podemos cirar outras perguntas para expandir quantidade de pares de perguntas e respostas:

- Pergunta: O que é Covid?
- Resposta: doença causada por um coronavírus chamado SARS-CoV-2
- Pergunta: Onde a Covid começou?

- Resposta: em Wuhan, República Popular da China.
- Pergunta: O que é Covid?
- Resposta: COVID-19 é a doença causada por um coronavírus chamado SARS-CoV-2.
- Pergunta: Como é chamada a Covid?
- Resposta: chamada SARS-CoV-2
- Pergunta: Quando foi encontrado pela primeira vez?
- Resposta: A OMS tomou conhecimento deste novo vírus em 31 de dezembro de 2019

### 4.1.6 Aumento da base de dados

Aplicamos transformações visando aumentar a diversidade dos dados de treinamento, gerando variações sintéticas das perguntas existentes. A aplicação dessa técnica envolvem alterações no vocabulário, sinônimos, substituição de palavras-chave, inversão da ordem das palavras ou frases, entre outros. Tais modificações visam enriquecer o conjunto de dados culminando num aprimorar da capacidade do modelo em generalizar respostas.

Para ilustrar o tipo de transformação realizada, consideremos algumas exemplos: termos como "COVID"podem ser substituídos por "Coronavirus". Expressões como *"shortness of breath"* podem ser reordenadas, resultando em *"breath shortness"*. Além disso, frases padrão como *"The most common symptoms of COVID-19"* podem ser reformuladas para *"Typical symptoms of COVID-19"*.

Importante destacar que, em nosso trabalho, optamos por não adotar variações com estruturas negativas, apesar de o modelo BERT ter o melhor desempenho entre os modelos pre-treinados, apresenta dificuldades em lidar eficientemente com este tipo de construção sintática [21].

### 4.1.7 Conversão das informações ao modelo SQUAD

Nessa etapa, precisa-se identificar as respostas corretas para cada pergunta contida nos contextos. No SQUAD, algumas perguntas podem não ter respostas explícitas no contexto, sendo necessário identificar esses casos e marcá-los como "impossível"como resposta. Nossa base de dados não contemplará respostas que não estejam no contexto.

Algumas das etapas anteriores são realizadas usando uma ferramenta da Haystack, que é uma biblioteca de funcionalidades para o processamento de linguagem natural e treinamento de modelos de perguntas e respostas. Ela oferece uma interface simplificada para realizar tarefas de preparação de dados e facilitar o ajuste fino de modelos como o BERT, para a tarefa de perguntas e respostas.

## 4.2 Escopo definido

O BERT é mais adequado quando há um escopo bem definido para as perguntas e respostas. Como o conjunto de dados de treinamento é limitado, o BERT pode ser treinado especificamente para responder perguntas dentro desse escopo, tornando-o mais preciso e relevante nesse contexto [22].

## 4.3 Identificação de trechos relevantes

O BERT tem a capacidade de identificar trechos relevantes do texto de entrada que correspondem à pergunta. Ele leva em consideração a relação entre as palavras e sua posição no contexto, permitindo encontrar a resposta correta com base nas informações presentes no conjunto de dados de treinamento [23].

## 4.4 *Fine-tuning* específico

Ao realizar o *fine-tuning* do BERT em um conjunto de dados específico para o QA, o modelo pode capturar tanto informações sintáticas, quanto outras informações relevantes presentes nos dados de treinamento. Isso significa que o BERT pode aproveitar seu conhecimento prévio sobre sintaxe, para melhor compreender as perguntas e encontrar respostas adequadas, dentro do escopo definido pelo conjunto de dados de treinamento. Portanto, essa capacidade de aprender informações sintáticas, de forma "natural"pode contribuir para respostas mais precisas e relevantes durante o textitfine-tuning do BERT para o QA. [23].

## 4.5 Eficiência na modelagem de *tokens*

O BERT é eficiente na modelagem de textittokens, o que é vantajoso quando se lida com conjuntos de dados limitados. Ele consegue capturar informações contextuais importantes, mesmo com uma quantidade menor de dados, o que é crucial para o desempenho do QA em um contexto específico.

## 4.6 Limitação de generalização

O BERT possui limitações em termos de generalização para perguntas fora do escopo do conjunto de treinamento [23]. Essa é justamente a característica necessária para uma plataforma de interação com o usuário: A manutenção do assunto dentro de um escopo definido, sem que outras perguntas não correlacionadas possam ser feitas.

A escolha do BERT em sua versão *large* foi fundamentada na capacidade de processar e compreender informações de forma mais abrangente e detalhada. O BERT *large* é conhecido por seu maior tamanho, camadas e número de parâmetros, em comparação com outros modelos BERT base, o que permite uma representação mais rica e complexa do texto. Consideramos que o modelo *large* é mais adequado para enfrentar o desafio de responder perguntas em um contexto de tamanho de grandes proporções, como o proposto neste estudo. Sua capacidade de processar e integrar informações de múltiplas camadas e contextos anteriores pode contribuir para uma melhor compreensão e resposta das perguntas apresentadas.

Ao optar pelo BERT *large* (Tabela 6), buscamos explorar seu potencial de capturar nuances e relações semânticas mais sutis presentes nas perguntas e contextos. A escolha de um modelo *large* passa pelo tamanho do dataset de *fine-tuning* que caracteristicamente é de porte reduzido, pois os FAQs raramente são de grande porte e por isso é necessário ter o modelo mais efetivo em capturar contextos, porém adequado ao poder de processamento disponível [15].

**Tabela 6: Modelos BERT disponíveis e suas camadas internas**

| **Modelo** | **Camadas (milhões)** | **Camadas** | **Lingua** |
|---|---|---|---|
| bert-base-uncased | 110 | 6 - 9 | Inglês |
| bert-large-uncased | 340 | 14 - 19 | Inglês |
| bert-base-cased | 110 | 6 - 9 | Inglês |
| bert-large-cased | 340 | 14 - 19 | Inglês |
| Fonte: https://huggingface.co/bert-base-uncased | | | |

## 4.7 Obtenção do modelo BERT pré-treinado

### 4.7.1 Fine-tuning do modelo

Nessa etapa, um modelo pré-treinado, em uma tarefa geral de linguagem em grandes quantidades de dado, é adaptado e refinado para uma tarefa específica, ou conjunto de dados de interesse. A intenção é aproveitar o conhecimento adquirido na etapa de pré-treinamento e utilizar o refinamento para ajustar o modelo às perguntas e respostas que lhe serão feitas.

### 4.7.2 Pré-processamento de dados

No processo de *fine-tuning*, a primeira etapa é carregar os arquivos JSON do base de dados destinado ao ajuste fino do modelo para a tarefa de perguntas e respostas. Durante essa etapa, os dados são convertidos e armazenados em estruturas de dados adequadas para o treinamento do modelo.

Ao mesmo tempo, é realizado um pré-processamento dos dados, eliminando todas as informações do JSON que não são relevantes para a tarefa em questão. Essa etapa de carregamento e conversão dos dados é essencial para preparar o conjunto de dados de treinamento, garantindo que apenas as informações necessárias sejam mantidas. Uma outra etapa da preparação do base de dados para a tarefa de perguntas e respostas é o cálculo das respostas com base no número associado a cada exemplo no conjunto de dados.

O processo de cálculo das respostas envolve localizar o trecho exato do contexto que corresponde à resposta. Usando o número de início fornecido pelo SQUAD, é possível delimitar o trecho de texto que contém a resposta correta. Isso pode ser feito encontrando o índice de início e o índice de fim do trecho de texto dentro do contexto. Uma vez que o trecho de texto da resposta é identificado, ele pode ser extraído do contexto e usado como a resposta correta para a pergunta específica.

### 4.7.3 *Fine-tuning*

Durante a etapa de *fine-tuning*, empregamos a linguagem de programação Python em conjunto com a renomada biblioteca *Transformers* da *HuggingFace* e o *framework* PyTorch. Iniciamos com modelos BERT pré-treinados, que serviram como substrato para o nosso processo de ajuste fino. O conjunto de dados empregado para essa finalidade era composto por 352 pares de perguntas e respostas, cuidadosamente preparados para amplificar a coleção original de apenas 26 pares. Este material provém de um FAQ específico sobre o tema abordado. O *fine-tuning* foi efetuado em um tempo total de cerca de um minuto e trinta segundos, um período bastante curto que pode ser atribuído tanto à escala reduzida do conjunto de dados, à capacidade de processamento do computador pessoal utilizado conforme a Tabela 7 e o poder do modelo BERT utilizado.

