

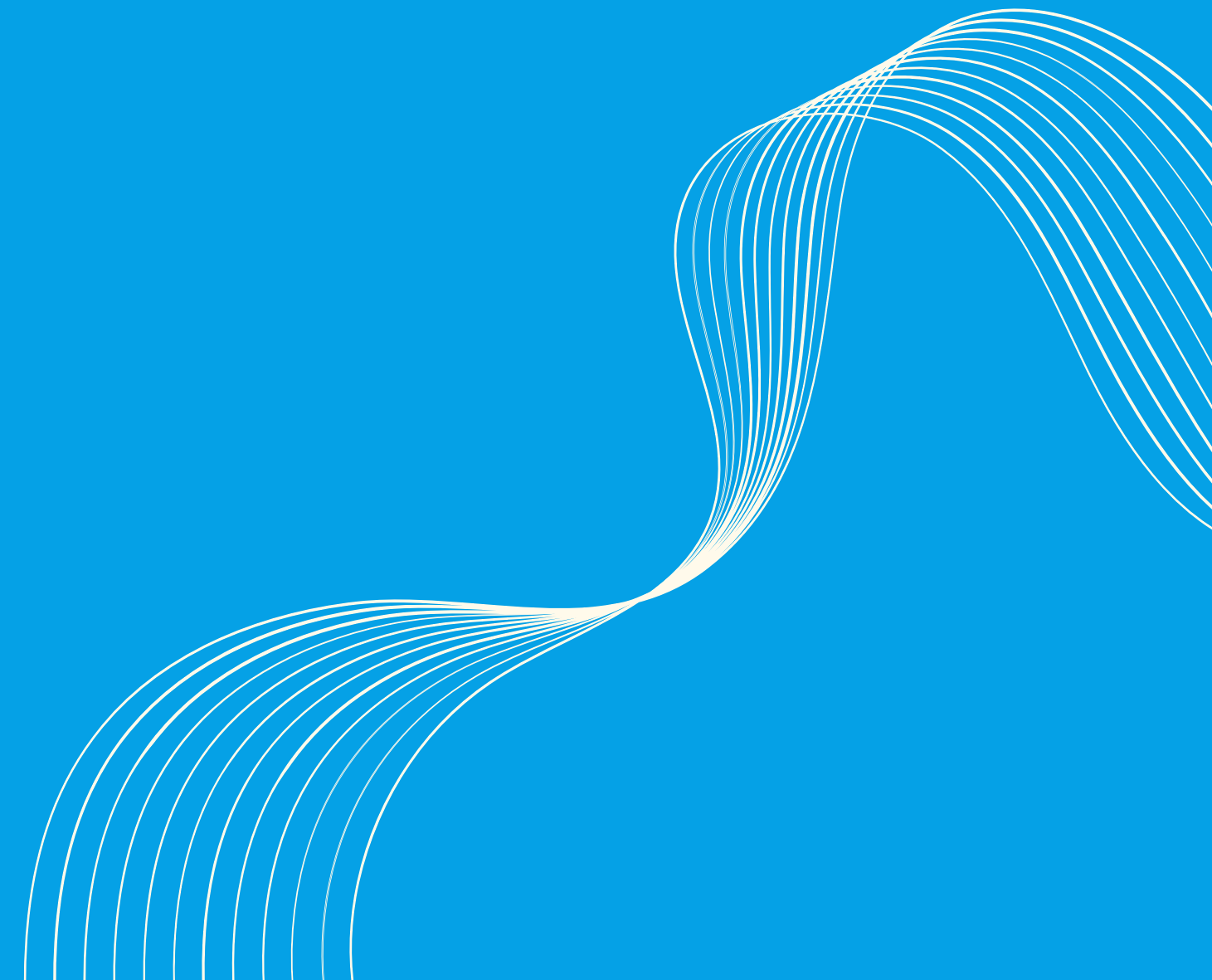
CAPÍTULO 04

DESIGNING MACHINE LEARNING SYSTEMS



Treinando os dados

- Dados de treinamento
- Use dados, mas não confie muito neles!

