SPRINT 80/20

SEFAZ/AL

MATERIAL COMPLEMENTAR · QUESTÕES DO CEBRASPE

# Análise de Dados, Estatística e IA Generativa

Tipos de dados · Tipos de gráficos · IA generativa

Ciência de Dados · 27/11

Inteligência Artificial · 30/11



**Prof. Rogério Araújo**

@profrogeraoaraujo · @destrinchandoti

destrinchandoTI

# Fundamentos de análise de dados

# Tipos de dados

Estruturados, semiestruturados e não estruturados

Quantitativos e qualitativos

# Estruturados, semiestruturados e não estruturados

Esquema rígido

## Estruturados

Organizados em linhas e colunas, com esquema definido antes do armazenamento

Exemplos

- Tabelas de bancos relacionais
- Planilhas e arquivos CSV
- Sistemas de folha e orçamento

Esquema flexível

## Semiestruturados

Trazem marcadores (tags, chaves) que descrevem o conteúdo, mas sem estrutura fixa

Exemplos

- Arquivos JSON e XML
- XML da nota fiscal eletrônica
- E-mails e logs de sistemas

Sem esquema

## Não estruturados

Não seguem modelo predefinido; o significado precisa ser extraído do conteúdo

Exemplos

- Textos, PDFs e documentos
- Imagens, áudios e vídeos
- Postagens em redes sociais

# Comparação

| Critério | Estruturado | Semiestruturado | Não estruturado |
|---|---|---|---|
| Esquema | **Fixo**, definido antes da gravação (schema-on-write) | **Autodescritivo** e flexível, varia entre registros | **Inexistente**; estrutura aplicada na leitura (schema-on-read) |
| Formatos | Tabelas SQL, CSV, planilhas | JSON, XML, YAML, logs | Texto livre, PDF, imagem, áudio, vídeo |
| Armazenamento | SGBD relacional, data warehouse | Bancos NoSQL de documentos, data lake | Data lake, armazenamento de objetos |
| Como analisar | Consultas SQL, BI, estatística | Parsing (XPath, JSONPath), NoSQL | OCR, processamento de linguagem natural, IA |
| Esforço de preparo | Baixo | Médio | Alto |

# Dados quantitativos e qualitativos



# Pegadinhas de prova

1

## Número nem sempre é quantitativo

CPF, CEP e matrícula são rótulos. Somar ou tirar a média deles não faz sentido: são qualitativos nominais

2

## Escala de 1 a 5 é ordinal

Uma avaliação de satisfação de 1 a 5 tem ordem, mas a distância entre os níveis não é mensurável

3

## Contar é discreto, medir é contínuo

Nº de auditorias realizadas é discreta; duração de uma auditoria em horas é contínua

4

## JSON, XML e e-mail são semiestruturados

Têm marcadores que descrevem o conteúdo, mas não exigem um esquema fixo de tabela

# Questões de concursos

[CESPE/CEBRASPE 2026 Câmera dos Deputados – Técnico Legislativo – Policial Legislativo Federal] Acerca de redes e de bancos de dados, julgue o item a seguir.

- Por não possuírem um formato predefinido, dados não estruturados são mais simples de manipular que dados estruturados.

# Questões de concursos

[CESPE/CEBRASPE 2026 Câmera dos Deputados – Técnico Legislativo – Policial Legislativo Federal] Acerca de redes e de bancos de dados, julgue o item a seguir.

- Por não possuírem um formato predefinido, dados não estruturados são mais ~~**simples**~~ **trabalhosos** de manipular que dados estruturados.
  - Gabarito: **ERRADO**.

# Questões de concursos

[CESPE/CEBRASPE 2025 PC/DF – Gestor de Apoio as Atividades Policiais Civis – Especialidade: Analista de Informática: Desenvolvimento de Sistemas] Acerca de API, julgue o item subsequente.

- Em conformidade com o formato JSON, os dados de um cliente podem ser representados da forma apresentada a seguir.
- {
  - “cliente:” {
  - “nome:” “João da Silva”,
  - “cpf:” “123.456.789-10”

# Comentários

- Objeto JSON:
  - {
    - "cliente": {
      - "nome": "João da Silva",
      - "cpf": "123.456.789-10"
    - }
  - }

# Questões de concursos

[CESPE/CEBRASPE 2025 PC/DF – Gestor de Apoio as Atividades Policiais Civis – Especialidade: Analista de Informática: Desenvolvimento de Sistemas] Acerca de API, julgue o item subsequente.