**Tabela 7: Configurações de hardware**

| Hardware | Tipo | Capacidade |
|---|---|---|
| Processador | AMD Ryzen 7 5700X | 8-CORE 3500Mhz |
| Memória RAM | Kingston Fury DDR3 | 16GB 3500Mhz |
| Chip Gráfico | GeForce RTX3060Ti | 8Gb GDDR6 |

Foram realizados testes com diferentes conjuntos de hiperparâmetros para buscar o melhor resultado no treinamento do modelo, considerando métricas como acurácia, F1-score e perda [1]. A variação dos hiperparâmetros permitiu explorar diferentes combinações e encontrar a configuração que resultou em um desempenho ótimo para o problema em questão.

**Tabela 8: Configurações de Hiperparâmetros**

| **Hiperparâmetro** | **Experimento 1** | **Experimento 2** |
|---|---|---|
| Tamanho do modelo | 24 | 24 |
| Tamanho do lote | 16 | 16 |
| Taxa de aprendizado | $5 \times 10^{-5}$ | $2 \times 10^{-5}$ |
| Taxa de otimizador | $1 \times 10^{-3}$ | $1 \times 10^{-3}$ |
| Número de épocas | 2 e 3 | 2 e 3 |

### 4.7.4 Métricas

1. **Perdas**: As perdas de treino e validação são métricas utilizadas para avaliar o desempenho de um modelo durante o processo de treinamento. A perda de treino é calculada com base nos erros cometidos pelo modelo ao realizar previsões nos dados de treinamento, para tarefas de perguntas e respostas, como no SQUAD, a perda deve ser calculada tendo em vista a diferença entre as previsões do modelo para o início e o fim de uma resposta em relação às posições reais da resposta no texto.

$$\text{Perda} = -\frac{1}{N}\sum_{i=1}^{N}(y_i \log(p_i) + (1 - y_i)\log(1 - p_i)) \tag{1}$$

   - $N$ é o número total de exemplos (no caso do SQUAD, seria o número de pares pergunta-resposta no conjunto de treinamento).
   - $y_i$ é a verdadeira label (rótulo) para o exemplo $i$, que indica se a resposta começa ou termina na posição $i$ no texto de contexto.
   - $p_i$ é a probabilidade prevista pelo modelo de que a resposta começa ou termina na posição $i$ no texto de contexto.

2. **Medida-F1**: F1-score é calculado a partir da média harmônica entre duas métricas. Ele fornece uma única medida que equilibra tanto a precisão quanto o recall, levando em consideração os falsos positivos e falsos negativos. O resultado é uma medida que varia de 0 a 1, sendo 1 o melhor desempenho possível.

$$\text{F1-score} = 2 \times \frac{\text{Precisão} \times \text{Recall}}{\text{Precisão} + \text{Recall}} \tag{2}$$

Onde a precisão e o recall são calculados conforme representado a seguir (2) e (3).

$$\text{Precisão} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

A precisão é a proporção de verdadeiros positivos (TP) em relação ao total de exemplos classificados como positivos pelo modelo, ou seja, a capacidade do modelo de classificar corretamente os exemplos positivos.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

O Recall é a proporção de verdadeiros positivos em relação ao total de exemplos que realmente são positivos, ou seja, a capacidade do modelo de encontrar todos os exemplos positivos.

3. **Acurácia**: A acurácia é uma métrica que mede a precisão geral de um modelo de classificação, ou seja, a taxa de acertos em relação ao total de exemplos. Ela varia de 0 a 1, sendo 1 o valor máximo que indica que o modelo classifica corretamente todos os exemplos. A acurácia pode ser interpretada como a porcentagem de acertos do modelo, onde valores mais altos indicam um desempenho melhor.

$$\text{Acurácia} = \left( \frac{\text{Número de previsões corretas}}{\text{Número total de exemplos}} \right) \times 100\% \tag{5}$$

# 5 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Esta seção está organizada a partir dos hiperparâmetros diferentes definidos na Tabela 8 e a análise e comparação dos resultados obtidos em cada experimento. Os experimento estão divididos em duas e três épocas.

- Modelo do experimento 1
  - Duas épocas
  - Três épocas
- Modelo do experimento 2
  - Duas épocas
  - Três épocas
- Análise dos resultados

No estudo, cada conjunto de valores da Tabela 8 e optamos por dividir o contexto em quatro categorias distintas, devido ao desempenho insatisfatório ao usar um contexto único, ocasionado pelas limitações de hardware. Essa abordagem reduziu o custo computacional e melhorou o desempenho dos modelos de treinamento, alcançando resultados mais satisfatórios.

Observamos que o desempenho do modelo, quando aplicado a contextos mais amplos, não atingiu o nível de eficácia esperado. Diante dos resultados obtidos, optamos por investigar alternativas para refinar os resultados alcançados. Uma tática que implementamos envolveu a divisão do contexto em perguntas e respostas baseadas no contexto mais restritos, proporcionando ao modelo a capacidade de focar em áreas mais delimitadas.

## 5.1 Modelo do Experimento 1

Foram realizados diálogos idênticos em todos os cenários de teste para permitir uma comparação visual direta entre os resultados. Isso significa que os mesmos conjuntos de diálogos foram usados nos experimentos com diferentes configurações de hiperparâmetros.

### 5.1.1 Duas épocas

Realizando o *fine-tuning* com o conjunto de hiperparâmetros do experimento 1, obtem-se a disponibilização do modelo melhorado para responder a perguntas do contexto

**Tabela 9: Resultado do *fine-tuning* do experimento 1 com duas épocas**

| Descrição | 1ª Época | 2ª Época |
|---|---|---|
| Perda do treino | 1,611 | 0,775 |
| Perda da validação | 1,070 | 1,100 |
| Média F1 | 0,643 | 0,675 |
| Acurácia | 0,714 | 0,698 |
| Tempo da época (s) | 85,5 | 85,6 |

**Tabela 10: Amostra de diálogo do experimento 1 com duas épocas**

| Perguntas | Respostas |
|---|---|
| O que é COVID-19? | COVID-19 é a doença causada por um coronavírus chamado SARS-CoV-2. |
| Onde a COVID foi detectada pela primeira vez? | Wuhan, República Popular da China |
| Quais são os sintomas mais comuns da COVID? | Febre, calafrios, dor de garganta |
| Quais são os sintomas menos comuns da COVID-19? | Dores musculares |
| Quem tem maior risco de contrair COVID? | Pessoas com problemas de saúde preexistentes |
| O que acontece com as pessoas que contraem COVID-19? | À medida que a taxa de testes diminui, torna-se mais difícil saber quantas pessoas têm COVID-19 e não procuram tratamento |
| Como posso me proteger contra a COVID-19? | Seguindo medidas preventivas |
| Quais são as medidas preventivas? | Manter distância, usar máscara em espaços lotados e mal ventilados, praticar higiene das mãos, etiqueta respiratória (cobrir a boca e nariz com o cotovelo dobrado ou um lenço quando tossir ou espirrar), vacinar-se e manter as doses de reforço em dia |
| Existe um teste que confirma infecções por COVID-19? | Torna-se mais difícil saber quantas pessoas têm COVID-19 e não procuram tratamento |
| Existem dois tipos de testes para COVID? | Qualquer pessoa com sintomas como início súbito de febre e tosse |

de Covid retirado do sitio da OMS. O desempenho é medido pela perda do treino, perda do conjunto de validação, média F1 e acurácia durante as épocas, como na Tabela 9.

Na primeira época, houve uma diminuição na perda do treinamento de 1,611 para 0,775, indicando uma melhora no desempenho do modelo nessa métrica. Já a perda da

validação aumentou de 1,07 para 1,10, sugerindo uma possível piora no ajuste do modelo a novos dados.

A medida F1 aumentou de 64,3% para 67,5%, indicando uma melhora na capacidade do modelo em encontrar um equilíbrio, a acurácia diminuiu de 71,4% para 69,8%, indicando uma taxa de acerto menor, em relação à primeira época.

Esses resultados indicam que houve uma melhora na perda do treinamento e na medida F1, mas com uma queda na acurácia e um possível aumento na perda da validação. É importante levar em consideração essas informações para avaliar o desempenho geral do modelo conforme verificamos na Tabela 10.

### 5.1.2 Três épocas

Podemos observar, conforme nos mostra a Tabela 11, que ao longo das épocas, algumas métricas apresentaram evolução enquanto outras tiveram uma degradação, enquanto a perda da validação se manteve quase inalterada ao longo das épocas, indo de 1,11 na primeira, para 1,12 na terceira época, alcançando 0,93. Podemos ver na 12 a amostra de diálogo obtida com a interação com o modelo, existem alguns espaços vazios, o que demonstra uma dificuldade em gerar respostas condizentes com as perguntas.

