



Identificação de sinais em Libras

Abordagem utilizando Visão Computacional

Integrantes



Elisa Ayumi Masasi de Oliveira
ayumi@discente.ufg.br



Evellyn Nicole Machado Rosa
nicole@discente.ufg.br



Iago Alves Brito
iagoalves@discente.ufg.br



Samuel França da Costa Pedrosa
samuelpedrosa@discente.ufg.br

Línguagem Brasileira de Sinais

Libras é a sigla para Língua Brasileira de Sinais, uma língua de modalidade gestual-visual em que é possível se comunicar por meio de gestos, expressões faciais e corporais.





NO BRASIL:

2,3 MILHÕES

DE PESSOAS POSSUEM ALGUM GRAU DE SURDEZ

IBGE, 2021

35,8%

DAS PESSOAS COM DEFICIÊNCIA AUDITIVA PROFUNDA
SABEM USAR A LIBRAS

PNS



04



Dataset

Desafios enfrentados

Poucos dados disponíveis

Para português, não foi encontrado nenhum dataset que satisfizesse a proposta

Aprender Libras para criar o Dataset

Foi necessário aprender algumas paravras e frases chave para criar o dataset

Diferença entre o alfabeto e frases

A proposta exigia a inferência em vídeo, pois apesar do alfabeto ter símbolos estáticos, as frases não tem



Identificando os sinais em Libras através de vídeos

O projeto propõe uma abordagem para identificação de sinais em Libras em tempo real por meio de vídeos.

Libras

Motivações para a escolha

Acessibilidade

Para muitos brasileiros, a única forma de se comunicar se dá através de libras

Inexistência de soluções

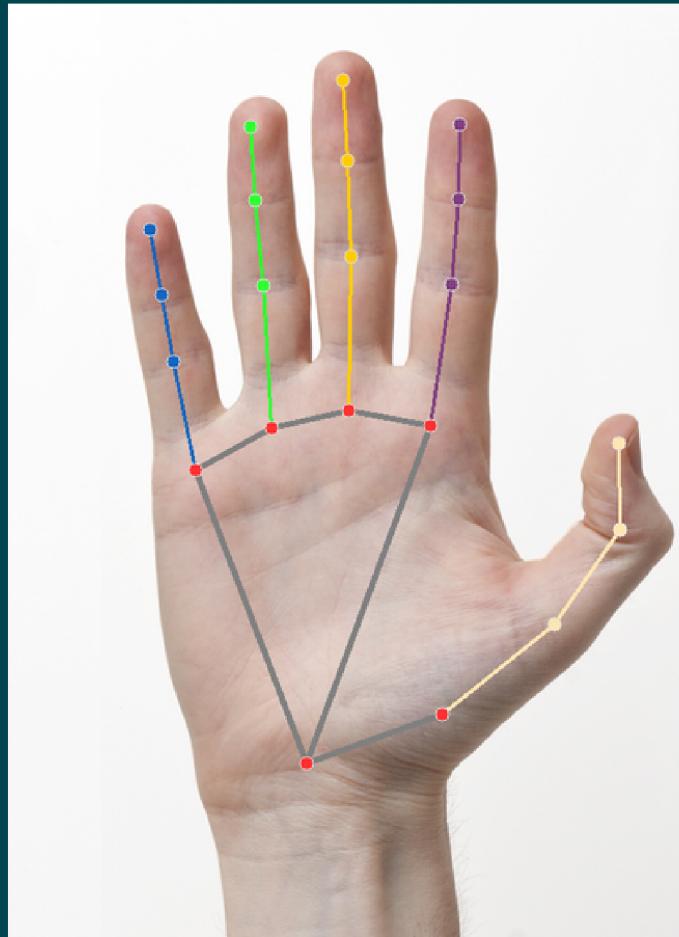
As principais soluções atuais contemplam apenas o alfabeto, sendo incapaz de lidar com palavras ou frases completas

