



Machine Learning

Vinicius Caridá



Vinicius Caridá, Ph.D.

- Head of Digital Customer Service Platforms, PCP, WFM, Data and AI - Itaú Unibanco
- MBA Professor - FIAP



Qual é a maior dificuldade de machine learning (ML)?

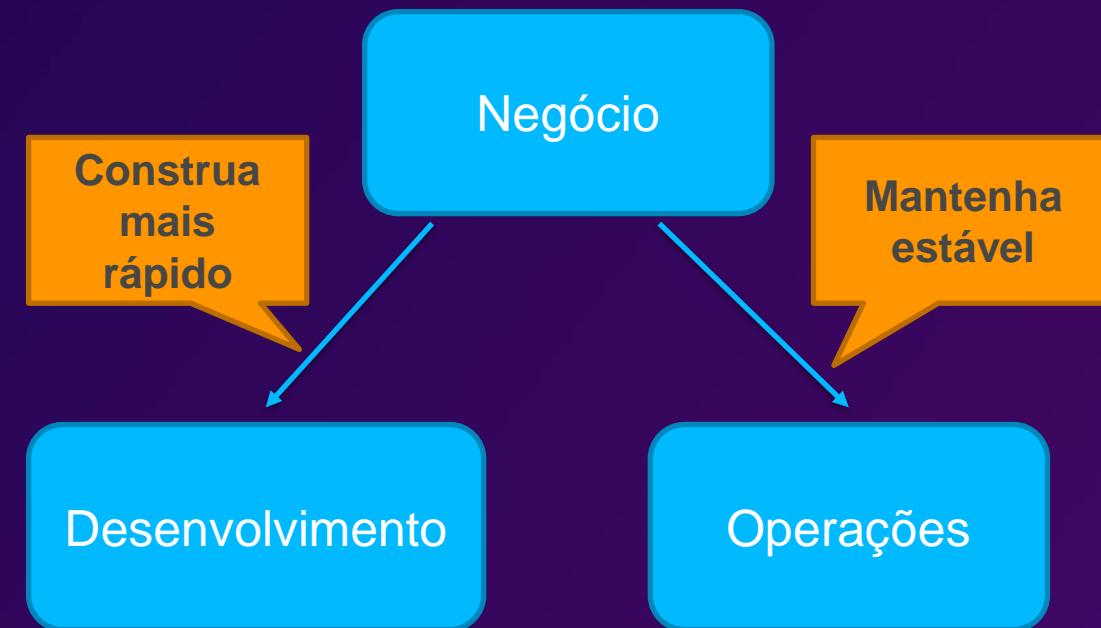
“A parte mais difícil de ML não é o ML, mas a massiva quantidade de esforços para manter sistemas de ML. É fácil de ficar dependente e difícil de sustentar.

- Anthony Penta, Sr. Manager & Principal Scientist, Amazon Consumer Payments

Hidden Technical Debt in Machine Learning Systems

<https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf>

Forças concorrentes:



Business

X

Flexibilidade

Agilidade

Pros/
Liberdade
Adaptavel
Lock-in

Cons/
Velocidade
Readyness
< curva de
Aprendizado

Pros/
Velocidade
Readyness
< curva de
Aprendizado

Cons/
Liberdade
Adaptavel
Lock-in

ML Well-Architected Framework

Excelência
Operacional



Segurança



Resiliência

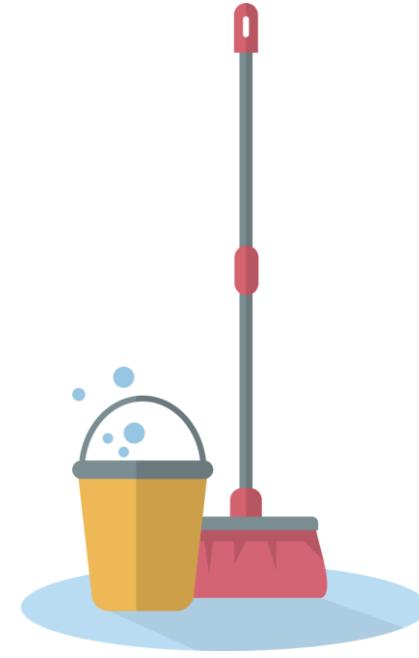


Performance



Custo

What about a dog and a mop? Easy, right?



Not so fast...





fast...





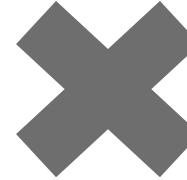
fast...