**Tabela 11: Resultado do *fine-tuning* do experimento 1 com três épocas**

| Descrição | 1ª Época | 2ª Época | 3ª Época |
|---|---|---|---|
| Perda do treino | 1,56 | 0,788 | 0,532 |
| Perda da validação | 1,110 | 0,930 | 1,120 |
| Média F1 | 0,658 | 0,722 | 0,682 |
| Acurácia | 0,714 | 0,793 | 0,746 |
| Tempo da época (s) | 80,5 | 80,9 | 85,6 |

Isso pode sugerir que o modelo está tendo dificuldades em generalizar para dados não vistos anteriormente, resultando em uma piora no ajuste aos novos dados. No geral, os indicadores que evoluíram indicam que o modelo está aprendendo e melhorando seu desempenho nas métricas de interesse. No entanto, é importante monitorar de perto a perda da validação para garantir que o modelo não esteja sofrendo sobreajuste aos dados de treinamento.

## 5.2 Modelo do Experimento 2

Testou-se o conjunto de hiperparâmetros do Experimento 2, que possui uma taxa de aprendizado diferente. Em vez de usar a taxa de aprendizagem de $5 \times 10^{-5}$, utilizou-se o valor de $2 \times 10^{-5}$ no *fine-tuning* do modelo. Essa alteração pode ter um impacto

**Tabela 12: Amostra de diálogo do experimento 1 com três épocas**

| **Perguntas** | **Respostas** |
|---|---|
| O que é COVID-19? | COVID-19 é a doença causada por um coronavírus chamado SARS-CoV-2. |
| Onde a COVID foi detectada pela primeira vez? | Wuhan, República Popular da China |
| Quais são os sintomas mais comuns da COVID? | Os sintomas mais comuns da COVID-19 são: febre, calafrios, dor de garganta. |
| Quais são os sintomas menos comuns da COVID-19? | - |
| Quem tem maior risco de contrair COVID? | - |
| O que acontece com as pessoas que contraem a COVID-19? | À medida que a taxa de testes diminui, torna-se mais difícil saber quantas pessoas têm COVID-19 e não procuram nenhum tratamento. |
| Como posso me proteger contra a COVID-19? | Seguindo medidas preventivas |
| Quais são as medidas preventivas? | Manter distância, usar máscara em locais lotados e mal ventilados, praticar higiene das mãos, etiqueta respiratória (cobrir a boca e o nariz com o cotovelo dobrado ou um lenço quando tossir ou espirrar), vacinar-se e manter-se atualizado com as doses de reforço |
| Existe um teste que confirma infecções por COVID-19? | - |
| Existem dois tipos de testes para COVID? | À medida que a taxa de testes diminui, torna-se mais difícil saber quantas pessoas têm COVID-19 e não procuram nenhum tratamento. |

significativo no processo de treinamento do modelo, pois influencia na rapidez com que o modelo aprende com os dados.

Durante o Experimento 2, fez-se uso dos mesmos conjuntos de diálogos utilizados nos experimentos anteriores. Isso permitiu realizar uma comparação visual direta entre os resultados obtidos com diferentes configurações de hiperparâmetros, como o número de épocas.

Essa abordagem sistemática de testar diferentes conjuntos de hiperparâmetros nos permite explorar as variações no desempenho do modelo e identificar quais configurações podem ser mais adequadas para o nosso problema específico. Prosseguiu-se com os testes para análise de resultados obtidos com o conjunto de hiperparâmetros do Experimento 2.

### 5.2.1 Duas épocas

Ao comparar os resultados obtidos com diferentes conjuntos de hiperparâmetros da tabela 13, podemos observar algumas diferenças, para o Experimento 2, a perda do treinamento diminuiu de 1,49 para 0,65, em relação ao Experimento 1. Isso indica uma melhora de aproximadamente 56,1% no desempenho do modelo em ajustar-se aos dados de treinamento. Ao avaliar as perguntas e respostas da tabela 14 vemos que há menos precisão nas respostas, incluindo algumas respostas vazias representadas por um hífen.

**Tabela 13: Resultado do *fine-tuning* do experimento 2 com duas épocas**

| Descrição | 1ª Época | 2ª Época |
|---|---|---|
| Perda do treino | 1.490 | 0.650 |
| Perda da validação | 0.990 | 0.830 |
| Média F1 | 0.794 | 0.777 |
| Acurácia | 0.714 | 0.793 |
| Tempo da época (s) | 87.2 | 86.1 |

A perda da validação diminuiu de 0,99 para 0,83 em comparação ao Experimento 1. Isso sugere uma melhora de aproximadamente 16,2% na capacidade do modelo de generalizar para dados não vistos durante o treinamento. A média F1 aumentou de 79,4% para 77,7% em relação ao Experimento 1. Isso indica uma melhora de aproximadamente 1,9% na capacidade do modelo de encontrar um equilíbrio entre precisão e recall. A acurácia permaneceu praticamente a mesma, com valores de 71,4% e 79,3% para os Experimento 1 e Experimento 2, respectivamente. O tempo da época aumentou ligeiramente, de 85,5 segundos para 87,2 segundos, comparando com o Experimento 1. Esses resultados sugerem que o Experimento 2 pode ter um desempenho ligeiramente melhor em relação à perda e à medida F1 quando comparados com os resultados do Experimento 1. No entanto, não houve uma melhora significativa na acurácia. É importante levar em consideração essas informações para avaliar o desempenho geral do modelo e selecionar o conjunto de hiperparâmetros mais adequado para o problema em questão.

É fundamental observar que, apesar das respostas apresentarem uma extensão considerável, elas permaneceram corretas e precisas. Essa abordagem mais abrangente fornece informações completas e aprofundadas. Contudo, é preciso equilibrar a concisão com a entrega de respostas diretas e sucintas, levando em consideração as preferências por respostas com mais precisão.

### 5.2.2 Três épocas

Obteremos aqui os resultados do *fine-tuning* com 3 épocas.

**Tabela 14: Amostra de diálogo do Experimento 2 com duas épocas**

| **Perguntas** | **Respostas** |
|---|---|
| O que é COVID-19? | COVID-19 é a doença causada por um coronavírus chamado SARS-CoV-2. |
| Onde a COVID foi detectada pela primeira vez? | Wuhan, República Popular da China |
| Quais são os sintomas mais comuns da COVID? | Os sintomas mais comuns da COVID-19 são: febre, calafrios, dor de garganta. |
| Quais são os sintomas menos comuns da COVID-19? | Dores musculares, fadiga severa ou cansaço, nariz escorrendo ou entupido, espirros, dor de cabeça, (...) dor persistente ou pressão no peito, pele fria ou pegajosa, ou mudança para cor pálida ou azulada, perda de fala ou movimento. |
| Quem tem maior risco de contrair COVID? | Pessoas com problemas de saúde preexistentes têm maior risco ao contrair COVID-19; |
| O que acontece com pessoas que contraem COVID-19? | Com a queda das taxas de testagem, fica mais difícil saber quantas pessoas têm COVID-19 e não buscam tratamento. |
| Como posso me proteger contra COVID-19? | Você pode se proteger e aos outros da COVID-19 seguindo medidas preventivas, como manter distância, usar máscara em locais lotados e mal ventilados, praticar higiene das mãos, etiqueta respiratória (cobrir boca e nariz com o cotovelo dobrado ou lenço ao tossir ou espirrar), vacinar-se e manter as doses de reforço em dia. |
| Quais são as medidas preventivas? | Você pode se proteger e aos outros da COVID (...) vacinando-se e mantendo as doses de reforço em dia. |
| Existe um teste que confirma infecções por COVID-19? | Qualquer pessoa com sintomas como início agudo de febre e tosse deve ser testada, sempre que possível, para garantir que receba os cuidados clínicos apropriados. |
| Existem dois tipos de testes para COVID? | Qualquer pessoa com sintomas como início agudo de febre e tosse deve ser testada, sempre que possível, para garantir que receba os cuidados clínicos apropriados. |

Ao comparar os dois conjuntos de resultados, podemos observar algumas diferenças significativas. No primeiro conjunto, conforme vemos na tabela 15, temos perdas de treino e validação mais baixas em comparação com o segundo conjunto, o que indica que o modelo está se ajustando melhor aos dados de treinamento. Além disso, a medida F1 e a acurácia também são ligeiramente mais altas no primeiro conjunto, ou seja, o primeiro conjunto de resultados possui um desempenho geral melhor, com perdas menores e métricas de desempenho mais altas. Pode-se verificar os diálogos resultantes do Experimento 2 com três épocas na tabela 16

**Tabela 15: Resultado do *fine-tuning* do Experimento 2 com três épocas**

| Descrição | 1ª Época | 2ª Época | 3ª Época |
|---|---|---|---|
| Perda do treino | 1.550 | 0.720 | 0.480 |
| Perda da validação | 1.150 | 1.00 | 0.970 |
| Média F1 | 0.675 | 0.722 | 0.730 |
| Acurácia | 0.714 | 0.777 | 0.825 |
| Tempo da época | 84.9 s | 83.1 s | 83.9 s |

## 5.3 Análise dos resultados

As análises dos dados e das amostras de diálogo revelaram que, para o objetivo de implementar um chatbot baseado em BERT para responder a perguntas e fornecer respostas em um contexto de QA, o modelo de Configuração de Experimento 1 e Configuração do Experimento 2, ambos com duas épocas, se destacaram como os melhores em termos de desempenho.