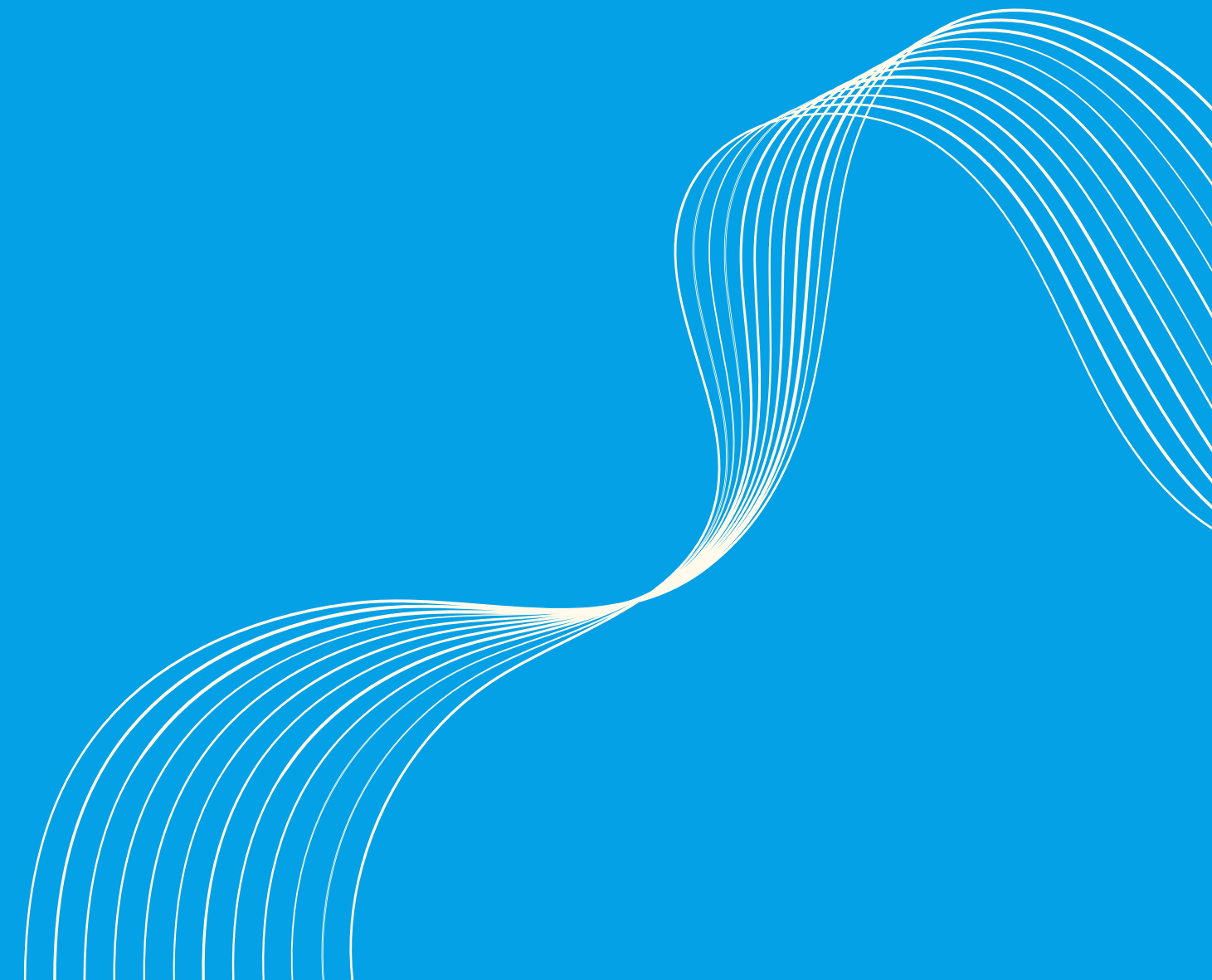


Amostragem

- Casos onde a amostragem é necessária - quase todos...
- Amostragem não probabilística e amostragem aleatória

