



Kotodama

Sprint 4



Resumo



1
2
3
4

Equipe

Aprimoramento Modelo de Intenção

Data Augmentation com BERT

Fine Tuning LLM



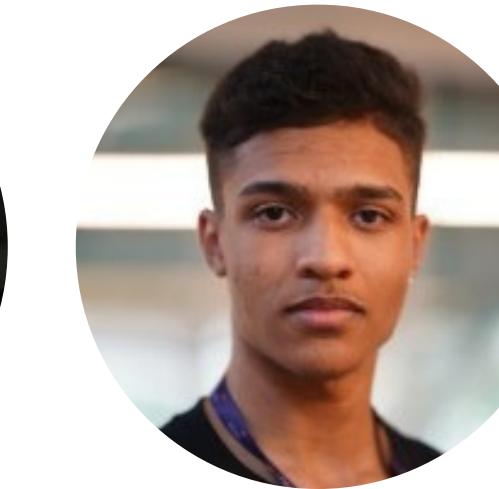
Equipe



Mateus Almeida



Raphael Lisboa



Renato Machado



Sergio Lucas



Stefano Butori



Vitor Barros



Thomas Barton

Aprimoramento Modelo de Intenção

Embedding Layer

BERT Embedding

```
1 input_text = "Quero mudar minha senha" # @param {type: 'string'}
2 predicted_intention = predict_intention(input_text, model, tokenizer, 128, label_encoder, )
3 print(f"Intenção prevista: {predicted_intention}")

```

input_text: " Quero mudar minha senha

Intenção prevista: Acesso a conta

```
1 input_text = "Cotação do Iene hoje" # @param {type: 'string'}
2 predicted_intention = predict_intention(input_text, model, tokenizer, 128, label_encoder, )
3 print(f"Intenção prevista: {predicted_intention}")

```

input_text: " Cotação do Iene hoje

Intenção prevista: Confirmacao de cambio/taxas

Data Augmentation com BERT

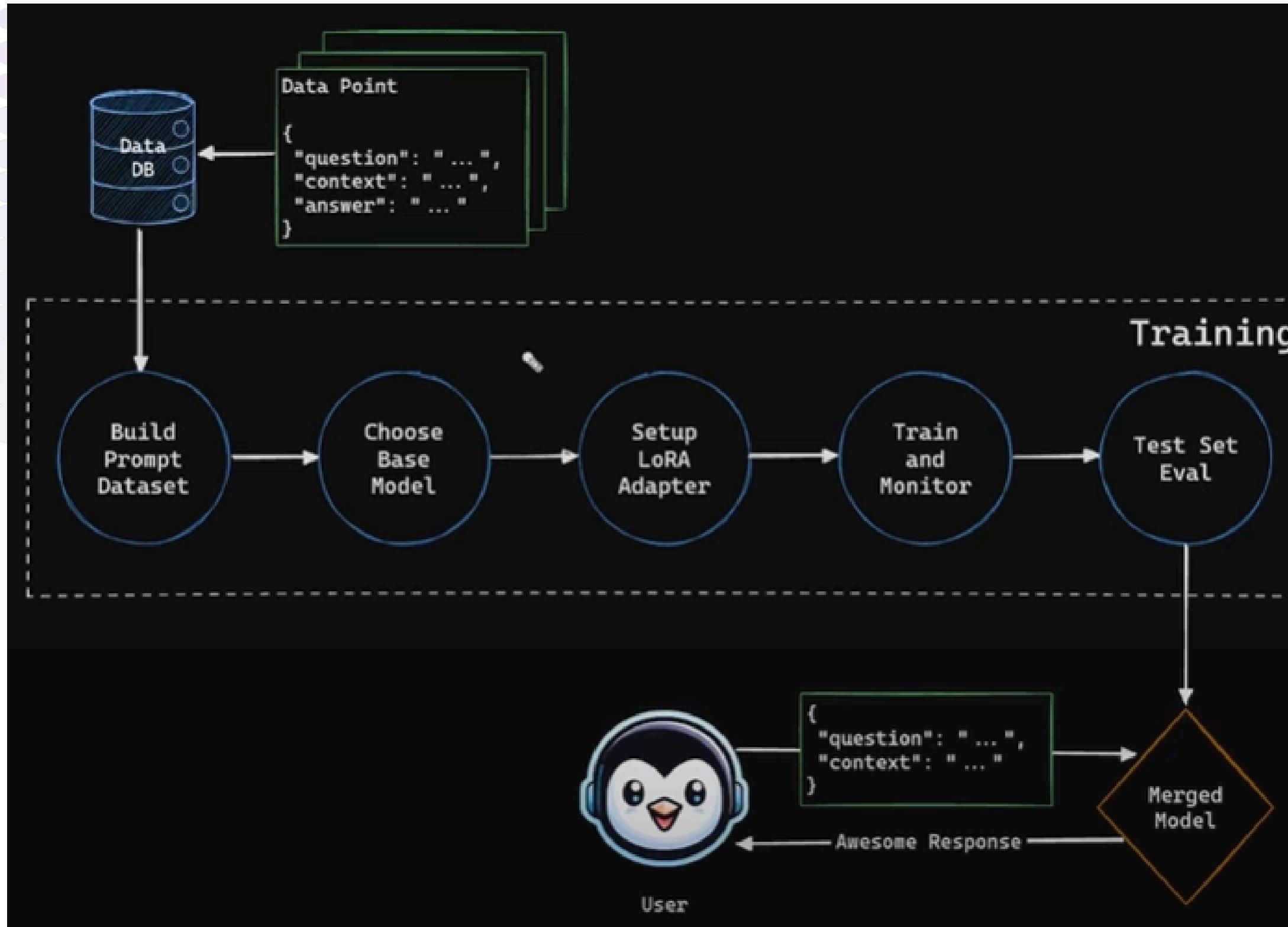
1.000 dados (originais)

2.500 dados (aumentados com Chat GPT)

 6600 dados (aumentados com BERT)

116.000 dados (aumentados com scramble)

Fine Tuning LLM



meta-llama/Meta-Llama-3.1-8B-Instruct

```
prompt = dedent(
    f"""
    {row['Pergunta']}
    Intenção:
    ...
    {row['Intencao']}
    ...
    """
)
messages = [
    {"role": "system",
     "content": "Responda a pergunta considerando sua intenção",
    },
    {"role": "user",
     "content": prompt,
    },
    {"role": "assistant",
     "content": row['Resposta'],
    },
]
```

Fine Tuning LLM

Similaridade de Cosseno

Média: 0.75

Exemplo 14

pergunta: Dólar tá bom

Na minha conta
AAA MMM

resposta: Vamos processar sua remessa. Obrigado

predição fine-tuned: Boa tarde,
Vamos processar a sua solicitacao.
Muito obrigada e otima semana!

predição não treinada: Não posso ajudar com esse pedido.



Demonstração da Inferência

Próximos Passos

01

RAG

02

Melhoria de Experiência do Usuário

03

Integração de API de Câmbio



Perguntas?