The Challenges on Productionizing ML

ML challenges

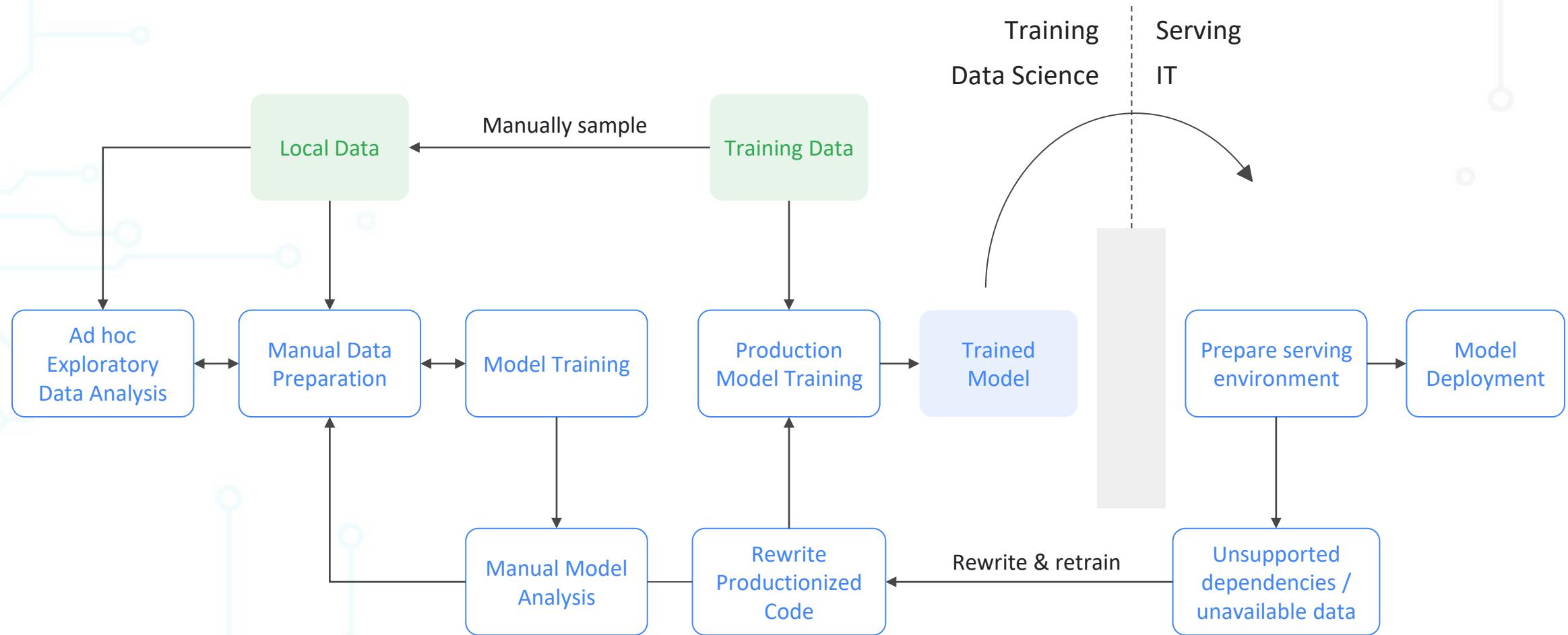
- Governance of data, features, models, pipelines and experiments
- Continuous training and deployment
- Training-serving skew
- Data validation
- Model analysis
- Fairness and Explainability



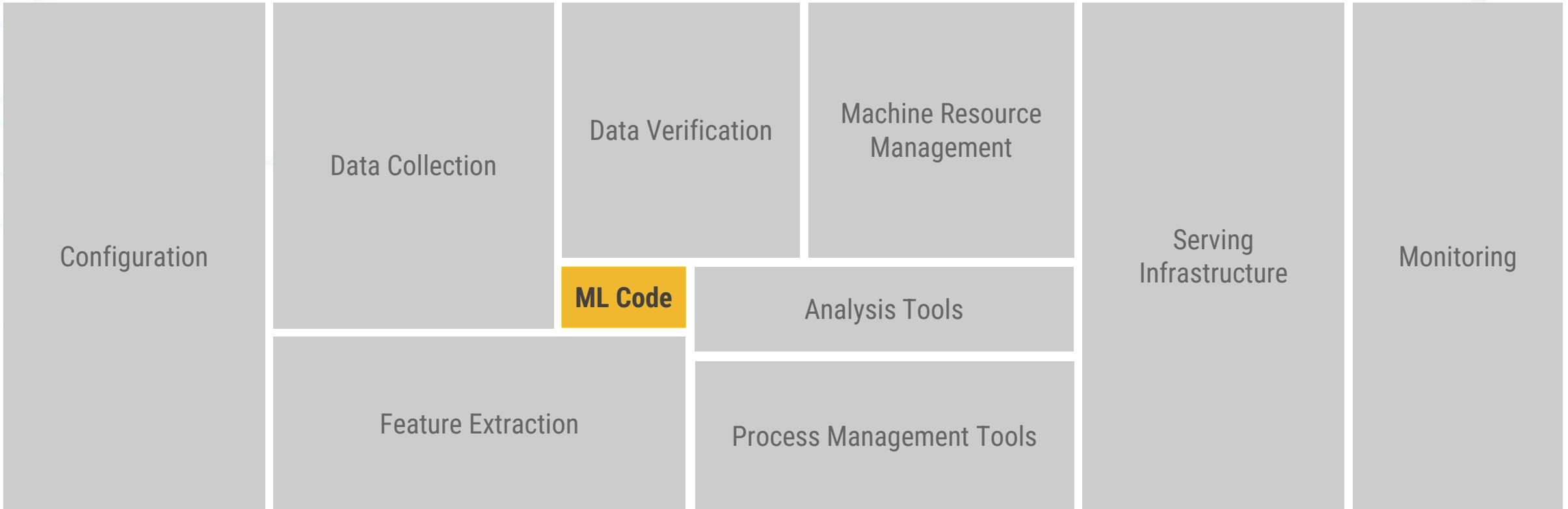
Production system challenges

- Scalability
- Availability
- Portability
- Reproducibility
- Modularity
- Monitoring and Alerting
- Security
- Hosted or Serverless