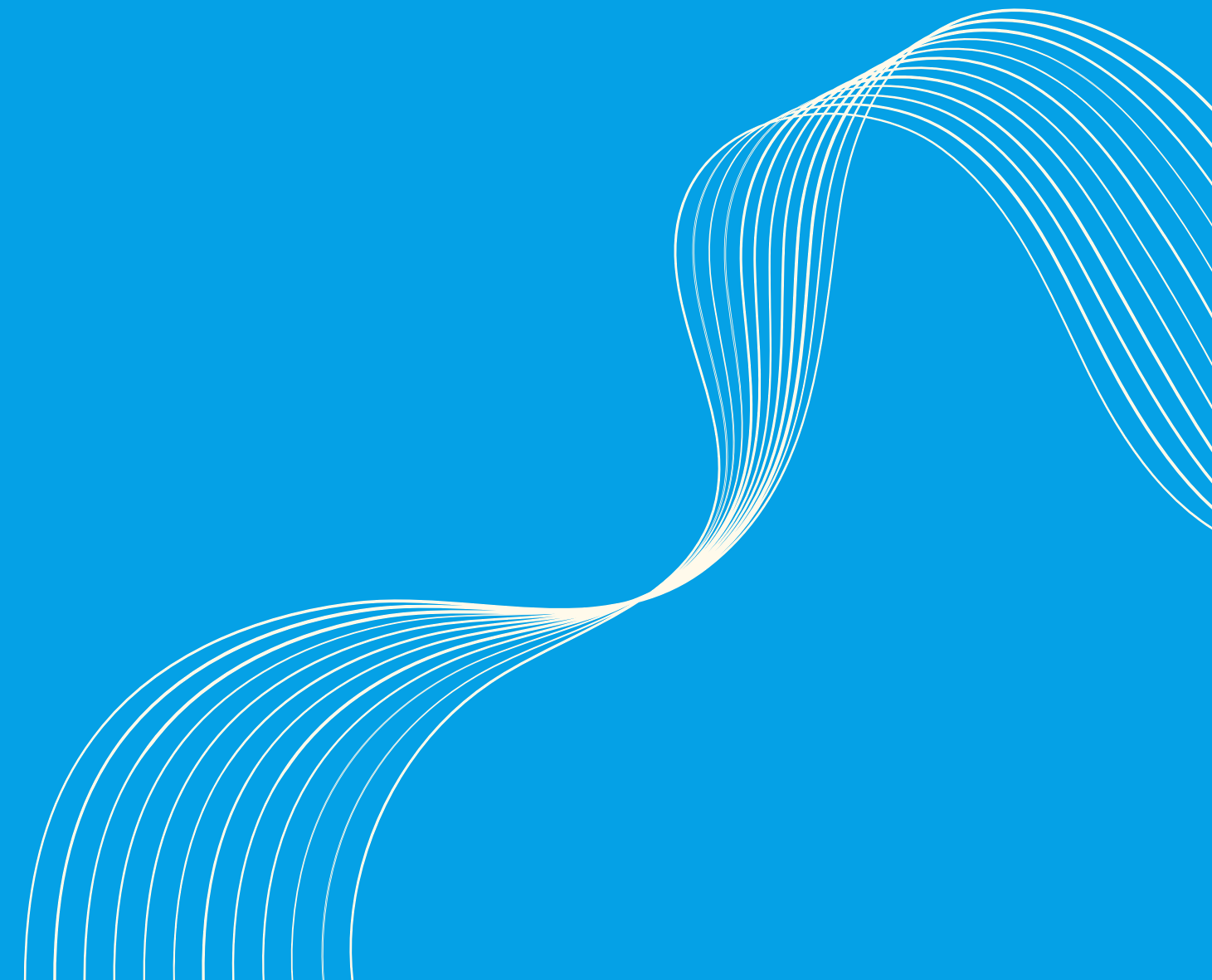


- Amostragem não probabilística
- Amostragem aleatória simples
- Amostragem estratificada
- Amostragem ponderada
- Amostragem de reservior
- Amostragem por importância



