

Então estruture o pipeline/escopo do projeto de maneira detalhada igual você fez ali em cima, de modo que ele seja um paper forte para o PKDD. Estou pensando em algo assim. Eu tenho os prompts. Posso processá-los usando uma Sentiment Analysis com BERT e classificá-los em positivo, negativo, neutro. Suponho que eles tendam para neutro-positivo. Metrificar isso. Fazer uma análise estatística desses dados. Depois promover as modificações nos dados de duas formas: primeiro omitindo os adjetivos. E depois reescrevendo as frases para que fiquem negativas e/ou só substituir o antônimo do adjetivo. (Essa parte de quais modificações fazer deixo a seu critério). Reproduzir as medidas que fiz no início com a base modificada. Fazer uma análise das notas atribuídas nas imagens originais e depois que elas foram regeradas usando o deepseek e os novos prompts. Comparar tudo.

O ChatGPT disse:

Aqui está um pipeline detalhado e estruturado para o seu projeto, levando em consideração o que você descreveu. Ele está pensado para gerar resultados sólidos para um paper no **PKDD**, focando na análise de viés e manipulação de prompts para modelos generativos de imagens.

1. Definição do Problema e Objetivo

O objetivo deste trabalho é investigar como diferentes tipos de modificações em prompts afetam a geração de imagens por modelos generativos, em particular a partir da perspectiva de viés de IA. Em específico, o estudo se foca na troca de adjetivos positivos por negativos, e como isso impacta a geração de imagens.

O foco principal será:

- **Identificar o viés de modelos generativos** ao manipular os prompts de entrada.
- **Analisar as diferenças nas imagens geradas** em função das modificações realizadas nos prompts.

2. Coleta e Preparação de Dados

2.1 Coleta de Prompts

- Obtenha um conjunto de **prompts originais** usados para gerar imagens. Estes podem ser extraídos de um repositório de prompts já existentes ou de uma base de dados de exemplos de prompts para modelos generativos, ou criados por você para o estudo.

2.2 Processamento Inicial dos Prompts

- **Análise de Sentimentos com BERT:**
 - Aplique uma análise de sentimentos sobre os prompts originais usando um modelo **BERT** pré-treinado para **classificar os prompts** em **positivo, negativo e neutro**.
 - Calcule uma **pontuação média de sentimentos** para o conjunto de prompts. Isso ajudará a entender o viés inicial dos prompts.
 - A análise estatística pode ser feita para avaliar a **distribuição** das categorias de sentimento e a **variabilidade** entre eles.

2.3 Estatísticas Descritivas

- **Distribuição de Sentimentos:** Analise a **distribuição de sentimentos** (positivo, negativo, neutro) nos prompts originais.
 - Quantifique a **proporção de sentimentos** para avaliar se os prompts tendem para neutro-positivo, como você sugeriu.
 - Utilize **testes estatísticos** para comparar as médias das pontuações de sentimento entre os prompts de diferentes categorias.

3. Modificação dos Prompts

Após o processamento inicial dos prompts, você fará duas modificações principais nos dados de entrada:

3.1 Modificação 1: Omissão de Adjetivos

- **Objetivo:** Remover os adjetivos dos prompts para observar como a **ausência de qualificadores** (como “bom”, “feio”, “bonito”) afeta a geração de imagens.
- **Processo:**
 - Realize a remoção dos **adjetivos** em cada prompt utilizando um **tagger linguístico** (ex: **spaCy**) ou outra ferramenta de processamento de linguagem natural (NLP).
 - Crie uma nova versão do prompt com adjetivos omitidos.
 - Avalie se essa modificação altera a **neutralidade** do prompt, possivelmente gerando imagens mais **neutras** ou **ambíguas**.