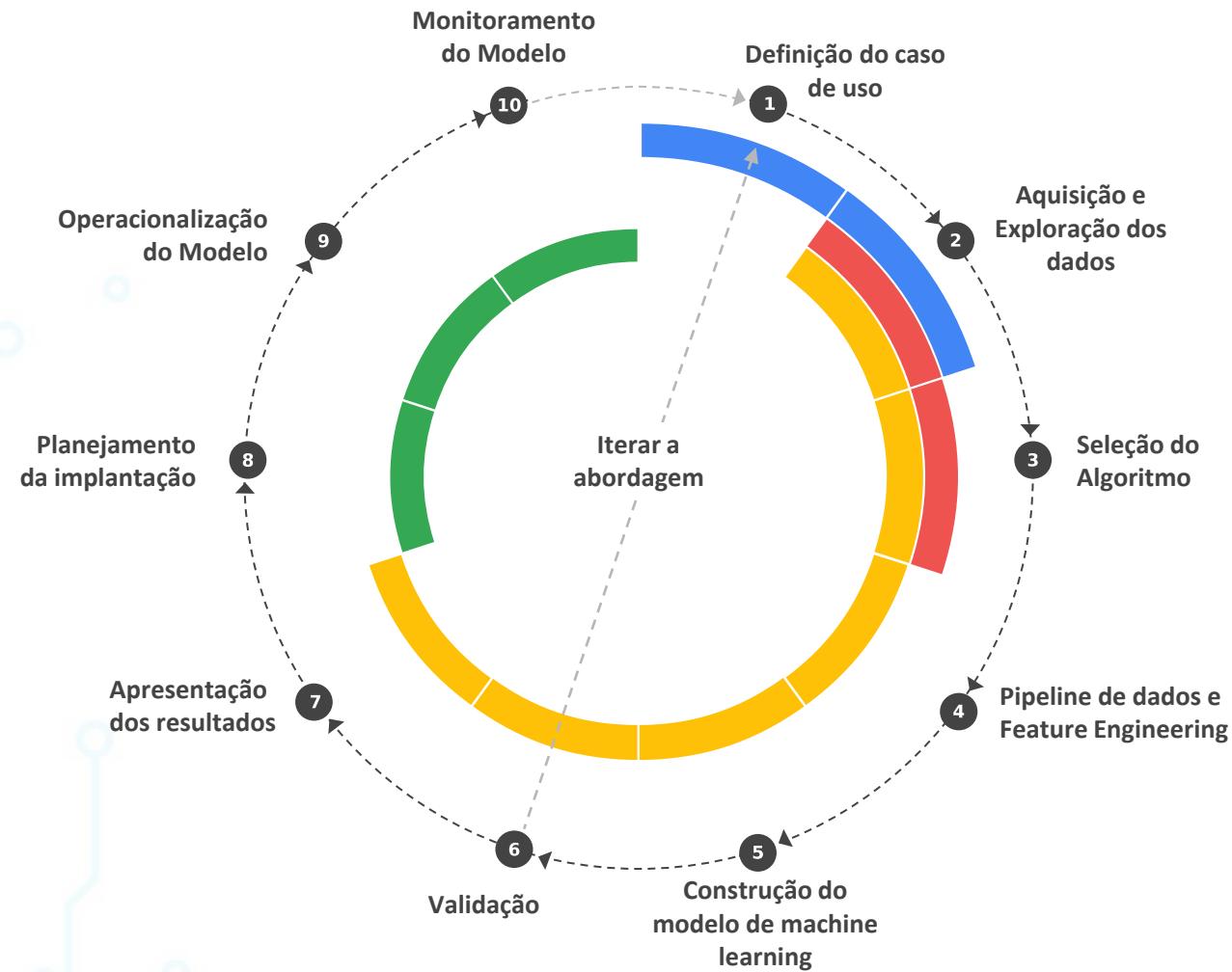
What's happening today: Data Science and IT (Ops) are isolated



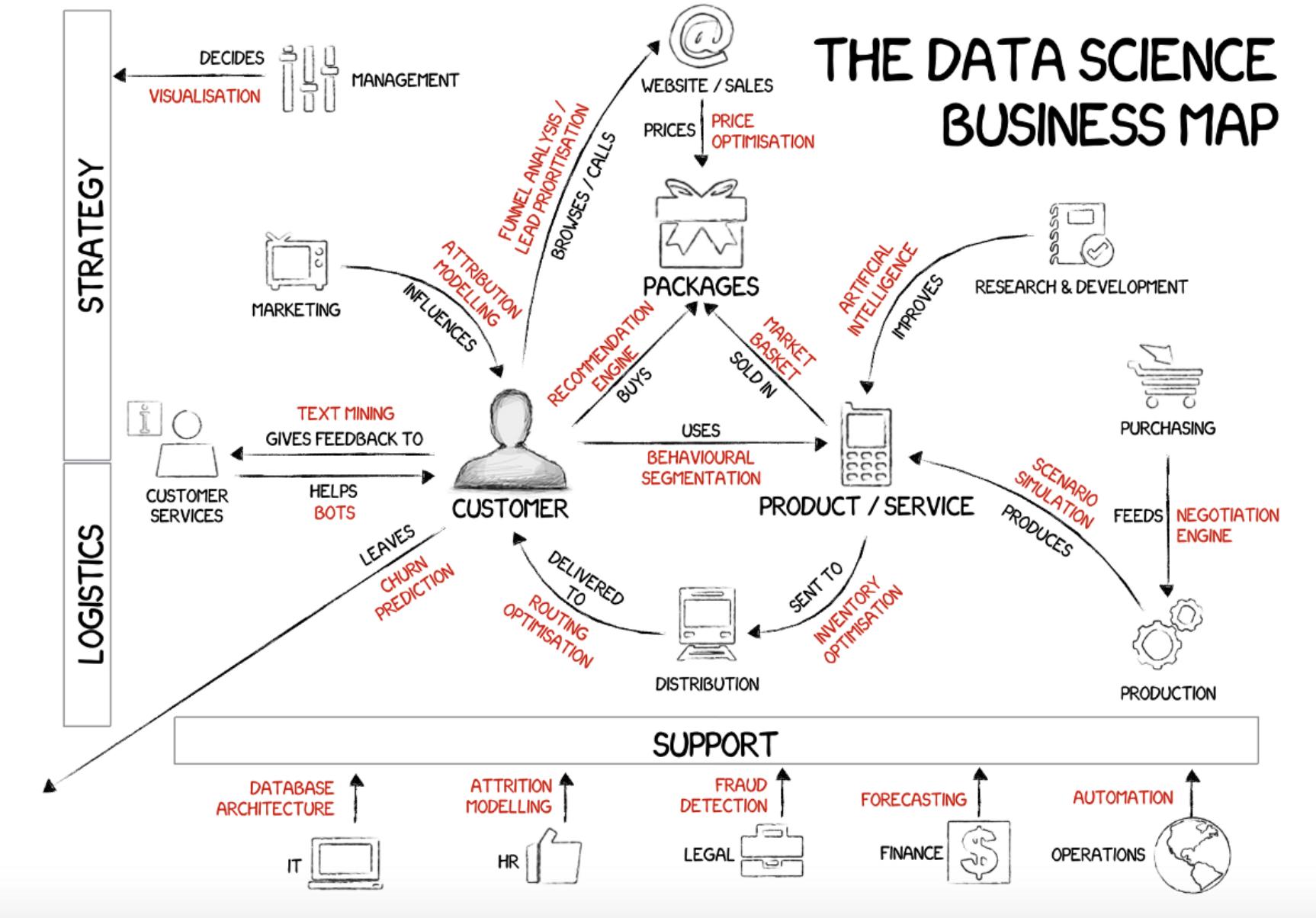
... a production solution requires so much more