- Em conformidade com o formato JSON, os dados de um cliente podem ser representados da forma apresentada a seguir.
- {
  - “cliente:” {
  - “nome:” “João da Silva”,
  - “cpf:” “123.456.789-10”
  - Gabarito: **ERRADO**.

# Questões de concursos

[CESPE/CEBRASPE 2022 TELEBRAS – Especialista em Gestão de Telecomunicações – Analista de TI] Julgue os próximos itens, acerca do mapeamento de fontes de dados e da arquitetura de business inteligence (BI).

- A coleta de dados visa obter informações estratégicas e pode ser realizada de diferentes formas. Nesse sentido: métodos de coleta de dados quantitativos são os menos indicados para previsões de longo prazo, pois são subjetivos; métodos de coleta de dados qualitativos são objetivos, pois se baseiam em experiência, julgamento, intuição, conjecturas, emoções etc.

# Questões de concursos

[CESPE/CEBRASPE 2022 TELEBRAS – Especialista em Gestão de Telecomunicações – Analista de TI] Julgue os próximos itens, acerca do mapeamento de fontes de dados e da arquitetura de business inteligence (BI).

- A coleta de dados visa obter informações estratégicas e pode ser realizada de diferentes formas. Nesse sentido: métodos de coleta de dados quantitativos são os menos indicados para previsões de longo prazo, pois são subjetivos; métodos de coleta de dados qualitativos são objetivos, pois se baseiam em experiência, julgamento, intuição, conjecturas, emoções etc.
  - Gabarito: **ERRADO**.

# Questões de concursos

[CESPE/CEBRASPE 2021 Polícia Federal – Escrivão de Polícia Federal] No que se refere aos conceitos de dados estruturados e não estruturados, julgue o próximo item.

- Os dados estruturados diferenciam-se dos dados não estruturados pela rigidez em seu formato e pelo fato de poderem ser armazenados em campos de tabelas de um banco de dados relacional.

# Questões de concursos

[CESPE/CEBRASPE 2021 Polícia Federal – Escrivão de Polícia Federal] No que se refere aos conceitos de dados estruturados e não estruturados, julgue o próximo item.

- Os dados estruturados diferenciam-se dos dados não estruturados pela rigidez em seu formato e pelo fato de poderem ser armazenados em campos de tabelas de um banco de dados relacional.
  - Gabarito: **CERTO**.

# Questões de concursos

[CESPE/CEBRASPE 2017 TCE/PE – Auditor de Controle Externo – Auditoria de Contas Públicas] Com relação a Big Data, julgue o item subsequente.

- Além de estar relacionado à grande quantidade de informações a serem analisadas, o Big Data considera o volume, a velocidade e a variedade dos dados estruturados — dos quais se conhece a estrutura de armazenamento — bem como dos não estruturados, como imagens, vídeos, áudios e documentos.

# Questões de concursos

[CESPE/CEBRASPE 2017 TCE/PE – Auditor de Controle Externo – Auditoria de Contas Públicas] Com relação a Big Data, julgue o item subsequente.

