

Resolvendo tarefas de NLP com few/one-shot learning

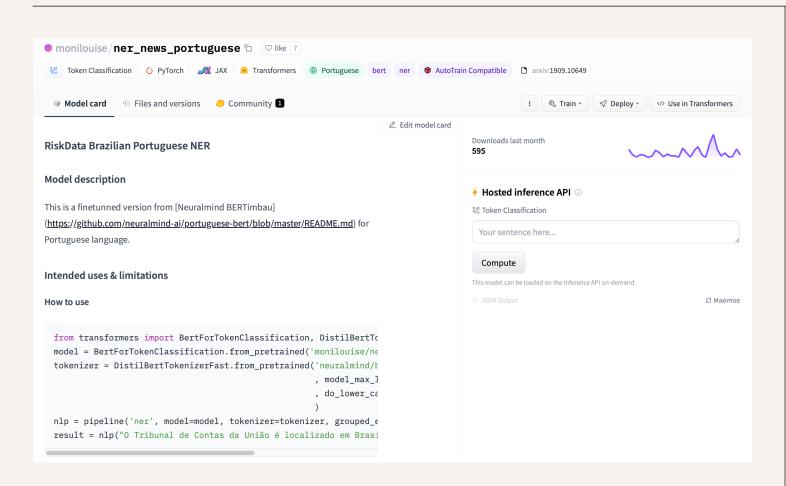
MONIQUE
MONTEIRO MONIQUELOUISE@
GMAIL.COM

# Objetivo

Validar técnicas de few/one-shot learning para a tarefa de reconhecimento de entidades nomeadas com o uso da API do ChatGPT (GPT 3.5 Turbo)

Comparar desempenho obtido com métricas coletadas em fine-tuning anterior

# Parâmetro de comparação



- ner\_news\_portuguese
- Fonte:
   <a href="https://huggingface.co/monilouise/ner\_news\_portuguese">https://huggingface.co/monilouise/ner\_news\_portuguese</a>
- Métricas:
  - precision: 0.86
  - recall: 0.91
  - f1: 0.88

#### Redução dos textos e do número de entidades para atender às restrições do modelo e da API

#### Em uma aplicação real:

 Quebra do texto em segmentos, sob pena de realização de várias chamadas

# Conversão do dataset no formato esperado pelo modelo

- Doccano 

   texto + entidades agrupadas por tipo, sem índices
- Ref.: <a href="https://sourajit16-02-93.medium.com/zero-shot-named-entity-recognition-using-openai-chatgpt-api-46738191f375">https://sourajit16-02-93.medium.com/zero-shot-named-entity-recognition-using-openai-chatgpt-api-46738191f375</a>

# Préprocessamento

# Exemplo

- Seleção de 2 documentos (restrição de tamanho para exemplos)
- A API assume um formato que emula um "diálogo":

Few-shot com instrução:

HEADER\_PROMPT = """Diga quais entidades são do tipo PESSOA, do tipo INSTITUIÇÃO
PÚBLICA, do tipo ORGANIZAÇÃO PRIVADA ou do tipo LOCAL."""

#### Resultados

Few-shot com instrução:

```
Recall = 0.5741603334150527

Precision = 0.7095980311730927

F1 = 0.6347347809487763
```

Few-shot sem instrução:

```
Recall = 0.5741603334150527

Precision = 0.7095980311730927

F1 = 0.6347347809487763
```

Zero-shot (apenas instrução)

Recall = 0.481000245158127

'Entidades do tipo INSTITUIÇÃO PÚBLICA: Ministério Público do Estado de Mato Grosso (MPMT), Prefeitura Municipal de Rondonópolis. \n\nEntidades do tipo PESSOA: José Carlos Junqueira de Araújo (prefeito), Izalba Diva de Albuquerque (secre tária municipal de Saúde), outros dois servidores públicos, empresários proprie tários das empresas envolvidas no contrato para aquisição de 22 ventiladores pu lmonares, Jesus de Oliveira Vieira de Souza (suposto proprietário da empresa Life Med Comércio de Produtos Hospitalares e Medicamentos Eireli), Ramos de Faria e Silva Filho (real proprietário da empresa Life Med Comércio de Produtos Hospitalares e Medicamentos Eireli).\n\nEntidades do tipo ORGANIZAÇÃO PRIVADA: empre sas envolvidas no contrato para aquisição de 22 ventiladores pulmonares (Life Med Comércio de Produtos Hospitalares e Medicamentos Eireli e outras empresas co tadas na suposta cotação de preços apresentada). \n\nEntidades do tipo LOCAL: Não há informações suficientes na matéria para determ...'

# Exemplos de saída (zero-shot learning)

- O ChatGPT não somente retorna entidades, como também contextualiza.
- Não é possível medir facilmente a precisão da saída:
  - Como identificar o que ele apontou como "entidade" dentro do texto de saída?

### Resultado inesperado:

- O ChatGPT faz automaticamente correção ortográfica ao retornar as entidades!
  - Impacto na coleta das métricas

Modo zero-shot demonstra mais "inteligência" (contextualização)

# Análise Qualitativa

# Tópico avançado



- Até que ponto ele apenas "prevê a próxima palavra"?
  - Haveria alguma etapa de parsing ou "pos-tagging", por exemplo?
  - Múltiplas inferências pelo modelo etc.