

vizinho mais próximo tem capturado a atenção para problemas com dados que possuem alta dimensionalidade (Wang; Shen, 2022).

### 2.3.1 Banco de Dados Vetorial Para Taipalus e Toni (2024),

um sistema de gerenciamento de banco de dados vetorial é um tipo especializado de sistema de gerenciamento de banco de dados que se concentra principalmente no gerenciamento eficiente de dados vetoriais de alta dimensão. Similarmente a outros tipos de sistemas de gerenciamento de banco de dados, espera-se que um SGBD direcionado a aplicação vetorial, seja um software funcional que possa gerenciar dados. O gerenciamento de dados inclui em geral a consulta e manipulação de dados, coleta de metadados, indexação, controle de acesso, backups, suporte para escalabilidade e interfaces com outros sistemas, como drivers de banco de dados, linguagens de programação, frameworks e sistemas operacionais. O surgimento de LLMs para tarefas como recuperação de informações, juntamente com o crescimento de dados não drivers de banco de dados, linguagens de programação, frameworks e sistemas operacionais. O surgimento de LLMs para tarefas como recuperação de informações, juntamente com o crescimento de dados não estruturados que sustentam impulsionadores econômicos como comércio eletrônico e plataformas de recomendação, demanda novos sistemas de gerenciamento de banco de dados que possam oferecer capacidades tradicionais, como otimização de consultas,22 transações, escalabilidade, tolerância a falhas e privacidade e segurança, mas para dados não estruturados (Pan; Jie; Wang; Li, 2023). Neste sentido, surgem os bancos de dados vetoriais que possibilitam o armazenamento e recuperação de qualquer unidade de informação como a Figura 2 apresenta a partir da representação do banco de dados vetorial Manu ® , desde que exista uma forma de transformar um texto, uma imagem, um vídeo ou um áudio, em uma representação numérica na forma de embeddings (Guo et al ., 2022).

Figura 2 - Banco de dados vetorial um modelo superior ao juridics/bertlaw-base-portuguese-sts-scale visto que seu treinamento é baseado no projeto SBERT que contém limitações quanto quantidade de tokens de entrada quando comparado à LLMs como GPT ® e Gemini ® . O modelo pode também sofrer o ajuste fino ( fine-tuning ) a partir do conjunto de jurisprudências coletado. Mais especificamente quanto ao banco de dados vetorial, foi utilizado o ranqueamento de cosseno, disponível na Seção 4.6. Porém, com a evolução do trabalho, é possível verificar o uso de outras formas de ranqueamento como distância euclidiana, distância de manhattan e produto dos pontos.62 Mostra-se relevante também a criação de uma aplicação web ou aplicativo de recuperação de jurisprudências e geração de defesa jurídica, tendo como base o protótipo apresentado na Seção 5.3. Por fim, pensando em termos de empreendedorismo, a disponibilização para o público poderia viabilizar a coleta de diferentes feedbacks pensando nas melhorias e, em momento futuro, na