Ao examinar as métricas relevantes, como perda de treinamento, perda de validação, medida F1, acurácia e tempo de época, observamos que ambos exibiram resultados superiores, em comparação com os não selecionados. Além disso, uma análise mais aprofundada das respostas geradas pelos modelos revelou que as respostas fornecidas pelos conjunto escolhido eram mais eficazes em termos de relevância e precisão.

Essas descobertas reforçam a importância de testar os conjuntos de hiperparâmetros e ajustar os modelos durante o treinamento para obter resultados otimizados. Ao considerar as nuances específicas do nosso cenário de QA, concluímos que os testes c com Hiperparâmentro 2 se destacaram como a melhor escolha, proporcionando respostas mais efetivas e satisfatórias aos usuários da ferramenta.

**Tabela 16: Amostra de diálogo do experimento 2 com três épocas**

| **Perguntas** | **Respostas** |
|---|---|
| O que é COVID-19? | COVID-19 é a doença causada por um coronavírus chamado SARS-CoV-2. |
| Onde a COVID foi detectada pela primeira vez? | Wuhan, República Popular da China |
| Quais são os sintomas mais comuns da COVID? | Os sintomas mais comuns da COVID-19 são: febre, calafrios, dor de garganta. |
| Quais são os sintomas menos comuns da COVID-19? | Dores musculares, fadiga severa ou cansaço, nariz escorrendo ou bloqueado, ou espirros, dor de cabeça, olhos doloridos, tontura, tosse nova e persistente, aperto no peito ou dor no peito, falta de ar, voz rouca, braços/pernas pesados, dormência, dificuldade em respirar, especialmente em repouso, ou incapacidade de falar em frases, confusão, sonolência ou perda de consciência, dor persistente ou pressão no peito, pele fria ou pegajosa, ou ficando pálida ou de cor azulada, perda de fala ou movimento. |
| Quem tem maior risco de contrair COVID? | Pessoas com problemas de saúde preexistentes estão em maior risco quando têm COVID-19. |
| O que acontece com as pessoas que contraem COVID-19? | À medida que a taxa de testes diminui, torna-se mais difícil saber quantas pessoas têm COVID-19 e não procuram tratamento. |
| Como posso me proteger contra a COVID-19? | Você pode se proteger e proteger os outros da COVID-19 seguindo medidas preventivas, como manter a distância, usar máscara em locais lotados e mal ventilados, praticar higiene das mãos, etiqueta respiratória (cobrir a boca e nariz com o cotovelo dobrado ou um lenço ao tossir ou espirrar), vacinar-se e manter as doses de reforço em dia. |
| Quais são as medidas preventivas? | Manter distância, usar máscara em espaços lotados e mal ventilados, praticar higiene das mãos, etiqueta respiratória (cobrir a boca e nariz com o cotovelo dobrado ou um lenço quando tossir ou espirrar), vacinar-se e manter as doses de reforço em dia. |
| Existe um teste que confirma infecções por COVID-19? | - |
| Existem dois tipos de testes para COVID? | - |

# 6 CONCLUSÕES E TRABALHOS FUTUROS

Na presente pesquisa, realizamos o *fine-tuning* de um modelo BERT "large"pre-treinado em um conjunto de dados construído especificamente para a tarefa de Question Answering. Os resultados obtidos foram bastante promissores, demonstrando a eficácia e o potencial do BERT na capacidade de responder perguntas em contextos limitados.

Ao avaliar o desempenho do modelo, observamos que o aumento da variedade de perguntas no conjunto de dados do *fine-tuning* contribuiu significativamente para a melhoria dos resultados. Isso indica a importância de garantir uma diversidade de perguntas, abrangendo diferentes estruturas e contextos, a fim de treinar o modelo de forma mais completa e representativa.

Além disso, adotamos estratégias para lidar com o desafio de contextos extensos, o que favoreceu a assertividade do modelo. Ao reduzir o tamanho dos contextos, conseguimos direcionar a atenção do modelo para informações mais relevantes e evitar possíveis ruídos ou confusões causados pelo tamanho do contexto somado à capacidade de processamento e memória disponível. Essa abordagem se mostrou eficiente para otimizar a capacidade de resposta do modelo, especialmente em cenários em que o contexto é extenso ou complexo.

Como trabalho futuro, deve-se explorar ainda mais a diversidade de perguntas e contextos, buscando aprimorar ainda mais o desempenho do modelo. Além disso, é importante investigar estratégias adicionais para lidar com contextos longos e complexos, como por exemplo, utilizar modelos BERT mais robustos pode ser uma direção promissora para melhorar ainda mais a assertividade do modelo.

Em suma, este estudo evidência a relevância e o potencial do uso de modelos BERT pré-treinados em SQuAD para a tarefa de QA e ajustado para o assunto específico. Os resultados alcançados estimulam novas pesquisas nessa área.

# REFERÊNCIAS


[1] DEVLIN, Jacob et al. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

[2] SHINDE, P.; SHAH, S.. A Review of Machine Learning and Deep Learning Applications. IEEE, 2018.

[3] LECUN, Y.; BENGIO, Y.; HINTON, G. Deep Learning. Nature, v. 521, n. 7553, p. 436-444, 2015.

[4] BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation Learning: A Review and New Perspectives. IEEE, 2013.

[5] RADFORD, A.; WU, J.; CHILD, R.; LUAN, D.; AMODEI, D.; SUTSKEVER, I. Language models are unsupervised multitask learners. 2019.

[6] VASWANI, A. et al. Attention is all you need. In: Advances in Neural Information Processing Systems. p. 6000-6010, 2017.

[7] RAJPURKAR, Pranav et al. Know What You Don't Know: Unanswerable Questions for SQuAD. arXiv preprint arXiv:1806.03822, 2016.

[8] LIDDY, Elizabeth D. Natural language processing. 2001.

[9] JUHN, Young; LIU, Hongfang. Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. Journal of Allergy and Clinical Immunology, v. 145, n. 2, p. 463-469, 2020. ISSN 0091-6749

[10] KATTE, Trupti. Recurrent Neural Network and its Various Architecture Types. International Journal of Research and Scientific Innovation (IJRSI), v. V, n. III, p. 124, Março 2018. ISSN 2321–2705.

[11] LI, Zewen; LIU, Fan; YANG, Wenjie; PENG, Shouheng; ZHU, Jun. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, [s.l.], v. XX, n. XX, p. 1, 2021. Disponível em: [https://ieeexplore.ieee.org/abstract/document/9451544]. Acesso em: [16 de Outubro de 2023].

[12] YIN, Wenpeng et al. Comparative study of CNN and RNN for natural language processing. arXiv preprint arXiv:1702.01923, 2017.

[13] WOLF, Thomas et al. Transformers: State-of-the-art natural language processing. In: Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations. p. 38-45, 2020.

[14] LI, M. Question Answering on SQuAD 2.0. 2021. Disponível em: https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1194/reports/default/15816213.pdf. Acesso em: 14 de junho de 2023.

[15] LIU, Yinhan et al. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692, 2019.

[16] LAN, Zhenzhong et al. Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942, 2019.

[17] SUN, Jiumeng et al. COVID-19: epidemiology, evolution, and cross-disciplinary perspectives. Trends in molecular medicine, v. 26, n. 5, p. 483-495, 2020.

[18] FALCÃO, Paula et al. Pandemia de desinformação: as fake news no contexto da Covid-19 no Brasil. 2021.

[19] MAHARANA, Kiran; MONDAL, Surajit; NEMADE, Bhushankumar. A review: Data pre-processing and data augmentation techniques. Global Transitions Proceedings, 2022.