3.2 Modificação 2: Transformação para Sentimento Negativo

- **Objetivo:** Substituir os adjetivos positivos por seus **antônimos negativos**, ou reescrever os prompts de forma que se tornem explicitamente **negativos**.
- **Processo:**
 - Para **adjetivos positivos**, utilize um dicionário de antônimos para substituir por adjetivos negativos (ex: "bonito" → "feio").
 - Para **prompts neutros** ou **positivos**, reescreva o conteúdo de forma a tornar as descrições **explicitamente negativas** ou **desfavoráveis**.
 - Para **testar diferentes abordagens**, além de substituir por antônimos, experimente fazer alterações contextuais que mudem a tonalidade do prompt (ex: "A imagem deve ser vibrante e alegre" → "A imagem deve ser sombria e monótona").

3.3 Geração de Imagens

- Utilize o modelo **DeepSeek** (ou outra alternativa como DALL-E, MidJourney, ou Stable Diffusion) para gerar imagens baseadas nos prompts modificados e nos prompts originais.
 - Se possível, use **diferentes modelos generativos** para comparar os efeitos da modificação de prompt.

- Armazene as **imagens geradas** para análise futura.

4. Avaliação das Imagens Geradas

4.1 Métricas de Avaliação Inicial (Imagens Originais)

- Gere **imagens a partir dos prompts originais**.
- Use métricas de **qualidade de imagem** (por exemplo, **Inception Score (IS)**, **Frechet Inception Distance (FID)**) para avaliar a **qualidade visual das imagens** originais.
- Armazene essas **pontuações de qualidade** como base para comparação.

4.2 Geração de Imagens com Prompts Modificados

- Gere **novas imagens** a partir dos prompts com adjetivos omitidos e com sentimentos negativos.
- Use as **mesmas métricas de avaliação** (IS, FID) para avaliar as imagens geradas a partir dos novos prompts.

4.3 Comparação das Imagens

- Compare as **imagens originais** e as **imagens modificadas** com base nas métricas de **qualidade**.
 - **Teste estatístico** (ex: **teste t de Student**) para comparar as médias das pontuações de qualidade entre as imagens originais e modificadas.
 - Avaliação de **mudanças significativas** no estilo, composição ou qualidade visual da imagem com base na modificação dos prompts.
 - Realização de uma análise **qualitativa** se possível, por meio da comparação visual das imagens geradas.

5. Análise Estatística

5.1 Comparação de Sentimentos e Imagens Geradas

- Realize uma **análise estatística** para comparar os efeitos dos diferentes tipos de modificações nos **prompts**:
 - **Testes de hipótese** para avaliar se as modificações nos prompts (omissão de adjetivos, reescrita para negativos) causam uma mudança significativa nas imagens geradas.
 - Avalie as diferenças nos **scores de qualidade** de imagem (IS, FID) antes e depois das modificações.
 - Compare a **distribuição de sentimentos** dos prompts originais e modificados e como essas modificações impactam os **scores de imagem**.

5.2 Análise do Viés

- Investigue se as modificações nos prompts têm um **efeito de viés** nas imagens geradas, como:

- **Mudança na estética** das imagens (ex: mudança de cores, formas, composição).
- **Mudança no conteúdo** das imagens (ex: uma imagem alegre se tornando sombria ou negativa).
- Análise do **viés cultural** ou de **estereótipos** nas imagens com modificações negativas.

6. Conclusões e Implicações

- **Interpretação dos resultados:** Conclua sobre os efeitos de diferentes tipos de modificação de prompts em modelos gerativos.
 - Como a **ausência de adjetivos** ou a **alteração para termos negativos** impactam as **imagens geradas**?
 - O que isso diz sobre o **viés dos modelos** e sua resposta a diferentes tipos de modificações em textos?
- **Implicações para práticas de IA generativa:** Como essas descobertas podem influenciar o uso de IA para geração de imagens em diferentes contextos (por exemplo, em criatividade computacional, design ou arte)?

Esse pipeline permitirá que você desenvolva um **paper forte para o PKDD**, com uma metodologia clara e uma análise rigorosa dos efeitos da modificação dos prompts na geração de imagens. As análises estatísticas e qualitativas fornecem a profundidade necessária para que o trabalho tenha impacto e relevância no campo de **Data Mining e Knowledge Discovery**.