

## Checkpoint 1 - Segundo semestre

Integrantes:

Henrico Nardelli Bela - Rm 95985

Daniel Faria - Rm 94026

Emilly Gabrielly - Rm 94437

Data da divulgação: 04/09/2023

Data da entrega: 22/09/2023

Esse Checkpoint deverá ser entregue em grupo de 3 indivíduos, e se possível ser mantido para os demais checkpoints até o fim do ano.

O checkpoint deve ser entregue em formato digital, preferencialmente .pdf

Para esse checkpoint será avaliado a capacidade de traduzir uma ideia de produto digital em uma estrutura prática para implementação.

Q1) O objetivo da disciplina neste segundo semestre é criar um GPT/LLM que seja funcional e utilize os conceitos aprendidos em aula sobre NLP. Para guiar a produção de vocês, responda as questões abaixo de forma a guiar a implementação

**a) Qual o nome do GPT/LLM?**

R: ResumeAI

**b) Qual o objetivo do GPT/LLM?**

R: Resumir textos

**c) Qual será a fonte de treino?**

R: Dataset do HuggingFace - "xsum" (v1.2.0)

**d) Qual arquitetura de modelo será utilizada?**

R: Modelo pré treinado do HuggingFace (T5) - Transformers - Pytorch

**e) O que o GPT do seu grupo não vai fazer?**

R: Tudo que não seja resumir textos/livros

Q2) Crie uma visão esquemática (end-to-end) do seu produto GPT/LLM, que contemple desde o input ao delivery. Passando por tratamento de dados, treino, backtest, métrica de performance. (Próxima página!)

## Início

### --- Aquisição de Dados

- |
- | --- Uso da biblioteca Datasets HuggingFace - "xsum"
- | --- Uso da Pipeline para processamento dos dados
- |

### --- Treinamento Inicial do Modelo

- |
- | --- Modelo pré treinado T5
- | --- Fine-Tuning com uso da Pipeline
- |

### --- Backtesting

- |
- | --- Avaliação de desempenho
- | |
- | | --- Métricas (F1 - Score, BLEU)
- |

### --- Implementação do Produto

- |
- | --- Desenvolvimento da Interface de Usuário (Streamlit)
- | --- Integração do Modelo
- | --- Conexão com Sistemas Externos
- |

### --- Pipeline de Processamento de Dados em Tempo Real

- |
- | --- Configuração do Pipeline
- | --- Pré-processamento em Tempo Real
- |

### --- Uso do Modelo em Tempo Real

- |
- | --- Aplicação em Tempo Real
- | --- Respostas Geradas
- |

### --- Monitoramento de Desempenho

- |
- | --- Métricas em Tempo Real
- | --- Análise de Desempenho
- |

### --- Treinamento Contínuo

- |
- | --- Atualização Periódica com Novos Dados
- | --- Fine-tuning Regular
- |

### --- Implantação e Escalabilidade

- |
- | --- Implantação em Servidores Escaláveis
- | --- Balanceamento de Carga
- | --- Monitoramento de Recursos
- |

### --- Entrega e Atendimento ao Cliente

- |
- | --- Entrega de Respostas/Produtos
- | --- Suporte ao Cliente
- |

## Fim