[20] KUMAR, V., CHOUDHARY, A., & CHO, E. Data augmentation using pre-trained transformer models. arXiv preprint arXiv:2003.02245, 2020.

[21] KASSNER, Nora; SCHÜTZE, Hinrich. Negated and misprimed probes for pretrained language models: Birds can talk, but cannot fly. arXiv preprint arXiv:1911.03343, 2019.

[22] ROBERTS, Adam; RAFFEL, Colin; SHAZEER, Noam. How much knowledge can you pack into the parameters of a language model?. arXiv preprint arXiv:2002.08910, 2020.

[23] ROGERS, Anna; KOVALEVA, Olga; RUMSHISKY, Anna. A Primer in BERTology: What We Know About How BERT Works. Transactions of the Association for Computational Linguistics, v. 8, p. 842-866, 2020. DOI: 10.1162/tacl_a_00349.

# ANEXO A – ANEXOS E APÊNDICES 1

Conforme mencionado nos resultados, analisou-se o desempenho do modelo (Tabela 17 e Figura 6) diante de um contexto extenso, composto por 11.697 caracteres. Essa análise buscou avaliar a capacidade do modelo em lidar com contextos mais amplos e complexos. Uma análise contribuiu para uma melhor compreensão das capacidades e limitações do modelo.

**Tabela 17: Números do treinamento do Experimento extra**

| Descrição | 1ª Época | 2ª Época |
|---|---|---|
| Perda do treino | 1.445 | 0.670 |
| Perda da validação | 1.32 | 1.91 |
| Média F1 | 0.193 | 0.145 |
| Acurácia | 0.177 | 0.169 |
| Tempo da época | 95.5 s | 92.6 s |

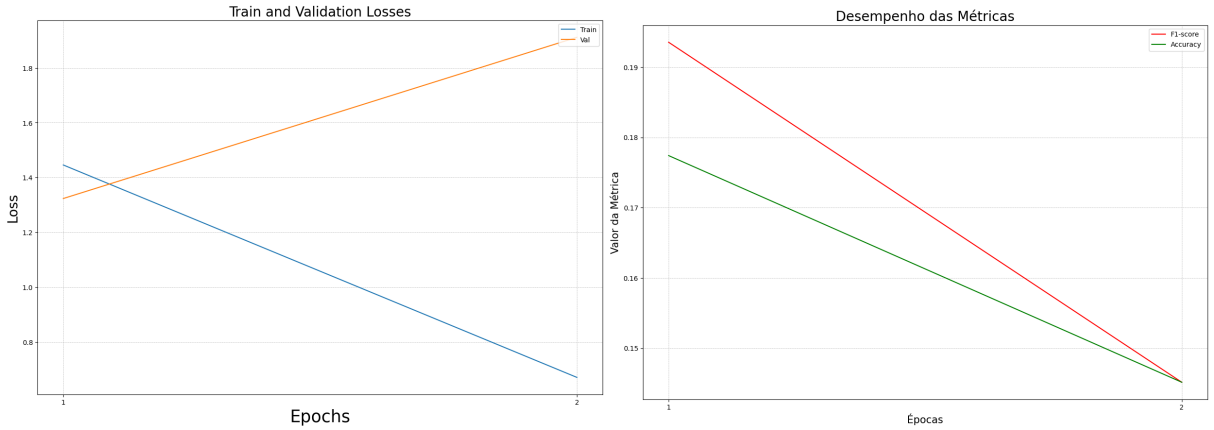


**Figura 6: Perdas durante o treino e Desempenho das métricas**

Constatamos que o desempenho do modelo treinado e utilizados em um contexto grande não foi tão efetivo quanto o esperado. Diante dessa limitação, decidimos explorar outras estratégias para aprimorar os resultados obtidos. Uma abordagem que adotamos foi a segregação do contexto em subtemas distintos, permitindo que o modelo se concentrasse em contextos mais específicos e facilitando a compreensão e geração de respostas mais precisas. Essa abordagem mostrou-se promissora, uma vez que melhorou significativamente o desempenho do modelo, proporcionando respostas mais adequadas e relevantes, conforme vemos na seção de conclusão.

Identificamos que o modelo utilizado apresentou limitações ao lidar com grandes contextos. Notamos na Tabela 18 que as respostas dadas não foram eficientes, as perdas do conjunto de validação cresceram e a acurácia diminuiu ao longo das épocas.

É crucial destacar que a obtenção de um desempenho ótimo exige o emprego de versões mais avançadas do modelo BERT, caracterizadas por um maior número de camadas e uma utilização do mecanismo *multi-head* para atenção. A necessidade de hardware de alta performance para o *fine-tuning* desses modelos robustos são decorrentes da intensidade computacional necessária para processar o aumento no volume e complexidade das camadas e *multi-heads*. Tal treinamento intensivo é demandante em termos de recursos computacionais, devido ao elevado número de operações inerentes a cada camada e à complexidade agregada pelos *multi-heads* [16].

Portanto, reconhecemos que, embora a utilização de modelos mais avançados fosse desejável para lidar com grandes contextos, a viabilidade dessa abordagem é condicionada à disponibilidade de hardware adequado.

**Tabela 18: Amostra de diálogo do Experimento extra**

| **Perguntas** | **Respostas** |
|---|---|
| O que é COVID-19? | COVID-19 é a doença causada por um coronavírus chamado SARS-CoV-2. |
| Onde foi detectado pela primeira vez o COVID? | Wuhan, República Popular da China |
| Quais são os sintomas mais comuns da COVID? | Os sintomas mais comuns da COVID-19 são: febre, calafrios, dor de garganta. |
| Quais são os sintomas menos comuns da COVID-19? | Dores musculares, fadiga severa ou cansaço, nariz escorrendo ou bloqueado, ou espirros, dor de cabeça, olhos doloridos, tontura, tosse nova e persistente, aperto no peito ou dor no peito, falta de ar, voz rouca, braços/pernas pesados, dormência (...), confusão, sonolência ou perda de consciência, dor persistente ou pressão no peito, pele fria ou pegajosa, ou ficando pálida ou de cor azulada, perda de fala ou movimento. |
| Quem tem maior risco de contrair COVID? | - |
| O que acontece com as pessoas que contraem COVID-19? | COVID-19 é a doença causada por um coronavírus chamado SARS-CoV-2. |
| Como posso me proteger contra a COVID-19? | COVID-19 é a doença causada por um coronavírus chamado SARS-CoV-2. |
| Quais são as medidas preventivas? | - |
| Existe algum teste que confirma as infecções por COVID-19? | COVID-19 é a doença causada por um coronavírus chamado SARS-CoV-2. |
| Existem dois tipos de testes para COVID? | Os sintomas mais comuns da COVID-19 são: febre, calafrios, dor de garganta. |

# ANEXO B – ANEXOS E APÊNDICES 2

Apresentamos os valores de referência utilizados para avaliar as respostas nos diálogos. Os valores de referência representam as respostas esperadas e consideradas corretas. Para realizar a comparação entre os resultados obtidos, adotamos um sistema de pontuação com os seguintes critérios:

**Tabela 19: Pontuação de análise**

| Pontuação | Descrição |
|---|---|
| 8 | Respostas 100% corretas e exatas |
| 6 | Respostas muito satisfatórias |
| 4 | Respostas satisfatórias |
| 2 | Respostas pouco satisfatórias |
| 0 | Respostas vazias ou incorretas |

Cada resultado obtido, considerando diferentes hiperparâmetros, quantidade de épocas e métodos de utilização do contexto, será avaliado de forma individualizada ao final comparada entre si.Para cada combinação de parâmetros, iremos analisar o desempenho subjetivo do modelo. Resultados de métricas como perda do treino, perda da validação, média F1 e acurácia serão avaliadas na conclusão da pesquisa, considerando os melhores resultados obtidos aqui:

- **T1**: Teste com conjunto de dados segregado usando hiperparâmetros listados do Experimento 1, presentes na Tabela 4, por 2 épocas.
- **T2**: Teste com conjunto de dados segregado usando hiperparâmetros listados do Experimento 1, presentes na Tabela 4, por 3 épocas.
- **T3**: Teste com conjunto de dados segregado usando hiperparâmetros listados do Experimento 2, presentes na Tabela 4, por 2 épocas.
- **T4**: Teste com conjunto de dados segregado usando hiperparâmetros listados do Experimento 1, presentes na Tabela 4, por 3 épocas.
- **T5**: Teste com conjunto de dados inteiro como contexto usando hiperparâmetros listados do Experimento 1, presentes na Tabela 4, por 2 épocas.
- **Tt**: Total de pontuação aferida para cada pergunta a fim de calcular a dificuldade de respondê-la no geral.