Escalabilidade

A solução proposta tem alta escalabilidade sem aumento nenhum de custo

DETECÇÃO E RASTREAMENTO

FERRAMENTA UTILIZADA PARA DETECÇÃO DAS MÃOS - MEDIPIPE



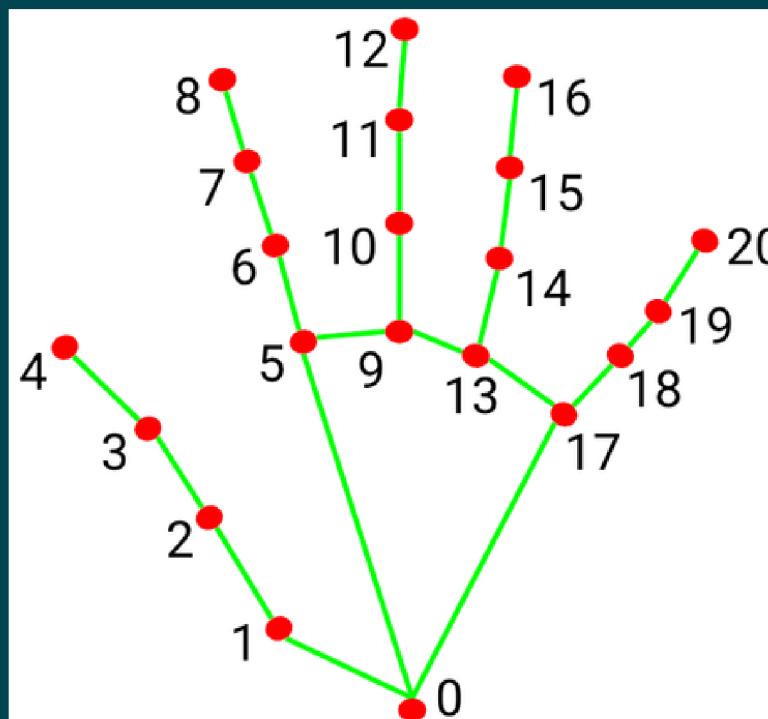
Utilizando como referência o Hands - MediaPipe:

- frames de vídeos passam por extração de features;
- Essas features passam por um modelo pré-treinado para detecção de objetos;
- Detecta-se pontos-chaves da mão para identificação do posicionamento e posteriormente do sinal representado.

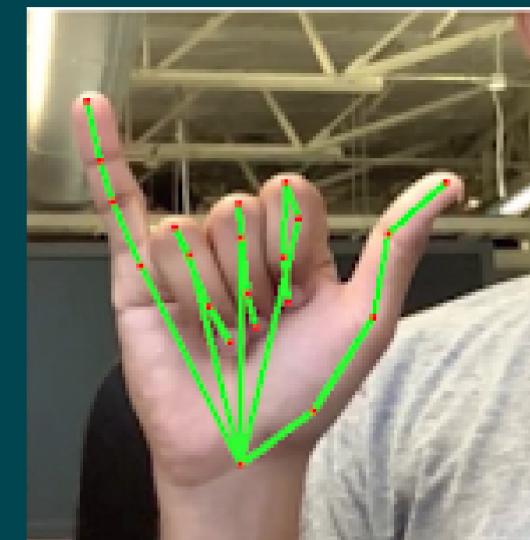
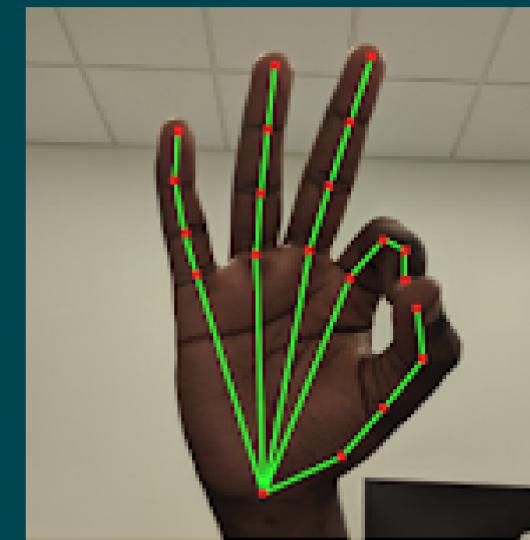
DETECÇÃO E RASTREAMENTO