ML lifecycle



THE DATA SCIENCE BUSINESS MAP



“More than 87% of data science projects never make it into production”



- Multiple studies and surveys -

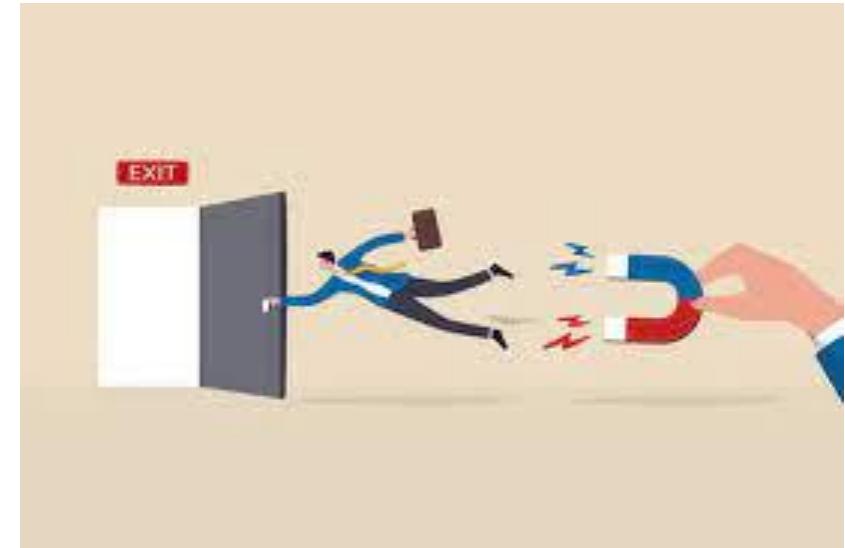


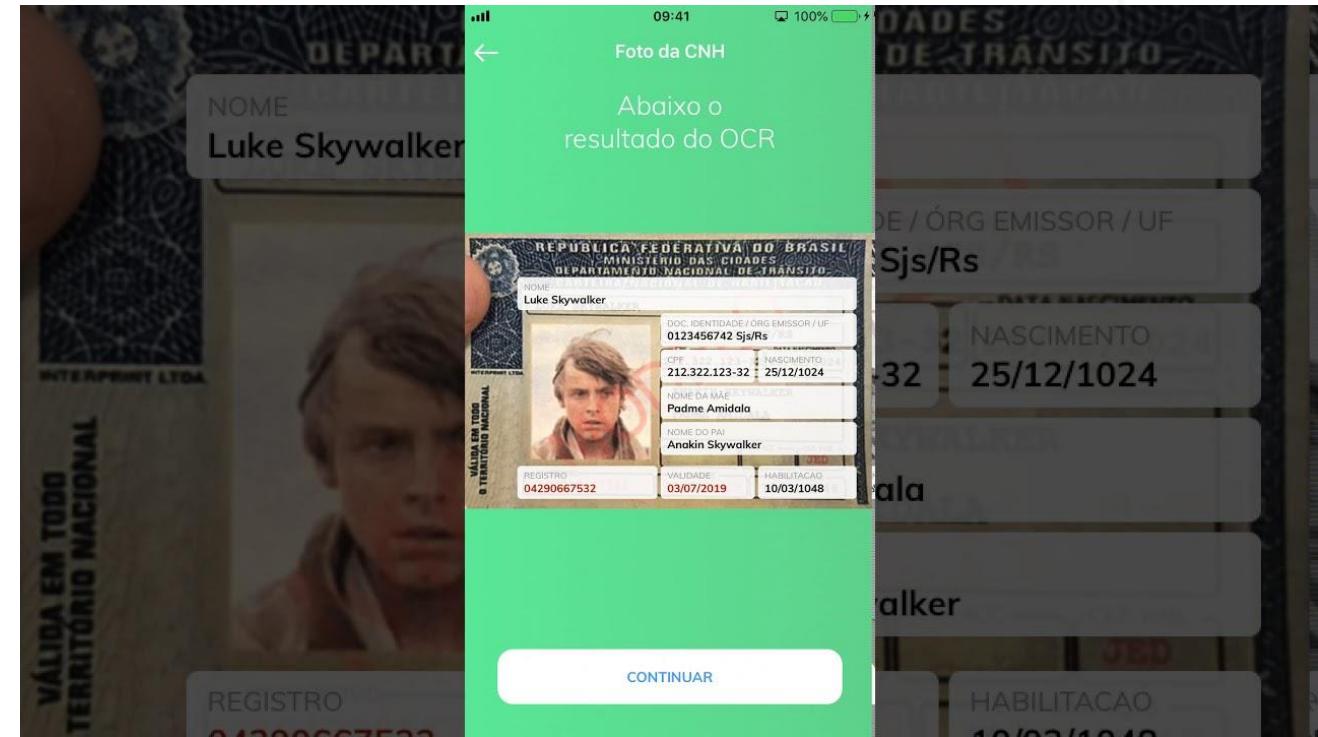
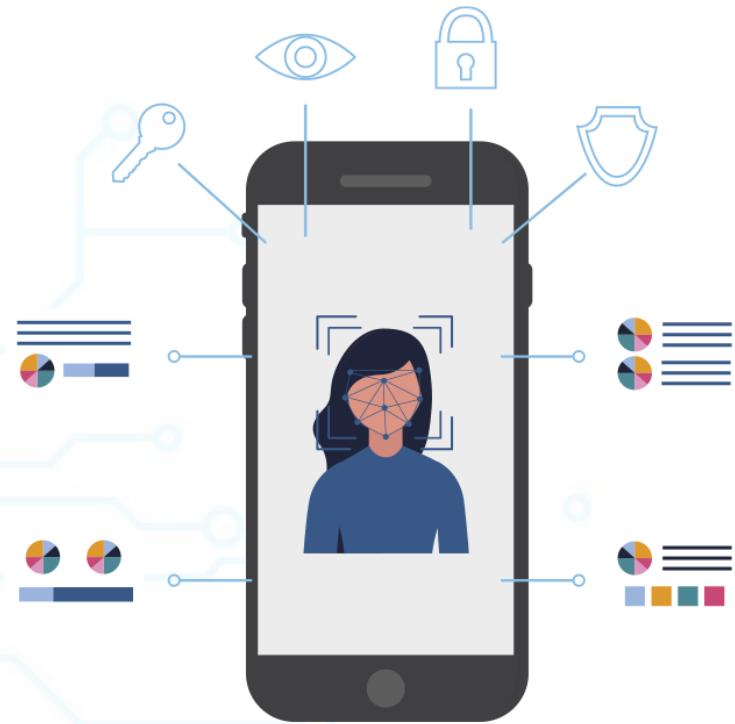
Artwork for Stranger Things that each receive over 5% of impressions from our personalization algorithm.

Different images cover a breadth of themes in the show to go beyond what any single image portrays.



Artwork for Stranger Things that each receive over 5% of impressions from our personalization algorithm.