As informações serão atribuídas a cada resposta dadas pelo modelo presentes na Tabela 21, levando em consideração as estratégias de contexto, hiperparâmetros e quantidade de épocas correspondentes.

- **TOTAL**: Soma total dos coeficientes obtidos em cada pergunta pelos modelos em seus conjuntos de teste, sendo 80 a pontuação máxima e 0 a mínima.
- **TOTAL (%)**: Total percentual obtido pelo conjunto de teste de acordo com o valor máximo a ser alcançado pelo modelo.

**Tabela 20: Desempenho subjetivo dos testes**

| **Perguntas** | **T1** | **T2** | **T3** | **T4** | **T5** | **Tt** |
|---|---|---|---|---|---|---|
| What is COVID-19? | 8 | 8 | 8 | 8 | 8 | 40 |
| Where was COVID first detected? | 8 | 8 | 8 | 8 | 8 | 40 |
| What are the most common COVID sympt.? | 6 | 8 | 8 | 8 | 8 | 30 |
| What are the less common sympt. of COVID? | 4 | 0 | 6 | 6 | 4 | 20 |
| Who has a higher risk of having COVID-19? | 8 | 0 | 4 | 4 | 0 | 16 |
| What happens to people who get COVID-19? | 4 | 8 | 8 | 8 | 0 | 28 |
| How can I protect myself against COVID-19? | 6 | 4 | 8 | 8 | 0 | 26 |
| What are the preventive measures? | 8 | 8 | 6 | 8 | 0 | 30 |
| Is there a test that confirms COVID infect.? | 0 | 0 | 0 | 0 | 0 | 0 |
| Are there two types of tests for COVID? | 0 | 0 | 0 | 0 | 0 | 0 |
| TOTAL | 52 | 44 | 56 | 58 | 28 | - |
| TOTAL (%) | 65,0 | 55,0 | 70,0 | 72,5 | 35,8 | - |

Diante dos resultados obtidos, procedemos à comparação dos índices percentuais de acurácia relativa entre diferentes modelos, resultando na conclusão de que o T5 apresentou o menor desempenho, obtendo uma taxa de acerto de apenas 35,8%. Isso é inferior aos 55% obtidos pelo T2, o qual demonstrou o pior desempenho, em termos de respostas dentro do conjunto de teste com contextos segregados.

Além disso, uma constatação clara é a diferença de desempenho entre contextos de diferentes tamanhos. Observou-se que o modelo escolhido não apresentou um bom desempenho em contextos extensos, enquanto obteve resultados mais satisfatórios em contextos mais reduzidos.

Essas conclusões demonstram a necessidade de avaliar cuidadosamente as características dos contextos e considerar a capacidade do modelo selecionado em lidar com diferentes volumes de informação. Tais observações são relevantes para futuras melhorias e aprimoramentos no desenvolvimento de modelos de processamento de linguagem natural.

**Tabela 21: Amostra de diálogo do Experimento extra**

| **Perguntas** | **Respostas** |
|---|---|
| O que é COVID-19? | COVID-19 é a doença causada por um coronavírus chamado SARS-CoV-2. |
| Onde foi detectada a primeira vez a COVID? | Wuhan, República Popular da China |
| Quais são os sintomas mais comuns da COVID? | Os sintomas mais comuns da COVID-19 são: febre, calafrios, dor de garganta. |
| Quais são os sintomas menos comuns da COVID-19? | Outros sintomas menos comuns que podem afetar alguns pacientes incluem: dores musculares, fadiga severa ou cansaço, nariz escorrendo ou bloqueado, ou espirros, dor de cabeça, olhos doloridos, tontura, tosse nova e persistente, aperto no peito ou dor no peito, falta de ar, voz rouca, braços/pernas pesados, dormência/formigamento, náusea, vômito, dor abdominal/na barriga, ou diarreia, perda de apetite, perda ou alteração do sentido do paladar ou olfato, dificuldade para dormir. |
| Quem tem maior risco de ter COVID? | Pessoas com 60 anos ou mais e aquelas com problemas médicos preexistentes como pressão alta, diabetes, outros problemas crônicos de saúde (por exemplo, afetando o coração, pulmões, rins e cérebro), função imunológica baixa/imunossupressão (incluindo HIV), obesidade, câncer e pessoas não vacinadas estão mais em risco de doença grave. |
| O que acontece com as pessoas que contraem COVID-19? | À medida que as taxas de testes caem, fica mais difícil saber quantas pessoas têm COVID-19 e não procuram tratamento. |
| Como posso me proteger contra a COVID-19? | Você pode se proteger e proteger os outros da COVID-19 seguindo medidas preventivas, como manter distância, usar máscara em locais lotados e mal ventilados, praticar higiene das mãos, etiqueta respiratória (cobrindo a boca e o nariz com o cotovelo dobrado ou um lenço ao tossir ou espirrar), vacinando-se e mantendo as doses de reforço em dia. |
| Quais são as medidas preventivas? | Manter distância, usar máscara em locais lotados e mal ventilados, praticar higiene das mãos, etiqueta respiratória (cobrindo a boca e o nariz com o cotovelo dobrado ou um lenço ao tossir ou espirrar), vacinar-se e manter as doses de reforço em dia. |
| Existe um teste que confirma infecções por COVID-19? | Existem dois principais tipos de testes que podem confirmar se você está infectado com SARS-CoV-2 |
| Existem dois tipos de testes para COVID? | Existem dois principais tipos de testes que podem confirmar se você está infectado com SARS-CoV-2, o vírus que causa a COVID-19. |

# ANEXO C – ANEXOS E APÊNDICES 3

As pergunta e respostas retiradas do sitio da OMS como referência para a criação do conjunto de dados utilizado no *fine-tuning*.

**Perguntas e Respostas**

**What is COVID-19**: COVID-19 is the disease caused by a coronavirus called SARS-CoV-2. WHO first learned of this new virus on 31 December 2019, following a report of a cluster of cases of so-called viral pneumonia in Wuhan, People's Republic of China.

**What are the symptoms of COVID-19?** The most common symptoms of COVID-19 are: fever, chills and sore throat. Other symptoms that are less common and may affect some patients include: muscle aches, severe fatigue or tiredness, runny or blocked nose, or sneezing, headache, sore eyes, dizziness, new and persistent cough, tight chest or chest pain, shortness of breath, hoarse voice, heavy arms/legs, numbness/tingling, nausea, vomiting, abdominal pain/ belly ache, or diarrhoea, appetite loss, loss or change of sense of taste or smell and difficulty sleeping. Symptoms of severe COVID-19 disease which need immediate medical attention include: difficulty in breathing, especially at rest, or unable to speak in sentences, confusion, drowsiness or loss of consciousness, persistent pain or pressure in the chest, skin being cold or clammy, or turning pale or a bluish colour and loss of speech or movement. If possible, call your health care provider, hotline or health facility first, so you can be directed to the right clinic. People who have pre-existing health problems are at higher risk when they have COVID-19; they should seek medical help early if worried about their condition. These include, but are not limited to: those taking immunosuppressive medication; those with chronic heart, lung, liver or rheumatological problems; those with HIV, diabetes, cancer or dementia.

**What happens to people who get COVID-19?** As testing rates fall, it is more difficult to know how many people have COVID-19 and do not seek any treatment. At the start of the pandemic, 15% of people were thought to become seriously unwell and require hospital treatment and oxygen. More recent estimates suggest that hospitalization is required in around 3% of people with COVID-19. This is partly due to immunization, partly due to changes in the virus (especially the Omicron variants), and partly due to the availability of targeted medical treatments. Most people make a full recovery without needing hospital treatment. For those with COVID-19 who are at high risk of severe illness (see question below), WHO has made recommendations on which drug treatments are effective in improving outcomes and preventing hospital admissions. It is also important to be vigilant in recognizing people with severe disease and those needing hospital treatment so that they are treated early. The consequences of severe COVID-19

include death, respiratory failure, sepsis, thromboembolism (blood clots), and multiorgan failure, including injury of the heart, liver or kidneys. In rare situations, children can develop a severe inflammatory syndrome a few weeks after infection.

**Who is most at risk of severe illness from COVID-19?** People aged 60 years and over, and those with underlying medical problems like high blood pressure, diabetes, other chronic health problems (for example those affecting the heart, lungs, kidneys, and brain), low immune function / immunosuppression (including HIV), obesity, cancer, and unvaccinated people are most at risk of severe illness. However, anyone at any age can get sick with COVID-19 and become seriously ill or die.