FERRAMENTA UTILIZADA PARA DETECÇÃO DAS MÃOS - MEDIAPIPE

São 21 pontos-chaves definidos pelo MediaPipe para identificação do posicionamento da mão. Capacitando o detector em não só localizar uma mão em uma imagem, mas também a forma como ela se encontra. Extremamente útil para identificação de sinais.

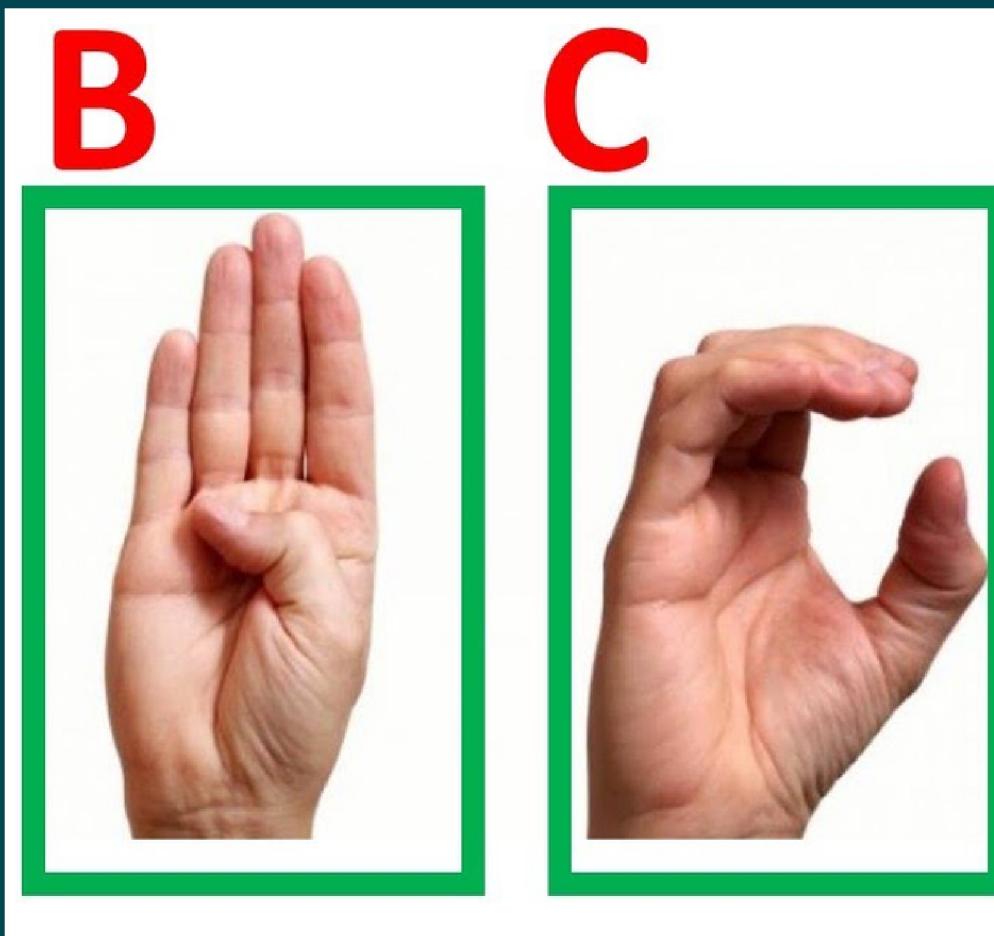


- | | |
|-----------------------|-----------------------|
| 0. WRIST | 11. MIDDLE_FINGER_DIP |
| 1. THUMB_CMC | 12. MIDDLE_FINGER_TIP |
| 2. THUMB_MCP | 13. RING_FINGER_MCP |
| 3. THUMB_IP | 14. RING_FINGER_PIP |
| 4. THUMB_TIP | 15. RING_FINGER_DIP |
| 5. INDEX_FINGER_MCP | 16. RING_FINGER_TIP |
| 6. INDEX_FINGER_PIP | 17. PINKY_MCP |
| 7. INDEX_FINGER_DIP | 18. PINKY_PIP |
| 8. INDEX_FINGER_TIP | 19. PINKY_DIP |
| 9. MIDDLE_FINGER_MCP | 20. PINKY_TIP |
| 10. MIDDLE_FINGER_PIP | |



DETECÇÃO E RASTREAMENTO

FERRAMENTA UTILIZADA PARA DETECÇÃO DAS MÃOS - MEDIPIPE



Para rastreamento, MediaPipe cria um Bounding Box da mão e utiliza os ponto-chaves já localizados para o próximo frame do vídeo, deixando o processo menos pesado e capacitado de ser utilizado em um vídeo/stream.

Cada frame coletado do vídeo fornece o posicionamento da mão e seus ponto-chaves que são utilizados em uma Rede Neural e uma rotulação de sinais com seus respectivos significados.

MediaPipe também possui ferramentas semelhantes para rosto e ombros, auxiliando na identificação dos articuladores dos sinais.

Redes Neurais Recorrentes

Escolhida devido à sua capacidade única de lidar com dados sequenciais e capturar relações temporais.

1

LSTM (LONG SHORT-TERM MEMORY)