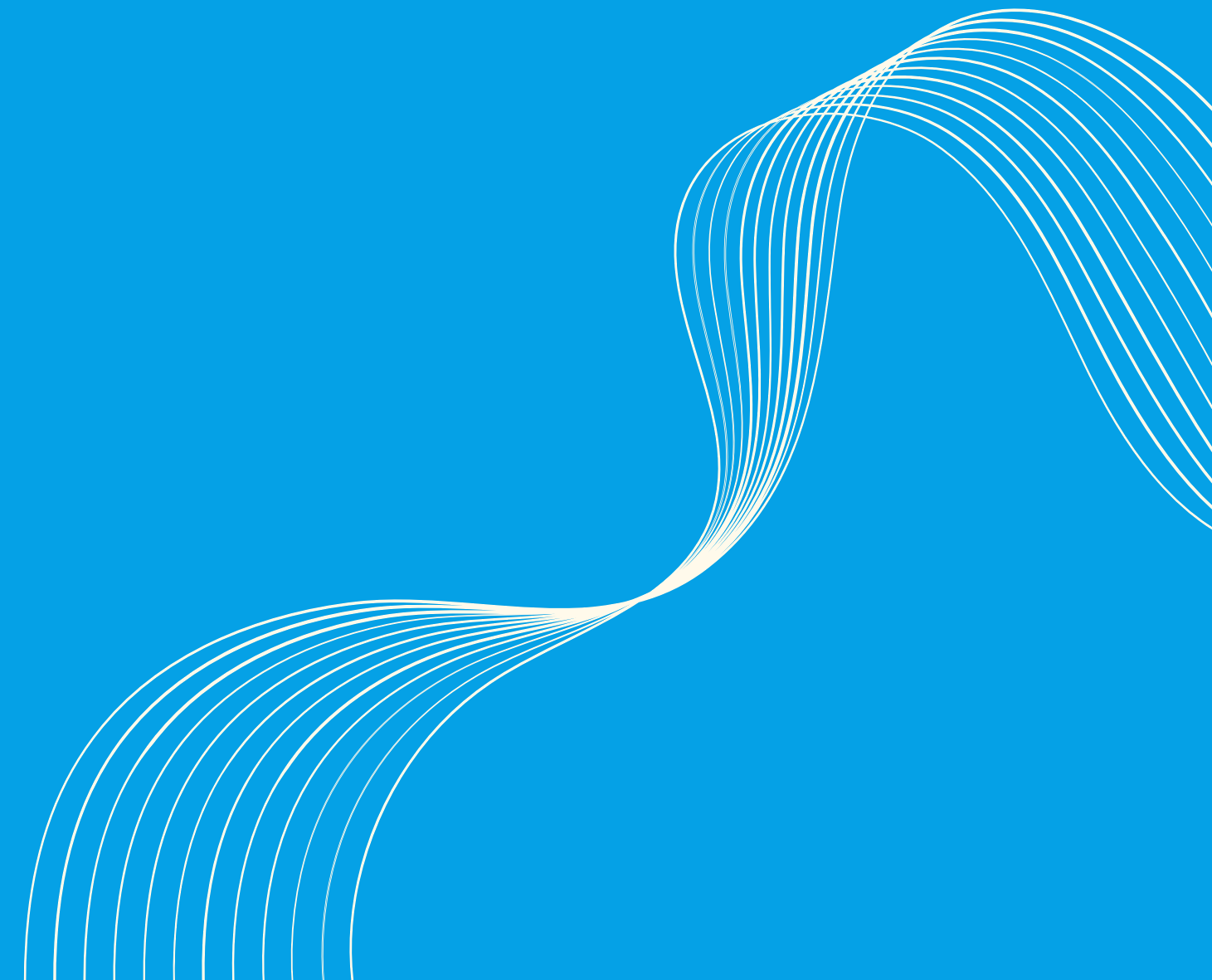
Rotulagem

- **Rótulos manuais**
 - Multiplicidade de rótulos
 - Linhagem dos dados
- **Rótulos naturais**
 - Extensão do loop de feedback