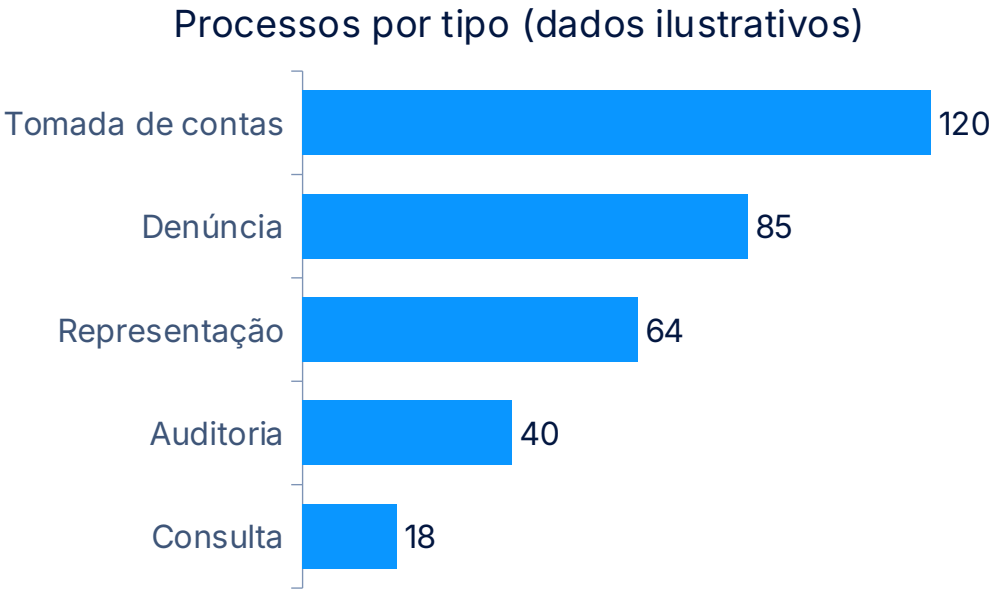
- Além de estar relacionado à grande quantidade de informações a serem analisadas, o Big Data considera o volume, a velocidade e a variedade dos dados estruturados — dos quais se conhece a estrutura de armazenamento — bem como dos não estruturados, como imagens, vídeos, áudios e documentos.
  - Gabarito: **CERTO**.

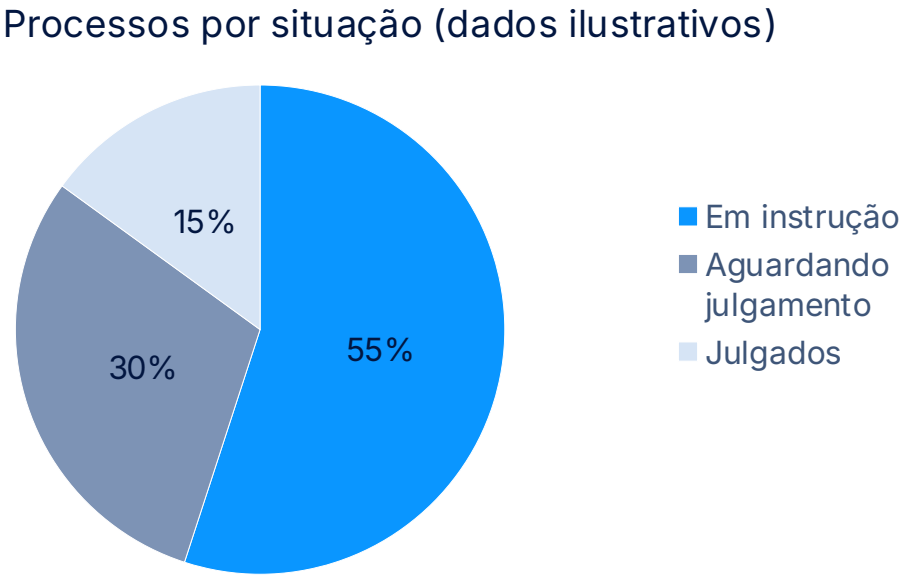
# Introdução à visualização de dados e storytelling

