

## Modelos Generativos

### Variational AutoEncoder

O Variational AutoEncoder (VAE) é um modelo de redes neurais generativas que gera dados em vez de classificá-los. Utilizando técnicas como convolução, pooling e flattening, o VAE comprime imagens em um espaço representado por um ponto  $Z$ , e então reconstrói as imagens a partir desse ponto. Durante o treinamento, o VAE busca minimizar duas funções de perda: uma que avalia as diferenças entre as imagens reconstruídas e as originais, e outra que modela as distribuições probabilísticas. Esse processo permite ao VAE gerar novas amostras que se assemelham às do conjunto de treinamento, possibilitando a geração de novos dados realistas.

### Generative Adversarial Networks

As Generative Adversarial Networks (GANs) são compostas por duas redes neurais em competição: uma geradora e outra discriminadora. A rede geradora produz imagens a partir de dados latentes, enquanto a discriminadora tenta distinguir entre imagens reais e geradas. Ao longo do treinamento, as redes evoluem em um processo adversarial, onde a rede geradora busca gerar imagens cada vez mais convincentes, enquanto a discriminadora aprimora sua capacidade de discernir entre imagens reais e falsas. Esse ciclo iterativo continua até que as imagens geradas se tornem indistinguíveis das reais. No entanto, desafios na classificação podem levar a instabilidades nesse processo, exigindo técnicas adicionais de regularização e controle para garantir a convergência adequada do modelo.

### Recriando o Chat GPT

O vídeo a respeito da recriação do modelo Chat GPT utiliza métodos distintos na criação de IA generativas em comparação aos mencionados anteriormente. O projeto mostrado no vídeo não faz uso de redes neurais; consequentemente, o modelo não se enquadra nem como GAN nem como VAE. O modelo desenvolvido tem como objetivo prever o próximo token com base nos dados do conjunto de dados de treinamento. Em outras palavras, após cada token, uma série de probabilidades é considerada para determinar o próximo token a ser gerado.

No entanto, este modelo não adota algumas das práticas comuns em Processamento de Linguagem Natural (NLP) que foram abordadas em momentos anteriores do bootcamp. Portanto, ele representa uma abordagem inovadora na criação de modelos generativos e na análise de linguagem.