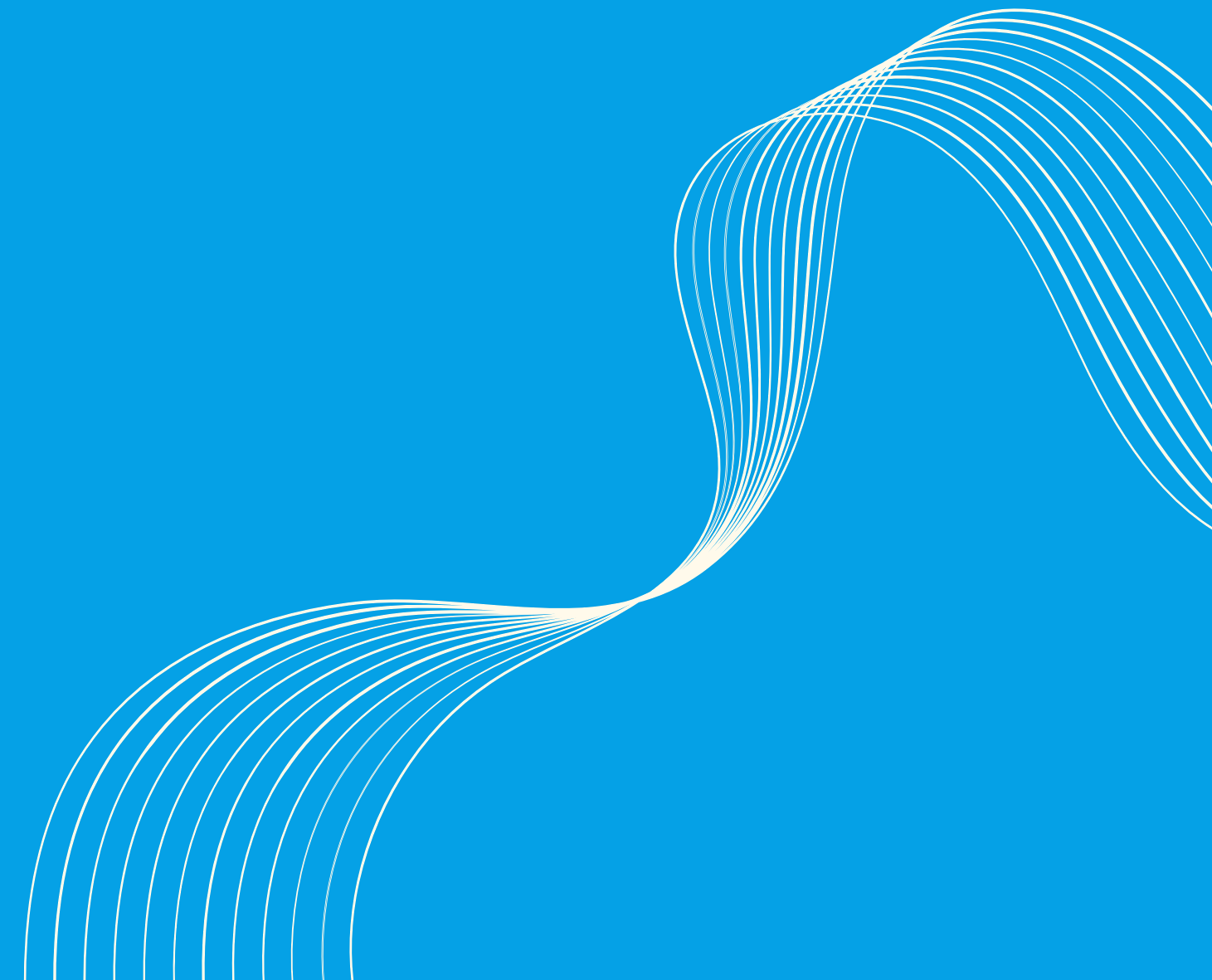
Rotulagem

- Lidando com a falta de rótulos
 - Weak supervision
 - Semisupervisão
 - Aprendizado por transferência
 - Aprendizado ativo




Classes desbalanceadas

- Como lidar?
 - Métricas adequadas
 - Reamostragem
 - oversampling
 - undersampling
 - Métodos de nível

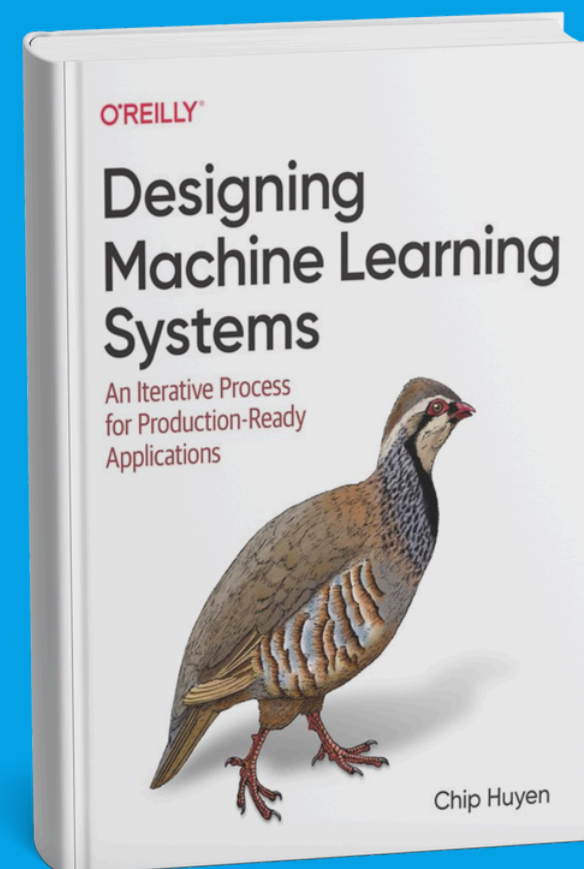


Data augmentation

- Aumenta a quantidade de dados de treinamento
 - Modelos mais robustos ao ruído
 - Ataques adversários
 - Transformação simples - preservação de rótulos
 - Perturbação
 - Síntese de dados
- 

Bibliografia

Designing Machine Learning Systems - Chip Huyen (O'Reilly, 2022)





That's all Folks!