# Tipos de gráficos

## Barras, pizza, linha, dispersão, histograma

# Barras e pizza: comparar categorias



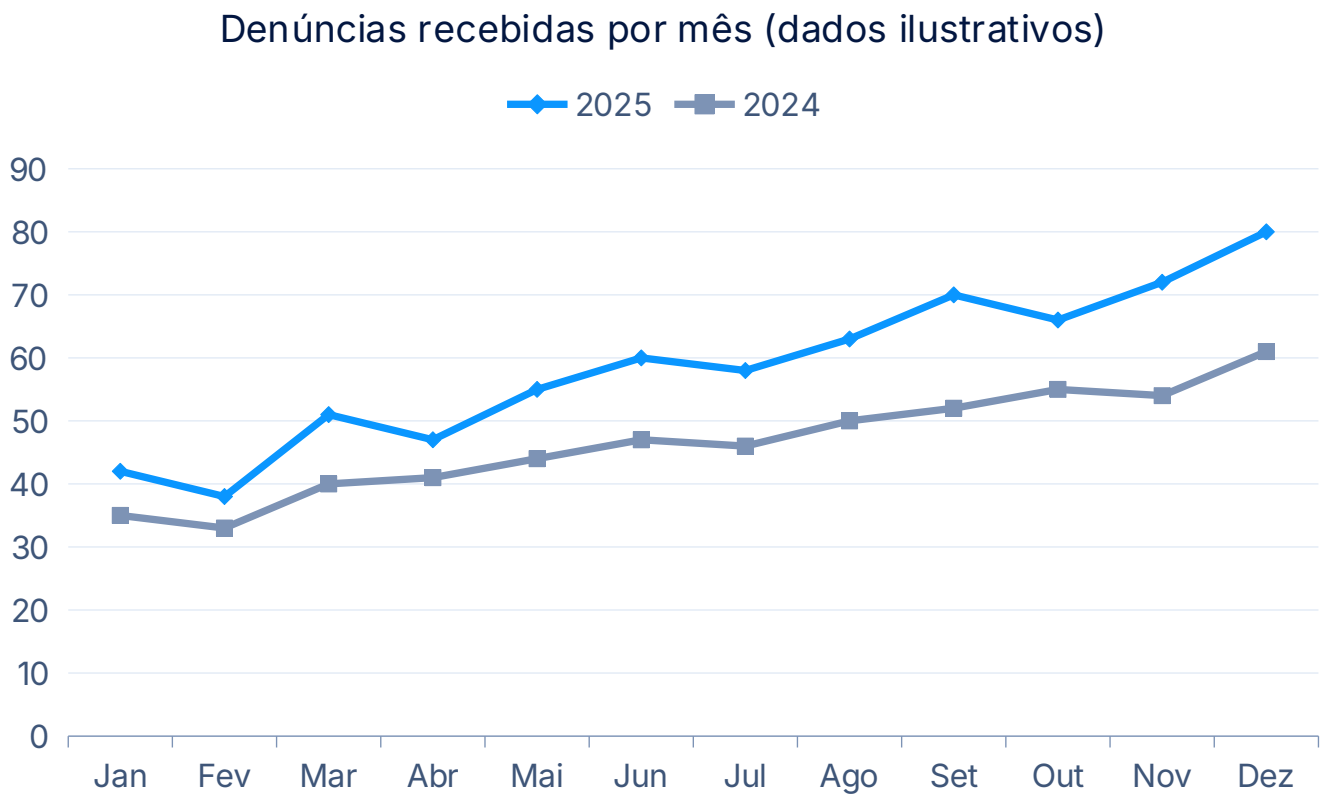

## Barras

- Objetivo: compara valores entre categorias
- Boa prática: Ordene as barras e inicie o eixo no zero

## Pizza (setores)

- Objetivo: mostra partes de um todo que somam 100%
- Boa prática: use poucas categorias