A LSTM é uma variante da RNN projetada para superar desafios de gradientes que desaparecem, permitindo capturar dependências de longo prazo. Ela usa memória especial para armazenar informações ao longo do tempo.

Redes Neurais Recorrentes

Escolhida devido à sua capacidade única de lidar com dados sequenciais e capturar relações temporais.

1 LSTM (LONG SHORT-TERM MEMORY)

A LSTM é uma variante da RNN projetada para superar desafios de gradientes que desaparecem, permitindo capturar dependências de longo prazo. Ela usa memória especial para armazenar informações ao longo do tempo.

2 FLEXIBILIDADE PARA SEQUÊNCIAS VARIÁVEIS

RNNs se adaptam a sequências variáveis, permitindo reconhecimento de sinais com diferentes durações. Sua estrutura ajustável às sequências facilita identificar fases de um sinal, tornando-as ideais para interpretar gestos complexos, como em linguagens de sinais.

Redes Neurais Recorrentes

Escolhida devido à sua capacidade única de lidar com dados sequenciais e capturar relações temporais.

1 LSTM (LONG SHORT-TERM MEMORY)

A LSTM é uma variante da RNN projetada para superar desafios de gradientes que desaparecem, permitindo capturar dependências de longo prazo. Ela usa memória especial para armazenar informações ao longo do tempo.

2 FLEXIBILIDADE PARA SEQUÊNCIAS VARIÁVEIS

RNNs se adaptam a sequências variáveis, permitindo reconhecimento de sinais com diferentes durações. Sua estrutura ajustável às sequências facilita identificar fases de um sinal, tornando-as ideais para interpretar gestos complexos, como em linguagens de sinais.

3 CAPTURA DE PADRÕES TEMPORAIS

RNNs eficazes para reconhecer sinais ao capturar padrões temporais em sequências, identificando relações complexas e variações ao longo do tempo, adequadas para interpretar sinais sequenciais.

