**Linguagem Natural e LLMs [25E2-25E2]**

**Aluna:** Rachel Reuters

**Fundamentos das LLMs**

1. Explique os seguintes conceitos fundamentais dos LLMs, fornecendo exemplos práticos e diagramas onde for relevante:

* Pre-training: Primeira etapa em que o modelo aprende sobre a linguagem humana de forma não supervisionada (sem classificação previa). O objetivo ‘e ter uma compreensão básica de gramática, semântica e fatos. Exemplo: O modelo recebe a frase "O céu é..." e sua tarefa é prever a próxima palavra mais provável, que seria "azul". Ao fazer isso bilhões de vezes, ele aprende as relações estatísticas entre as palavras.
* Transfer Learning: estratégia de transferir o conhecimento obtido na fase de pre-treinamento para uma nova tarefa mais específica. Em vez de treinar um modelo do zero para cada problema, aproveita-se a base de conhecimento geral já existente. Exemplo: diagnosticar se uma tomografia do útero representa sinais de câncer ou não. A primeira etapa seria o treinamento de tomografia de úteros, sabendo identificar se representa ou não um útero. Na etapa de transfer learning seria utilizar esse modelo e aplicar o conhecimento de câncer através da imagem de tomografia.
* Embeddings: são representações numéricas de palavras, frases ou textos inteiros em um espaço vetorial. Quanto maior semelhança mais próximos serão os vetores nesse espaço. Exemplo: representações vetoriais entre casa e palácio como pode ser visto no diagrama [Embedding projector - visualization of high-dimensional data](https://projector.tensorflow.org/).
* Transformers: Arquitetura projetada para processar dados sequenciais, como texto, e entender o contexto e as relações entre todas as partes dessa sequência. Ele possui um mecanismo de auto atenção que olha para todas as palavras da frase em paralelo e determina as mais importantes para entender o significado de cada palavra individualmente. Internamente ele funciona com encoders e decoders. Exemplo:

Diagrama

O conteúdo gerado por IA pode estar incorreto.

* Attention: A técnica que permite que o modelo pese a importância de diferentes palavras na sequência de entrada ao processar uma palavra específica. Ele aprende a prestar atenção nas palavras mais relevantes para entender o contexto, não importando o quão distantes elas estejam na frase. Exemplo: Identificar o que o “a” representa na frase O jogador pegou a bola e a jogou.

Diagrama

O conteúdo gerado por IA pode estar incorreto.

* Fine-Tunning: Segunda fase do Transfer Learning. Após o pré-treinamento, o modelo de conhecimento geral é adaptado para uma tarefa específica. Isso é feito treinando-o um pouco mais com um conjunto de dados menor e rotulado, específico para o problema que se quer resolver.

**Quizzes do Curso de NLP da Hugging Face**

2) Acesse os quizzes dos capítulos 1, 2 e 3 do curso de NLP da Hugging Face através do link: Curso de NLP.

Resolva os quizzes e capture screenshots dos resultados.

Anexe as screenshots a esta avaliação e explique brevemente os conceitos abordados em cada quiz.

Parte 1) O foco principal dessa primeira parte foi apresentar o Hugging face, explicar a diferença entre LLM e NLP, explicar como funciona a biblioteca Transformers, e também apresentar as arquiteturas possíveis da mesma (encoder, decoder e misto).

Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Interface gráfica do usuário, Texto, Aplicativo

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Interface gráfica do usuário, Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Parte 2) Na segunda parte já entra mais no detalhe de implementação da biblioteca Transformers, explica a diferença entre usar com Pytorch e Tensorflow, mostra alguns casos de uso, explica sobre os Automodels, o processo de tokenização que pode ser aplicado utilizando o modulo AutoTokenizer, explica algumas técnicas para evitar que o modelo quebre devido a textos muito longos.

Texto

O conteúdo gerado por IA pode estar incorreto.

Interface gráfica do usuário, Texto

O conteúdo gerado por IA pode estar incorreto.

Interface gráfica do usuário, Site

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Interface gráfica do usuário, Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Parte 3) Esse modulo foi mais focado em explicar como fazer o fine-tuning, que pega um modelo pre- treinado e especializa para a necessidade desejada. Ensina como utilizar o modulo Trainer e a biblioteca Accelerate.