# Linha: evolução ao longo do tempo



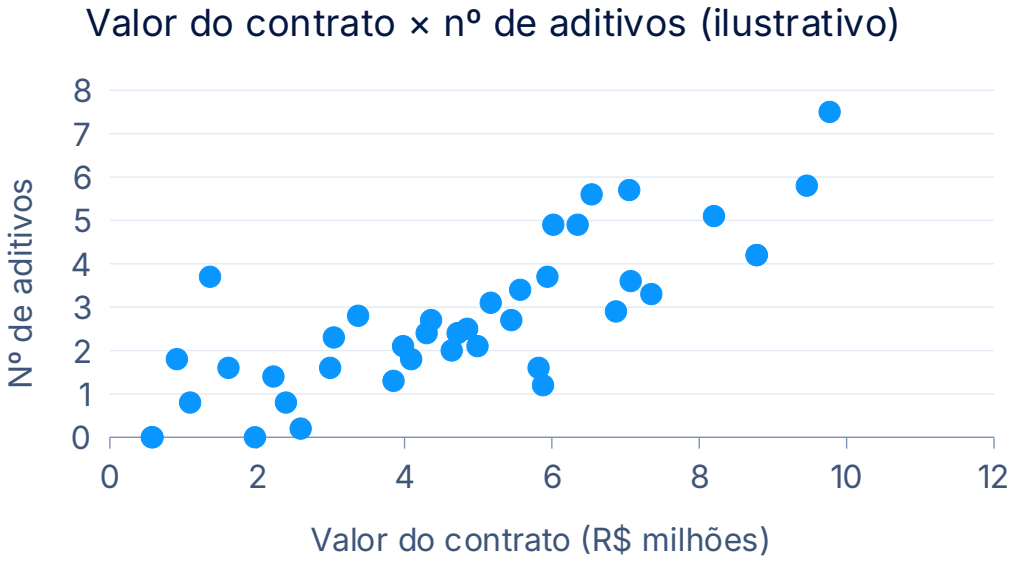

## Quando usar

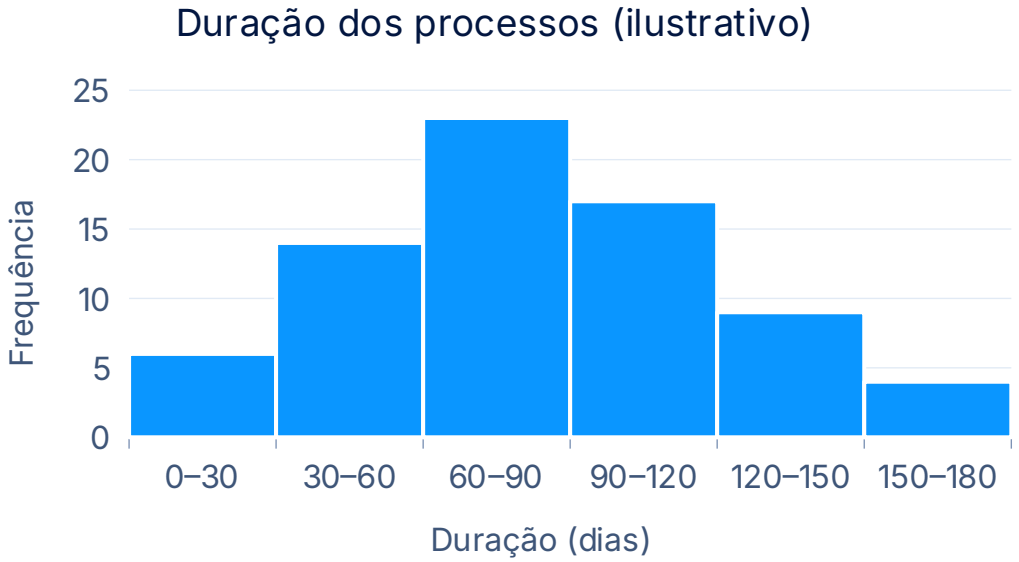
- Mostrar tendência, crescimento ou queda
- Identificar sazonalidade e picos

## Cuidados

- Eixo X temporal, em ordem e com intervalos regulares
- Não ligar categorias sem ordem
- Poucas séries por gráfico

# Dispersão e histograma: relação e distribuição




## Dispersão

- Relação entre duas variáveis quantitativas
- Revela correlação e outliers

## Histograma

- Distribuição de uma variável contínua em classes
- Barras encostadas.

# Qual gráfico usar?

| Objetivo | Gráfico | Variáveis envolvidas | Atenção |
|---|---|---|---|
| Comparar categorias | Barras | 1 qualitativa + 1 quantitativa | Ordenar as barras; eixo começa no zero |
| Mostrar composição | Pizza | Partes de um todo (100%) | Poucas categorias; evitar 3D |
| Mostrar tendência | Linha | Tempo + 1 quantitativa | Eixo temporal em ordem e regular |
| Mostrar relação | Dispersão | 2 quantitativas | Correlação não implica causalidade |
| Mostrar distribuição | Histograma | 1 quantitativa contínua | Barras encostadas; escolha das classes |

# Questões de concursos

[CESPE/CEBRASPE 2026 TCE/RN – Analista Administrativo – Especialidade: Tecnologia da Informação] No que se refere à classificação da mineração de dados, à mineração de texto, à modelagem preditiva e à visualização e análise exploratória de dados, julgue os item subsecutivo.

- Histograma é uma representação da distribuição dos dados de certa base, conforme um ou mais atributos desses dados, por meio de um gráfico de barras.

# Questões de concursos

[CESPE/CEBRASPE 2026 TCE/RN – Analista Administrativo – Especialidade: Tecnologia da Informação] No que se refere à classificação da mineração de dados, à mineração de texto, à modelagem preditiva e à visualização e análise exploratória de dados, julgue os item subsecutivo.