Resultados

Matriz de confusão

Classe "Oi"

Real

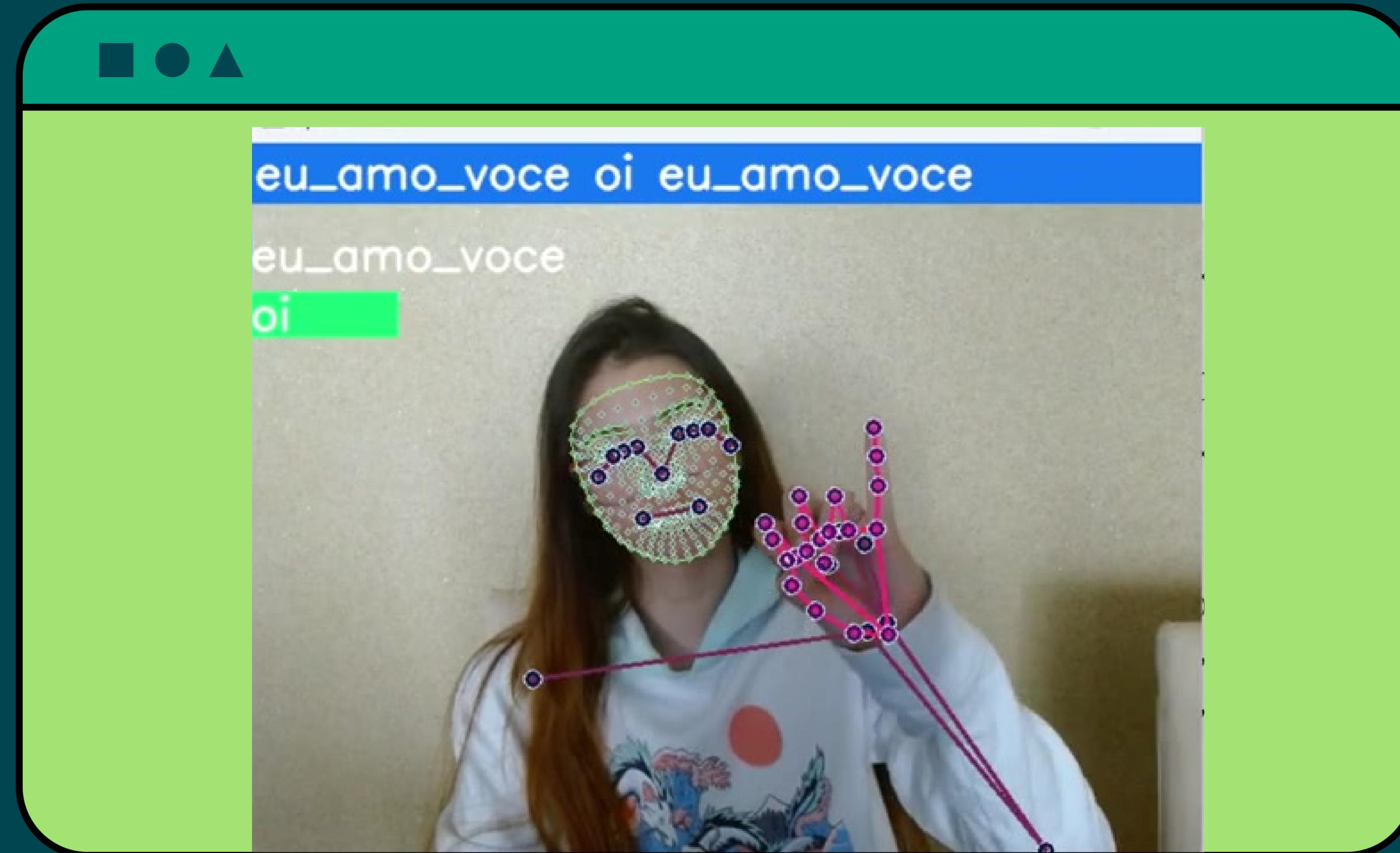
	Positivo	Negativo
Positivo	1	0
Negativo	0	2

Classe "Eu te amo"