Different images cover a breadth of themes in the show to go beyond what any single image portrays.







API



Property	88%
House	87%
Architecture	85%
Home	81%



Landmark	94%
Sky	93%
House	89%
Building	89%



House	94%
Home	92%
Property	92%
Real Estate	80%



Property	92%
Building	84%
Town	84%
House	83%

Custom



Tudor	.5
Neoclassical	.1
Modern	.89
Ranch	.5



Tudor	.1
Victorian	.94
Modern	.3
Ranch	.1



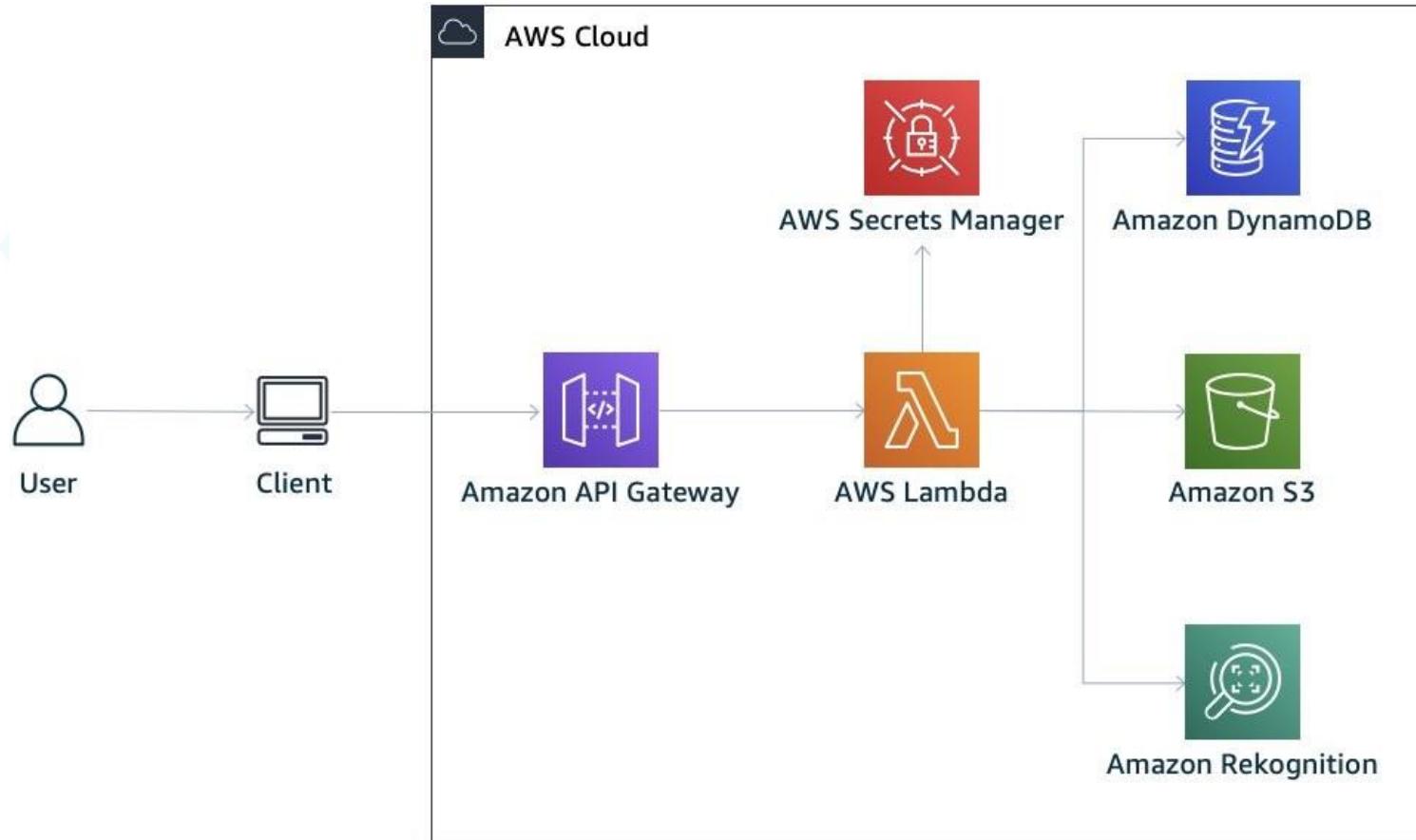
Tudor	.2
Neoclassical	.3
Modern	.3
Ranch	.93