- Histograma é uma representação da distribuição dos dados de certa base, conforme um ou mais atributos desses dados, por meio de um gráfico de barras.
  - Gabarito: **CERTO**.

# Questões de concursos

[CESPE/CEBRASPE 2025 TRF 6ª Região – Analista Judiciário – Área: Apoio Especializado – Especialidade: Análise de Dados] A respeito de técnicas de visualização de dados, julgue o item subsequente.

- Na visualização de um conjunto complexo de dados, o uso de gráficos permite a exibição de plotagens e diagramas, organizados ao longo de dois eixos.

# Questões de concursos

[CESPE/CEBRASPE 2025 TRF 6ª Região – Analista Judiciário – Área: Apoio Especializado – Especialidade: Análise de Dados] A respeito de técnicas de visualização de dados, julgue o item subsequente.

- Na visualização de um conjunto complexo de dados, o uso de gráficos permite a exibição de plotagens e diagramas, organizados ao longo de dois eixos.
  - Gabarito: **CERTO**.

# Questões de concursos

[CESPE/CEBRASPE 2024 SEBRAE-NACIONAL – Analista Técnico II – Dados – Business Intelligence] Julgue os itens a seguir, a respeito de ferramentas de visualização de dados.

- [I] Os gráficos de linhas são os mais adequados para acompanhar a evolução de uma variável ao longo do tempo.
- [II] Os histogramas classificam os dados de acordo com o valor de várias categorias.
- [III] Os gráficos de dispersão são usados para visualizar a relação entre duas variáveis contínuas.

# Questões de concursos

[CESPE/CEBRASPE 2024 SEBRAE-NACIONAL – Analista Técnico II – Dados – Business Intelligence] Julgue os itens a seguir, a respeito de ferramentas de visualização de dados.

- **[I] Os gráficos de linhas são os mais adequados para acompanhar a evolução de uma variável ao longo do tempo.**
- [II] Os ~~**histogramas**~~ **gráficos de barras** classificam os dados de acordo com o valor de várias categorias.
- **[III] Os gráficos de dispersão são usados para visualizar a relação entre duas variáveis contínuas.**

# Questões de concursos

[CESPE/CEBRASPE 2024 SEBRAE-NACIONAL – Analista Técnico II – Dados – Business Intelligence] Assinale a opção correta.

- [A] Apenas os itens I e II estão certos.
- [B] Apenas os itens I e III estão certos.
- [C] Apenas os itens II e III estão certos.
- [D] Todos os itens estão certos.

# Questões de concursos

[CESPE/CEBRASPE 2024 SEBRAE-NACIONAL – Analista Técnico II – Dados – Business Intelligence] Assinale a opção correta.

- [A] Apenas os itens I e II estão certos.
- **[B] Apenas os itens I e III estão certos.**
- [C] Apenas os itens II e III estão certos.
- [D] Todos os itens estão certos.

# Questões de concursos

[CESPE/CEBRASPE 2024 MPO – Analista de Planejamento e Orçamento – Especialidade: Gestão de Dados Orçamentários] Acerca do Microsoft Power BI, julgue o item a seguir.

- O uso de gráficos de indicadores é ideal para exibir o status atual em relação a uma meta, enquanto os gráficos de barras e as colunas são mais adequados para comparar valores.

# Questões de concursos

[CESPE/CEBRASPE 2024 MPO – Analista de Planejamento e Orçamento – Especialidade: Gestão de Dados Orçamentários] Acerca do Microsoft Power BI, julgue o item a seguir.

- O uso de gráficos de indicadores é ideal para exibir o status atual em relação a uma meta, enquanto os gráficos de barras e as colunas são mais adequados para comparar valores.
  - Gabarito: **CERTO**.

# Inteligência artificial generativa

# Conceituação

É um ramo da Inteligência Artificial focado na criação de novos conteúdos

Como

A partir de padrões aprendidos

Textos, imagens, áudios, vídeos ou códigos

Em grandes volumes de dados

# Conceituação

Produzem algo novo

São diferentes das de IAs tradicionais

Simulando a criatividade humana

Essas apenas analisam ou classificam informações

# Visão geral




## Modelos que geram conteúdo novo a partir de padrões aprendidos em grandes volumes de dados

### IA preditiva

Classifica ou prevê a partir de dados existentes

Exemplo: estimar o risco de uma licitação

### IA generativa

Produz conteúdo que não existia

Exemplo: redigir a minuta de um relatório

# Características

## Gera

- Conteúdos originais
- Não apenas copia dados existentes

## Baseia-se em

- Modelos de linguagem
- Redes neurais avançadas

## Utiliza

- Aprendizado profundo (deep learning)
- Para compreender e replicar padrões complexos