Real

	Positivo	Negativo
Positivo	2	0
Negativo	0	1

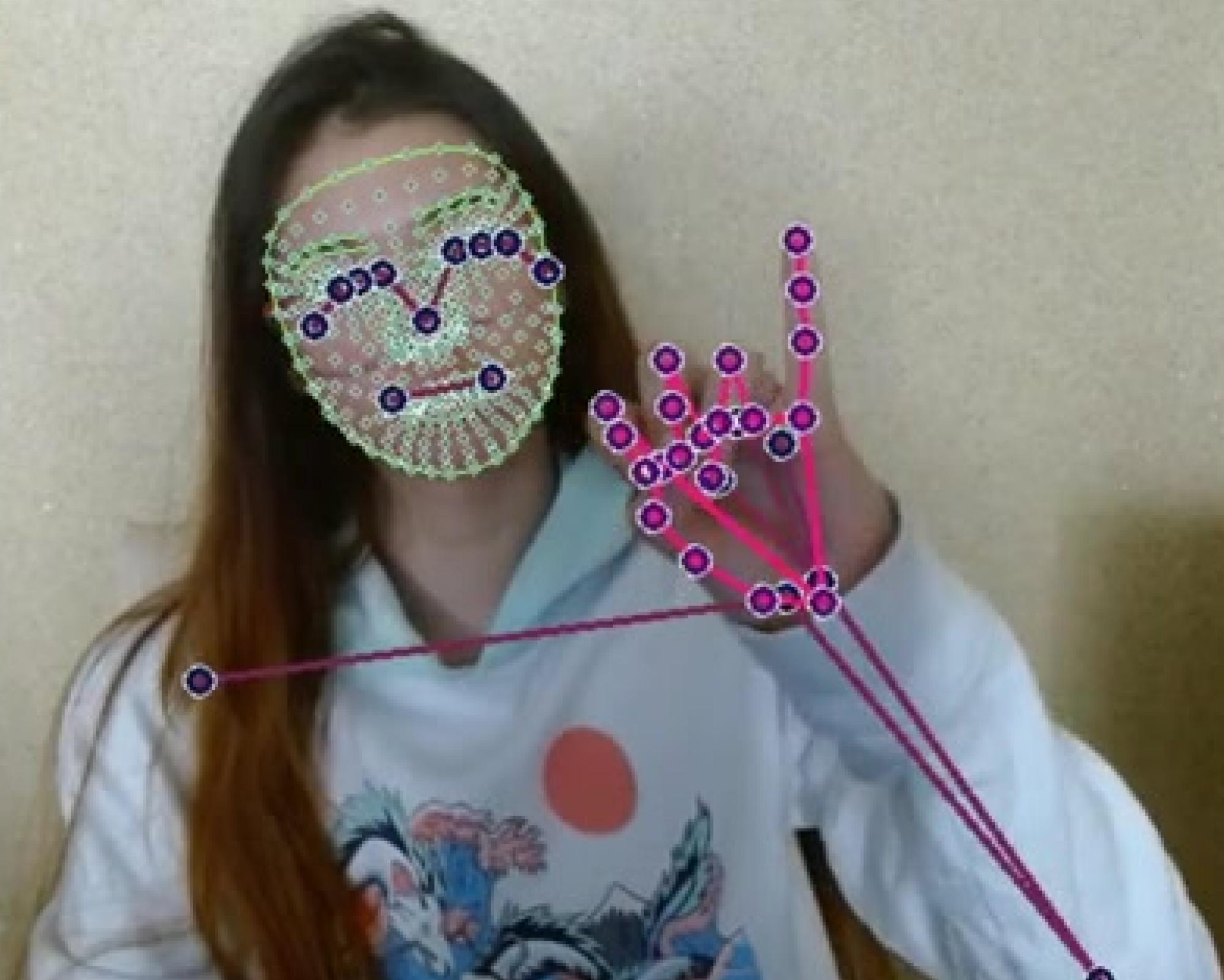
Resultados