**Are there long-term effects of COVID-19?** Some people who have had COVID-19, whether they have needed hospitalization or not, continue to experience symptoms, including fatigue, respiratory and neurological symptoms. These long-term effects are called post COVID-19 condition (also called long COVID). For further information on long COVID, see post COVID-19 condition and questions and answers.

**How can we protect others and ourselves if we don't know who is infected?** ou can protect yourself and others from COVID-19 by following preventive measures, such as keeping a distance, wearing a mask in crowded and poorly ventilated spaces, practicing hand hygiene, respiratory etiquette (covering your mouth and nose with a bent elbow or a tissue when you cough or sneeze), getting vaccinated and staying up to date with booster doses. Check local advice where you live and work. Read our public advice page for more information.

**When should I get a test for COVID-19?** Anyone with symptoms such as acute onset of fever and cough should be tested, wherever possible, to ensure that they receive appropriate clinical care. People who do not have symptoms but have had close contact with someone who is, or may be, infected may also consider testing. Contact your local health guidelines and follow their guidance. While a person is waiting for test results, they should preferably wear a mask when interacting with others in or outside of their household or sharing space with others. Where testing capacity is limited, tests should first be done for those at higher risk of infection, such as health workers, and those at higher risk of severe illness such as older people, especially those living in seniors' residences or long-term care facilities. Individuals with signs and symptoms suggestive of COVID-19 or those who test positive for the virus should wear a mask when interacting with others in or outside of one's household or sharing space with others.

**What test should I get to see if I have COVID-19?** There are two main types of tests that can confirm whether you are infected with SARS-CoV-2, the virus that causes COVID-19. Molecular tests, such as polymerase chain reaction (PCR), are the most accurate tests for diagnosing SARS-CoV-2 infection. Molecular tests detect virus

in the sample by amplifying viral genetic material to detectable levels. Rapid antigen tests (sometimes known as rapid diagnostic tests or RDTs) detect viral proteins (known as antigens). RDTs are a simpler and faster option than molecular tests and are available for testing by trained operators or by the individual themselves (sometimes called self-tests). They perform best when there is more virus circulating in the community and when sampled from an individual during the time they are most infectious, generally within the first 5–7 days following symptom onset. Samples for both types of tests are collected from the nose and/or throat with a swab.Learn more about what kind of COVID-19 tests are available Why testing for SARS-CoV-2 is important Use of Ag-RDTs What you need to know about self-testing

**I want to find out if I had COVID-19 in the past, what test could I take?** Antibody tests can tell us whether someone has had an infection in the past, even if they have not had symptoms. Also known as serological tests, these tests detect antibodies produced in response to an infection or vaccination. In most people, antibodies start to develop after days to weeks and can indicate if a person has had past infection or has been vaccinated. Antibody tests cannot be used to diagnose SARS-CoV-2 infection in the early stages of infection or disease but can indicate whether or not someone has had the disease in the past. If you have been vaccinated, many antibody tests are not able to distinguish between whether you were previously infected or have been vaccinated (or both), so in such situations you will test positive in either case.

**What is the difference between isolation and quarantine?** Both isolation and quarantine are methods of preventing the spread of COVID-19. Quarantine is used for certain persons who are a contact of someone infected with SARS-CoV-2, the virus that causes COVID-19, whether the infected person has symptoms or not. Quarantine means that you remain separated from others because you have been exposed to the virus and you may be infected and can take place in a designated facility or at home. For COVID-19, this means staying in the facility or at home for several days. Isolation is used for people with COVID-19 symptoms or who have tested positive for the virus. Being in isolation means being separated from other people, ideally in a medical facility where you can receive clinical care. If isolation in a medical facility is not possible and you are not in a high-risk group for developing severe disease, isolation can take place at home. If you have symptoms, you should remain in isolation for at least 10 days. If you are infected and do not develop symptoms, you should remain in isolation for 5 days from the time you test positive. You can be discharged from isolation early if you test negative on a rapid antigen test.

**What should I do if I have been exposed to someone who has COVID-19?** If you have been exposed to someone with COVID-19, you may become infected, even if you feel well. After exposure to someone who has COVID-19, do the following:

Call your healthcare provider to get tested or test yourself. Stay home if you feel unwell. Wear a mask when interacting with others in or outside of your household or sharing space with others. Clean your hands regularly. Practice respiratory etiquette: cover your mouth and nose with your bent elbow or a tissue when you cough or sneeze. Avoid crowded, enclosed or poorly ventilated spaces. Get vaccinated and stay up to date with booster doses.

**How long does it take to develop symptoms?** The time from exposure to COVID-19 to the moment when symptoms begin is, on average, 5–6 days and can range from 1–14 days. This is why people who have been exposed to the virus are advised to remain at home and stay away from others in order to prevent the spread of the virus.

**Is there a vaccine for COVID-19?** Yes. There are several COVID-19 vaccines validated for use by WHO (given Emergency Use Listing) and from other stringent national regulatory agencies (SRAs). The first mass vaccination programme started in early December 2020 and the number of vaccination doses administered is updated on a regular basis here. For more information on vaccine for COVID-19, see the vaccine questions and answers: Coronavirus disease (COVID-19): Vaccines (who.int)

**What should I do if I have COVID-19 symptoms?** If you are unwell, stay at home. If you have any symptoms suggestive of COVID-19, wear a mask when interacting with others in or outside of your household or sharing space with others. If you have shortness of breath or pain or pressure in the chest, seek medical attention at a health facility immediately. Call your health care provider or hotline in advance for direction to the right health facility. Get tested for COVID-19, regardless of your vaccination status, and especially if you are at high-risk for severe illness and could therefore be eligible for drug treatments. Practice protective and preventive measures. Wear a mask, avoid crowded and poorly ventilated areas, improve ventilation in indoor spaces, keep a distance, practice hand hygiene, and respiratory etiquette (covering your mouth and nose with a bent elbow or a tissue when you cough or sneeze), get vaccinated and stay up to date with booster doses.

**Are there treatments for COVID-19?** There has been huge progress in developing treatments for patients with COVID-19. Treatments for COVID-19 should be decided on an individual basis between the person and the healthcare professional looking after them. The choice will depend on the severity of disease and the risk of disease worsening (including the person's age and whether they have any health problems). WHO maintains a list of recommended treatments, together with the evidence for each, available at https://app.magicapp.org/#/guideline/nBkO1E. These currently include: for non-severe COVID-19: nirmatrelvir-ritonavir; molnupiravir; remdesivir for severe COVID-19: corticosteroids (including dexamethasone); IL-6 receptor blockers (tocilizumab or sarilumab); baricitinib; remdesivir. Apart from these medications, amongst

the most commonly and globally used, and important treatments is oxygen for severely ill patients. WHO leads work which aims to improve global capacity and access to oxygen production, distribution and supply to patients.

For corticosteroids, further specific information can be found in the Corticosteroids, including dexamethasone questions and answers.

**Are antibiotics effective in preventing or treating COVID-19?** Antibiotics do not work against viruses; they only work on bacterial infections. COVID-19 is caused by a virus, so antibiotics do not work. Antibiotics should not be used as a means of prevention or treatment of COVID-19.

In hospitals, physicians will sometimes use antibiotics to treat secondary bacterial infections, which can be a complication of COVID-19 in severely ill patients. They should only be used as directed by a physician to treat a bacterial infection.

# ANEXO B – ANEXOS E APÊNDICES 4

Nas Tabelas 22, 23, 24, 25 e 26 contém os diálogos, compostos de perguntas feitas e as respectivas respostas fornecidas, de acordo com o conjunto de hiperparâmetros de cada experimento em inglês, idioma original do modelo, com duas ou três épocas.