## Capaz de

- Criar multimodalidade
- Texto, imagem, e áudio

# Exemplos de aplicações

## Textos

- ChatGPT
- Gera Redações, respostas e resumos

## Imagens

- DALL·E
- Cria ilustrações a partir de descrições em texto

## Áudio

- Modelos que
- Sintetizam vozes humanas
- Criam músicas

## Vídeo

- Geração de
- Cenas animadas ou realistas para entretenimento e treinamento

## Código

- GitHub Copilot
- Sugere e escreve códigos automaticamente

# Como funciona um modelo de linguagem

1 **Pré-treinamento**

Aprende a prever o próximo token em grandes volumes de texto

2 **Ajuste e alinhamento**

Refinado com exemplos e feedback humano para seguir instruções

3 **Inferência (uso)**

Recebe um prompt e gera a resposta token a token

## Token

Unidade de texto processada pelo modelo, em geral um pedaço de palavra

## Janela de contexto

Quantidade de texto que o modelo considera de uma só vez

## Transformer

Arquitetura baseada no mecanismo de atenção, base dos LLMs

## Temperatura

Parâmetro que controla o quanto as respostas variam

# Aplicações na análise de dados

1

## Gerar código e consultas

Escrever SQL ou Python a partir de uma pergunta em linguagem natural

2

## Estruturar dados não estruturados

Extrair campos de editais, contratos e notas em PDF e organizá-los em tabelas

3

## Resumir e comparar documentos

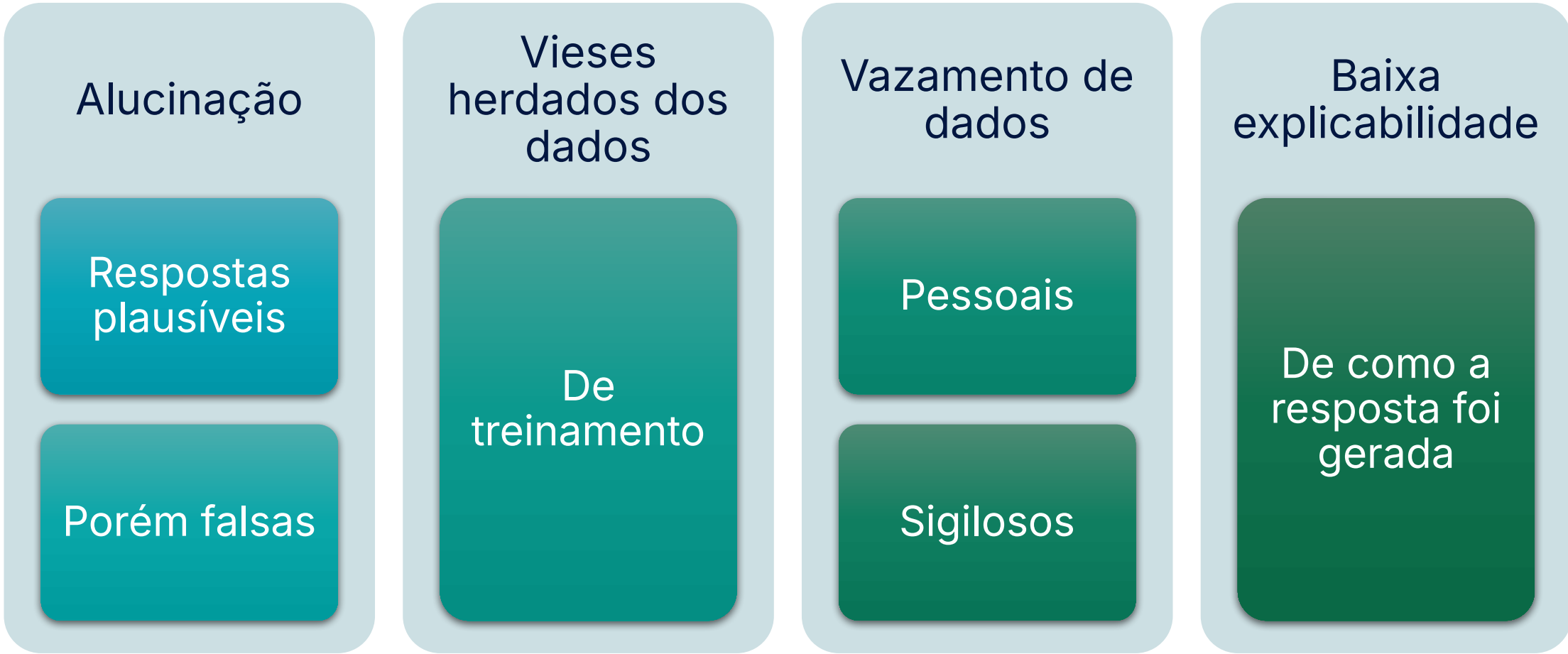
Sintetizar processos longos e apontar diferenças entre versões de um texto