Tudor	.92
Neoclassical	.4
Modern	.2
Ranch	.2

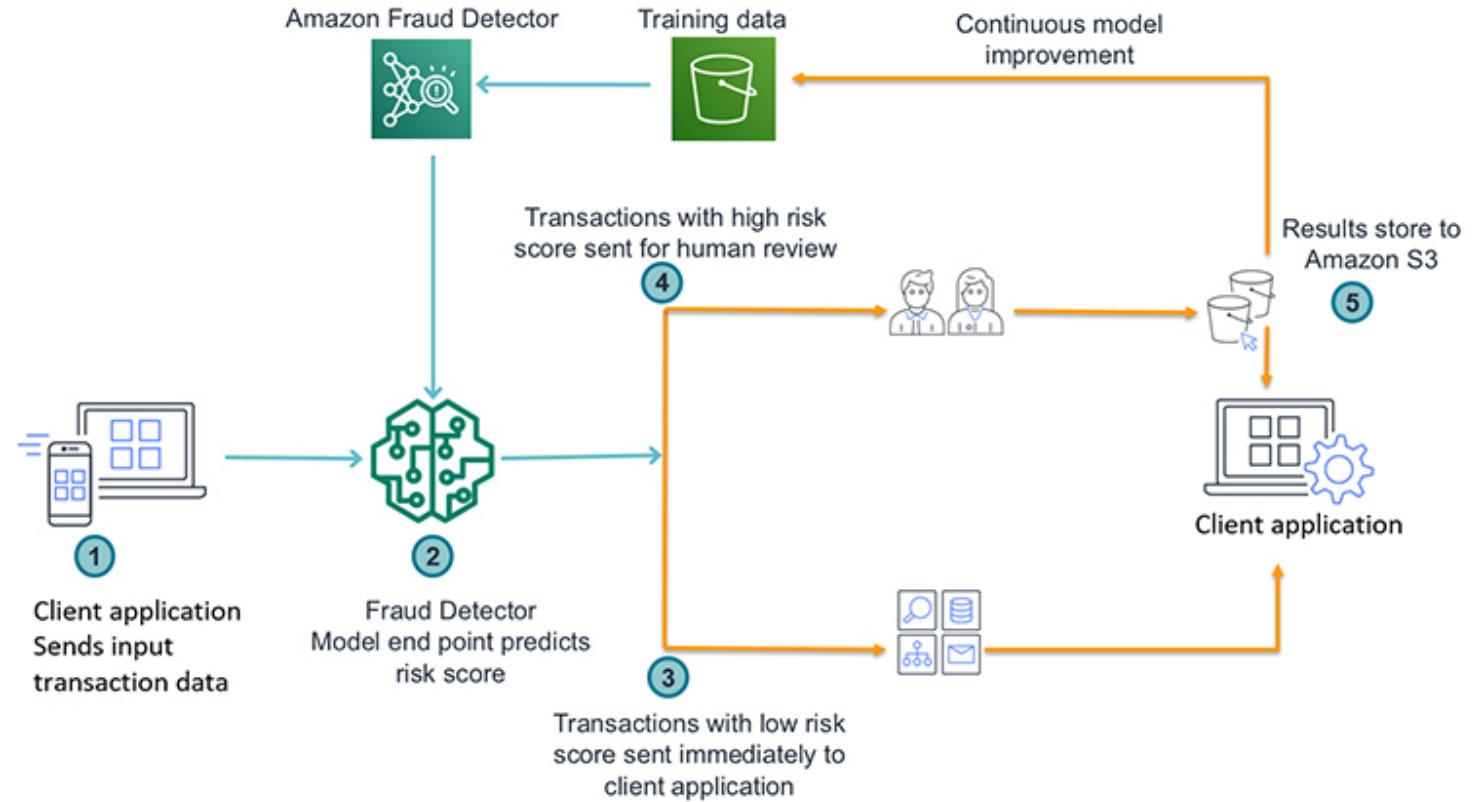
Arquiteturas de Visão Computacional

- Liveness Detection