**Tabela 22: Amostra de diálogo do Experimento 1 com duas épocas em inglês**

| **Perguntas** | **Respostas** |
|---|---|
| What is COVID-19? | COVID-19 is the disease caused by a coronavirus called SARS-CoV-2. |
| Where was COVID first detected? | Wuhan, People's Republic of China |
| What are the most common COVID symptoms? | Fever, chills, sore throat |
| What are the less common symptoms of COVID-19? | Muscle aches |
| Who has a higher risk of having COVID? | People who have pre-existing health problems |
| What happens to people who get COVID-19? | As testing rates fall, it is more difficult to know how many people have COVID-19 and do not seek any treatment |
| How can I protect myself against COVID-19? | By following preventive measures |
| What are the preventive measures? | Keeping a distance, wearing a mask in crowded and poorly ventilated spaces, practicing hand hygiene, respiratory etiquette (covering your mouth and nose with a bent elbow or a tissue when you cough or sneeze), getting vaccinated and staying up to date with booster doses |
| Is there a test that confirms COVID-19 infections? | It is more difficult to know how many people have COVID-19 and do not seek any treatment |
| Are there two types of tests for COVID? | Anyone with symptoms such as acute onset of fever and cough |

**Tabela 23: Amostra de diálogo do Experimento 1 com três épocas em inglês**

| **Perguntas** | **Respostas** |
|---|---|
| What is COVID-19? | COVID-19 is the disease caused by a coronavirus called SARS-CoV-2. |
| Where was COVID first detected? | Wuhan, People's Republic of China |
| What are the most common COVID symptoms? | The most common symptoms of COVID - 19 are : fever, chills, sore throat. |
| What are the less common symptoms of COVID-19? | - |
| Who has a higher risk of having COVID? | - |
| what happens to people who get covid-19? | As testing rates fall, it is more difficult to know how many people have COVID - 19 and do not seek any treatment. |
| How can I protect myself against COVID-19? | By following preventive measures |
| What are the preventive measures? | Keeping a distance, wearing a mask in crowded and poorly ventilated spaces, practicing hand hygiene, respiratory etiquette (covering your mouth and nose with a bent elbow or a tissue when you cough or sneeze), getting vaccinated and staying up to date with booster doses |
| Is there a test that confirms COVID-19 infections? | - |
| Are there two types of tests for COVID? | As testing rates fall, it is more difficult to know how many people have COVID - 19 and do not seek any treatment. |

**Tabela 24: Amostra de diálogo do Experimento 2 com duas épocas em inglês**

| **Perguntas** | **Respostas** |
|---|---|
| What is COVID-19? | COVID-19 is the disease caused by a coronavirus called SARS-CoV-2. |
| Where was COVID first detected? | Wuhan, People's Republic of China |
| What are the most common COVID symptoms? | The most common symptoms of COVID - 19 are : The most common symptoms of COVID - 19 are : fever, chills, sore throat. |
| What are the less common symptoms of COVID-19? | muscle aches, severe fatigue or tiredness, runny or blocked nose, or sneezing, headache, (...) persistent pain or pressure in the chest, skin being cold or clammy, or turning pale or a bluish colour, loss of speech or movement. |
| Who has a higher risk of having COVID? | People who have pre - existing health problems are at higher risk when they have COVID - 19 ; |
| what happens to people who get covid-19? | As testing rates fall, it is more difficult to know how many people have COVID - 19 and do not seek any treatment. |
| How can I protect myself against COVID-19? | You can protect yourself and others from COVID - 19 by following preventive measures, such as keeping a distance, wearing a mask in crowded and poorly ventilated spaces, practicing hand hygiene, respiratory etiquette ( covering your mouth and nose with a bent elbow or a tissue when you cough or sneeze ), getting vaccinated and staying up to date with booster doses. |
| What are the preventive measures? | You can protect yourself and others from COVID (...) getting vaccinated and staying up to date with booster doses. |
| Is there a test that confirms COVID-19 infections? | Anyone with symptoms such as acute onset of fever and cough should be tested, wherever possible, to ensure that they receive appropriate clinical care. |
| Are there two types of tests for COVID? | Anyone with symptoms such as acute onset of fever and cough should be tested, wherever possible, to ensure that they receive appropriate clinical care. |

**Tabela 25: Amostra de diálogo do Experimento 1 com três épocas em inglês**

| **Questions** | **Answers** |
|---|---|
| What is COVID-19? | COVID-19 is the disease caused by a coronavirus called SARS-CoV-2. |
| Where was COVID first detected? | Wuhan, People's Republic of China |
| What are the most common COVID symptoms? | The most common symptoms of COVID-19 are: fever, chills, sore throat. |
| What are the less common symptoms of COVID-19? | Muscle aches, severe fatigue or tiredness, runny or blocked nose, or sneezing, headache, sore eyes, dizziness, (...) persistent pain or pressure in the chest, skin being cold or clammy, or turning pale or a bluish color, loss of speech or movement. |
| Who has a higher risk of having COVID? | People who have pre-existing health problems are at higher risk when they have COVID-19; |
| What happens to people who get COVID-19? | As testing rates fall, it is more difficult to know how many people have COVID-19 and do not seek any treatment. |
| How can I protect myself against COVID-19? | You can protect yourself and others from COVID-19 by following preventive measures such as keeping a distance, wearing a mask in crowded and poorly ventilated spaces, practicing hand hygiene, respiratory etiquette (covering your mouth and nose with a bent elbow or a tissue when you cough or sneeze), getting vaccinated and staying up to date with booster doses. |
| What are the preventive measures? | Keeping a distance, wearing a mask in crowded and poorly ventilated spaces, practicing hand hygiene, respiratory etiquette (covering your mouth and nose with a bent elbow or a tissue when you cough or sneeze), getting vaccinated and staying up to date with booster doses. |
| Is there a test that confirms COVID-19 infections? | - |
| Are there two types of tests for COVID? | - |

**Tabela 26: Amostra de diálogo do Experimento extra em inglês**

| **Perguntas** | **Respostas** |
|---|---|
| What is COVID-19? | COVID-19 is the disease caused by a coronavirus called SARS-CoV-2. |
| Where was COVID first detected? | Wuhan, People's Republic of China |
| What are the most common COVID symptoms? | The most common symptoms of COVID - 19 are : fever, chills, sore throat. |
| What are the less common symptoms of COVID-19? | muscle aches, severe fatigue or tiredness, runny or blocked nose, or sneezing, headache, sore eyes, dizziness, new and persistent cough, tight chest or chest pain, shortness of breath, hoarse voice, heavy arms / legs, numbness (...) difficulty in breathing, especially at rest, or unable to speak in sentences, confusion, drowsiness or loss of consciousness, persistent pain or pressure in the chest, skin being cold or clammy, or turning pale or a bluish colour, loss of speech or movement. |
| Who has a higher risk of having COVID? | - |
| What happens to people who get COVID-19? | COVID - 19 is the disease caused by a coronavirus called SARS - CoV - 2. |
| How can I protect myself against COVID-19? | COVID - 19 is the disease caused by a coronavirus called SARS - CoV - 2. |
| What are the preventive measures? | - |
| Is there a test that confirms COVID-19 infections? | COVID - 19 is the disease caused by a coronavirus called SARS - CoV - 2. |
| Are there two types of tests for COVID? | The most common symptoms of COVID - 19 are : fever, chills, sore throat. |

**Tabela 27: Amostra de diálogo do Experimento extra em inglês**

| **Perguntas** | **Respostas** |
|---|---|
| What is COVID-19? | COVID-19 is the disease caused by a coronavirus called SARS-CoV-2. |
| Where was COVID first detected? | Wuhan, People's Republic of China |
| What are the most common COVID symptoms? | The most common symptoms of COVID - 19 are : The most common symptoms of COVID-19 are: fever, chills, sore throat. |
| What are the less common symptoms of COVID-19? | Other symptoms that are less common and may affect some patients include: muscle aches, severe fatigue or tiredness, runny or blocked nose, or sneezing, headache, sore eyes, dizziness, new and persistent cough, tight chest or chest pain, shortness of breath, hoarse voice, heavy arms/legs, numbness/tingling, nausea, vomiting, abdominal pain/ belly ache, or diarrhoea, appetite loss, loss or change of sense of taste or smell, difficulty sleeping. |
| Who has a higher risk of having COVID? | People aged 60 years and over, and those with underlying medical problems like high blood pressure, diabetes, other chronic health problems (for example those affecting the heart, lungs, kidneys, and brain), low immune function / immunosuppression (including HIV), obesity, cancer, and unvaccinated people are most at risk of severe illness. |
| What happens to people who get COVID-19? | As testing rates fall, it is more difficult to know how many people have COVID-19 and do not seek any treatment. |
| How can I protect myself against COVID-19? | You can protect yourself and others from COVID-19 by following preventive measures, such as keeping a distance, wearing a mask in crowded and poorly ventilated spaces, practicing hand hygiene, respiratory etiquette (covering your mouth and nose with a bent elbow or a tissue when you cough or sneeze), getting vaccinated and staying up to date with booster doses. |
| What are the preventive measures? | keeping a distance, wearing a mask in crowded and poorly ventilated spaces, practicing hand hygiene, respiratory etiquette (covering your mouth and nose with a bent elbow or a tissue when you cough or sneeze), getting vaccinated and staying up to date with booster doses. |
| Is there a test that confirms COVID-19 infections? | There are two main types of tests that can confirm whether you are infected with SARS-CoV-2 |
| Are there two types of tests for COVID? | There are two main types of tests that can confirm whether you are infected with SARS-CoV-2, the virus that causes COVID-19. |