4

## Explicar e comunicar resultados

Interpretar gráficos e redigir a primeira versão de relatórios e pareceres

# Riscos


Alucinação
Respostas plausíveis
Porém falsas
Vieses herdados dos dados
De treinamento
Vazamento de dados
Pessoais
Sigilosos
Baixa explicabilidade
De como a resposta foi gerada

# Boas práticas

## Revisão humana

Antes de usar qualquer resultado

## Conferir

Fontes

Números citados

## Não inserir dados sigilosos

Em ferramentas externas

## RAG

Ancorar as respostas em documentos oficiais

# Questões de concursos

[CESPE/CEBRASPE 2026 TCE/RN – Auditor de Controle Externo – Especialidade: Administração – Gestão de Políticas Públicas – Gestão Pública] Julgue o próximo item, relativo ao Microsoft Office, à Política Nacional de Educação Digital (PNED), bem como a inteligência artificial generativa (IAG), a ameaças digitais e ao Portal Brasileiro de Dados Abertos.

- Para a adoção de ferramentas de IAG no serviço público, é fundamental avaliar riscos como as alucinações, que são informações falsas ou fabricadas que, embora pareçam coerentes, resultam de limitações nos dados de treinamento ou na arquitetura do modelo.

# Questões de concursos

[CESPE/CEBRASPE 2026 TCE/RN – Auditor de Controle Externo – Especialidade: Administração – Gestão de Políticas Públicas – Gestão Pública] Julgue o próximo item, relativo ao Microsoft Office, à Política Nacional de Educação Digital (PNED), bem como a inteligência artificial generativa (IAG), a ameaças digitais e ao Portal Brasileiro de Dados Abertos.

- Para a adoção de ferramentas de IAG no serviço público, é fundamental avaliar riscos como as alucinações, que são informações falsas ou fabricadas que, embora pareçam coerentes, resultam de limitações nos dados de treinamento ou na arquitetura do modelo.
  - Gabarito: **CERTO**.

# Questões de concursos

[CESPE/CEBRASPE 2025 TJ/PA – Analista Judiciário – Especialidade: Análise de Suporte] Acerca dos tipos de aprendizado, da inteligência artificial generativa e das redes neurais, julgue o item subsequente.

- A inteligência artificial generativa inclui modelos como as redes generativas adversárias e os modelos baseados em transformadores, os quais são capazes de criar novos conteúdos, como imagens, textos ou áudios.

# Comentários

## GANs

Generative Adversarial Networks

Um gerador produz amostras, enquanto um discriminador aprende a distinguir as amostras reais das geradas

O treinamento adversário ajuda o gerador a produzir resultados mais realistas

## Modelos baseados em Transformers

Em modelos como o GPT, o texto é gerado token a token, com cada novo token previsto a partir do contexto anterior

Outras variantes podem gerar imagens e áudios

# Questões de concursos

[CESPE/CEBRASPE 2025 TJ/PA – Analista Judiciário – Especialidade: Análise de Suporte] Acerca dos tipos de aprendizado, da inteligência artificial generativa e das redes neurais, julgue o item subsequente.

- A inteligência artificial generativa inclui modelos como as redes generativas adversárias e os modelos baseados em transformadores, os quais são capazes de criar novos conteúdos, como imagens, textos ou áudios.
  - Gabarito: **CERTO**.