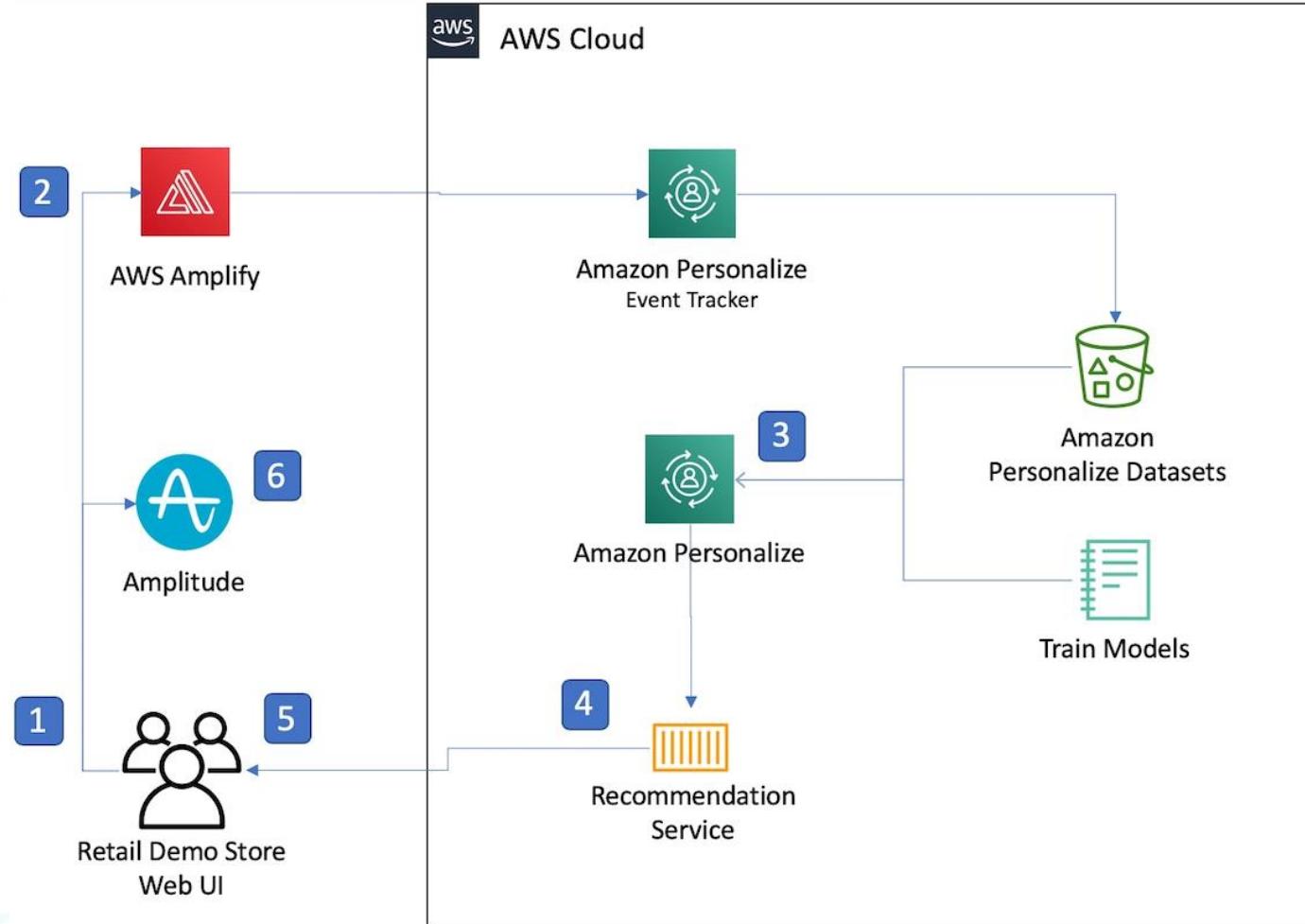
Arquiteturas de Fraude

- Detecção de fraude



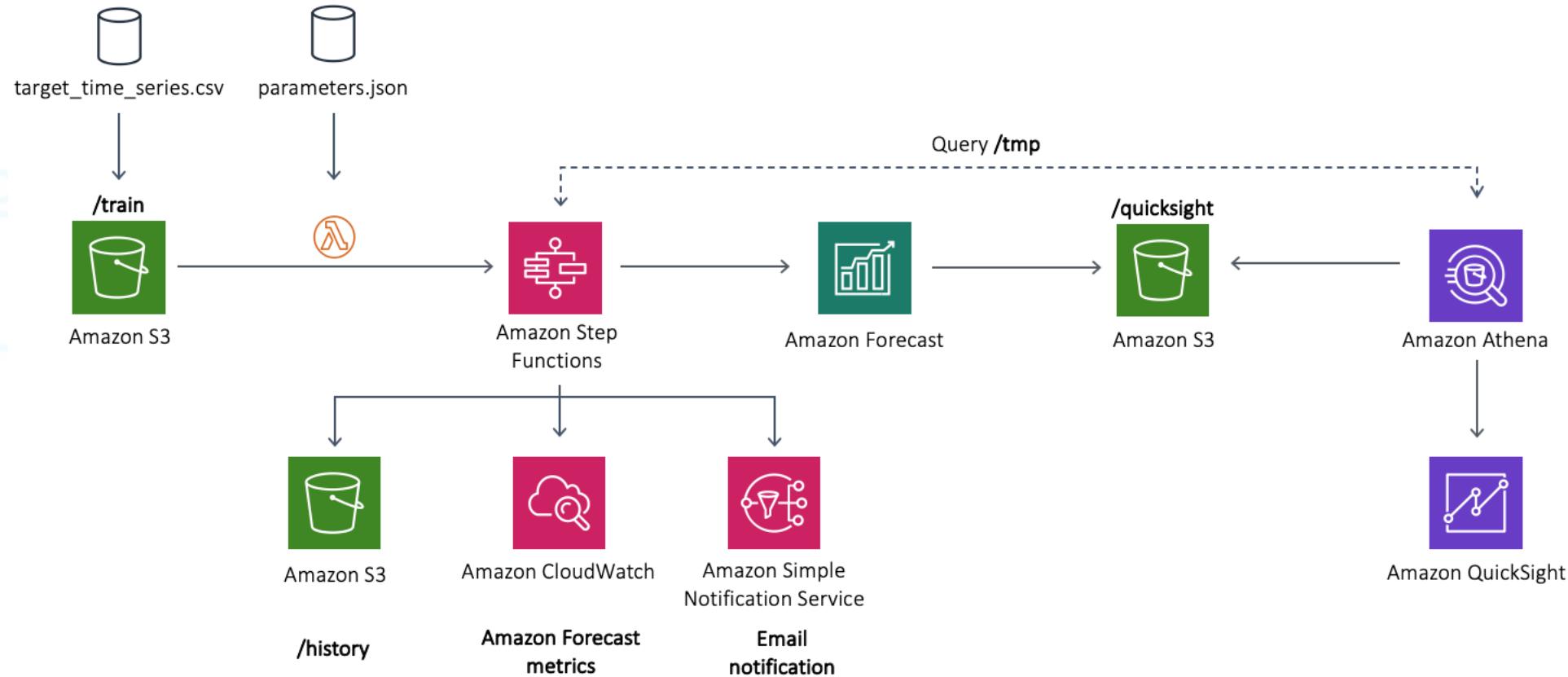