Interface gráfica do usuário, Texto

O conteúdo gerado por IA pode estar incorreto.

Interface gráfica do usuário, Texto, Aplicativo

O conteúdo gerado por IA pode estar incorreto.

Interface gráfica do usuário, Texto, Aplicativo

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

Interface gráfica do usuário, Texto

O conteúdo gerado por IA pode estar incorreto.

Texto

O conteúdo gerado por IA pode estar incorreto.

**Análise de Dados com NER**

3) Baixe o conjunto de dados de notícias disponível em: [Folha UOL News Dataset](https://www.kaggle.com/datasets/marlesson/news-of-the-site-folhauol)..

Utilize o modelo <https://huggingface.co/monilouise/ner_news_portuguese> para identificar e extrair entidades mencionadas nas notícias.Crie um ranking das organizações que mais apareceram na seção "Mercado" no primeiro trimestre de 2015.Apresente os resultados em um relatório detalhado, incluindo a metodologia utilizada e visualizações para apoiar a análise.

Tive algumas dificuldades com os resultados, pois a partir da tokenização, apareceram muitos resultados com caractere ### que pode indicar que a estratégia de agregação do modelo em questão não foi capaz de entender como uma palavra somente. Tentei utilizar outros métodos de aggregation\_strategy como max, average e simple, em todas elas continuei obtendo resultados com ##, como pode ser visto abaixo.

Porem nota-se que Brad e ###esco era para ser uma única palavra.

Algumas outras ficam muito difíceis de entender como ##er, ##u , ##i, S, O.

Outro tratamento pós processamento foi de ignorar organização Folha, pois é o jornal que está emitindo a mensagem, então acaba que o aparecimento fica bem obvio e não gera novos insights pra a análise.

Inferências :

**Setor Financeiro e Bancário (mais presente):**

* Brad : Provavelmente se refere ao Bradesco.
* Itaú, Ita : Itaú Unibanco.
* BTG, Pactual : BTG Pactual.
* Unibanco : Banco que se fundiu com o Itaú.
* Santa : Provavelmente se refere ao banco Santander.
* Moody's : Uma das maiores agências de classificação de risco de crédito.
* Investimentos : A própria palavra "investimentos".
* Brasil : Banco do Brasil

**Mercado de Ações e Empresas Brasileiras:**

* FBovespa : se refere ao Ibovespa, o principal índice da bolsa de valores brasileira.
* BM & : Provavelmente BM&F (Bolsa de Mercadorias e Futuros), que faz parte da B3, a bolsa do Brasil.
* Vale : Uma das maiores empresas da bolsa brasileira.
* Galvão : possivelmente a Galvão Engenharia, uma construtora.

**Setor Corporativo Geral:**

* Google : Gigante da tecnologia.
* Volks : Volkswagen.

O texto provavelmente fala sobre o desempenho do mercado de ações brasileiro (Ibovespa), mencionando a performance das principais empresas listadas, como Vale, e o papel dos grandes bancos (Bradesco, Itaú, BTG Pactual). Também pode ser um relatório de uma corretora avaliando o cenário de investimentos no Brasil, recomendando ações e discutindo o risco-país (mencionado pela presença da Moody's).

Interface gráfica do usuário

O conteúdo gerado por IA pode estar incorreto.

Realizando um filtro para remover esses itens que não são facilmente compreensíveis:

Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.

Uma imagem contendo Logotipo

O conteúdo gerado por IA pode estar incorreto.

**Engenharia de Prompts**

4) Analise os seguintes prompts e identifique por que eles poderiam gerar respostas insatisfatórias ou irrelevantes. Reformule cada prompt utilizando técnicas de engenharia de prompts para torná-los mais específicos e direcionados. Explique as melhorias feitas em cada caso e os motivos por trás das reformulações:

* Exemplo 1: "Escreva sobre cachorros."

Esse prompt é extremamente genérico. Não se sabe exatamente o que o usuário deseja, que tipo de informação. A resposta disso seria algo bem genérico igualmente. Utilizando um método de engenharia de prompt “Self-Ask” : "Eu quero que você escreva sobre cachorros. Tentando responder algumas perguntas:

Quais espécies são as mais queridas pelos humanos?