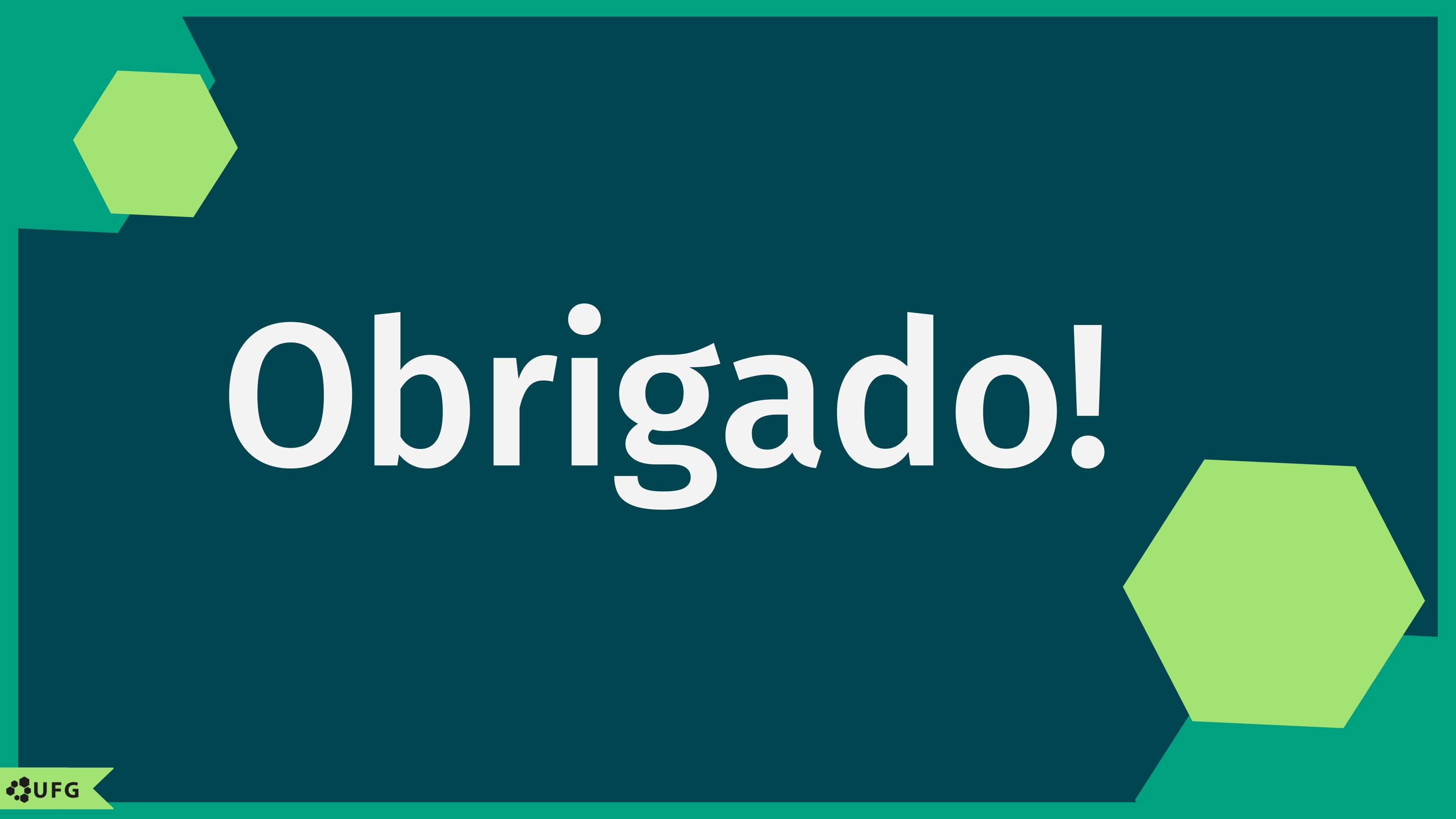


eu_amo_voce oi eu_amo_voce

eu_amo_voce

oi





Obrigado!