Quais espécies são as mais raras de se encontrar?

Porque cachorros são considerados os melhores amigos do homem?

Depois de responder a essas perguntas, use as respostas para gerar um artigo bem estruturado sobre cachorros."

Decompõe em perguntas, delimitando a resposta gerada, além disso guia o “raciocínio” do algoritmo.

* Exemplo 2: "Explique física."

Assim como o anterior, é genérico. Física é um conceito muito amplo que pode abordar uma infinidade de coisas. Utilizando um método de engenharia de prompt “Zero-shot” : “Crie uma tabela comparativa que explique as principais diferenças entre a Física Clássica e a Física Quântica. A tabela deve ter três colunas: 'Conceito', 'Física Clássica' e 'Física Quântica'. Compare pelo menos três pontos fundamentais, como 'Escala de Aplicação' (macroscópico vs. subatômico), 'Natureza do Resultado' (determinístico vs. probabilístico) e 'Visão sobre a Energia'.” Com isso, foi definido o formato, criou um tipo de limite de conteúdo e forneceu critérios.

5) O prompt "Descreva a história da internet." foi mal formulado. Aplique técnicas de engenharia de prompts para melhorá-lo. Reformule o prompt para melhorar a especificidade e a qualidade da resposta. Justifique as mudanças feitas e explique como elas contribuem para obter uma resposta mais eficaz e relevante.

Utilizando a técnica "Least-To-Most" dividindo o problema em uma série de problemas mais simples e guiar o modelo para resolv Descreva a história da internet em etapas, do conceito mais simples ao mais complexo. Por favor, responda cada pergunta em ordem.

1. A Origem Militar: foque no seu propósito original durante a Guerra Fria e por que sua arquitetura era tão inovadora.

2. A Transição para o Meio Acadêmico: descreva como essa tecnologia evoluiu. Explique como foi adaptada a rede para conectar universidades e centros de pesquisa.

3. A Explosão para o Público: cite duas inovações que tornaram a internet acessível ao público geral nos anos 90, explicando qual foi o ponto de virada.

4. A Conclusão: Resuma como a base militar, a expansão acadêmica e a revolução de usabilidade da WWW e dos navegadores se combinaram para formar a internet que conhecemos hoje.er em sequência.

Essa técnica dividiu a história da internet em fases lógicas, tornando as perguntas mais gerenciáveis. Cada etapa construindo um contexto progressivo. Isso melhora a qualidade e profundidade da resposta, evitando explicações superficiais e permitindo uma narrativa estruturada. Também fica mais fácil controlar os tópicos mais relevantes de acordo com a necessidade final.

6) Aplique a técnica de Chain of Thought (CoT) para melhorar o prompt "Explique como funciona a energia solar.", detalhando o raciocínio necessário para que o modelo forneça uma resposta completa e coerente. Explique como a aplicação da técnica CoT melhora a resposta do modelo.

"Explique como funciona a energia solar, seguindo esta cadeia de pensamento:

O Princípio Físico: explique como funciona a corrente elétrica, os elétrons e fótons.

Explique como são feitos os painéis solares

Aplicando o princípio físico nos painéis solares: como os painéis solares conseguem interagir com os fótons e como que funciona a liberação de energia nesses painéis.

Por fim, como que a energia é transmitida para o consumidor final depois de passar dos painéis solares.”

Utilizando a CoT, é possível garantir que todos os estágios do processo sejam abordados, desde a física fundamental até a aplicação prática em uma residência, resultando em uma resposta muito mais completa.

**Projeto Prático com Streamlit, LLM e LangChain**

7) Escolha uma aplicação para desenvolver utilizando Streamlit, LLM e LangChain. Crie um aplicativo interativo que demonstre o uso de LLMs para resolver um problema específico.

Descreva a aplicação escolhida e os objetivos principais do projeto.

Explique a arquitetura do aplicativo, incluindo como o Streamlit, LLM e LangChain são utilizados.

Implemente o aplicativo e forneça o código-fonte, junto com instruções para execução.

Apresente evidências e exemplos de uso do aplicativo e discuta os resultados obtidos.