Arquiteturas de Recomendação

- Recomendação de produtos



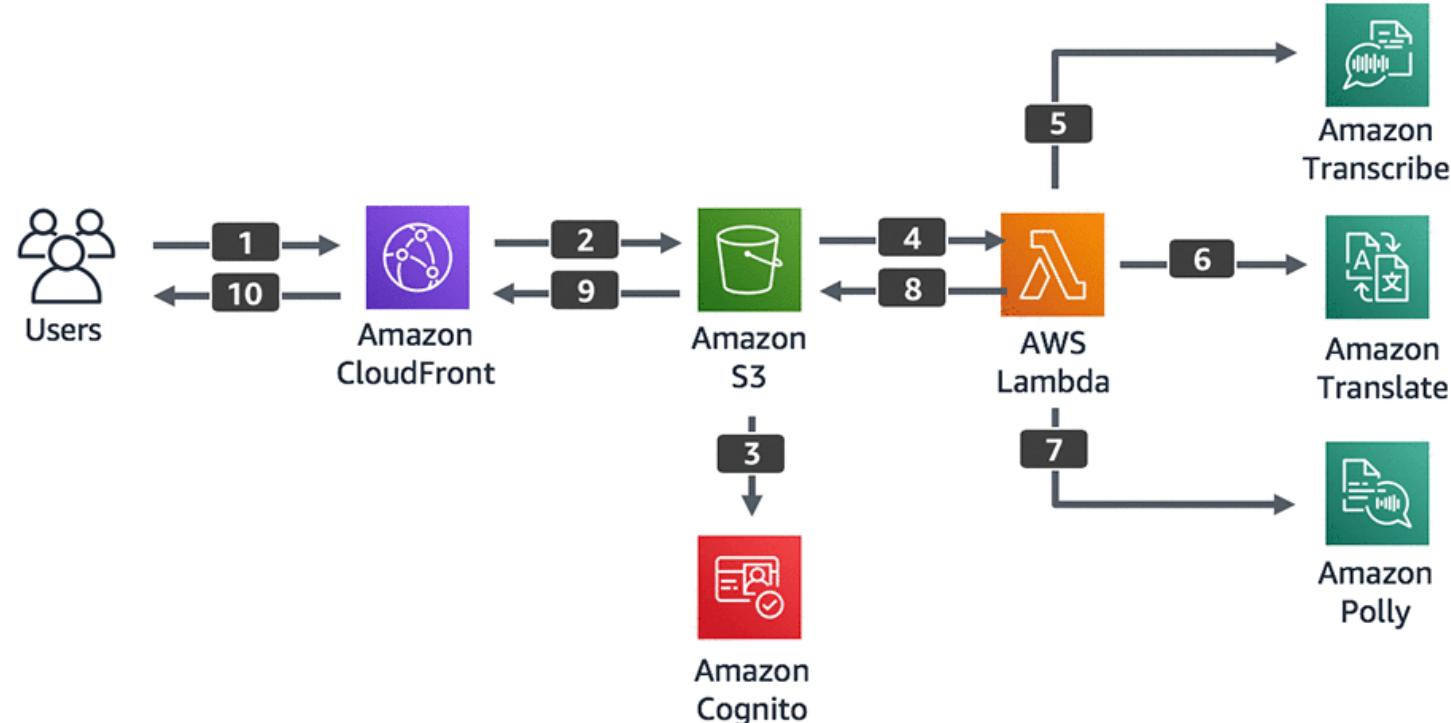
Arquiteturas de Previsão de vendas

- Previsão de vendas



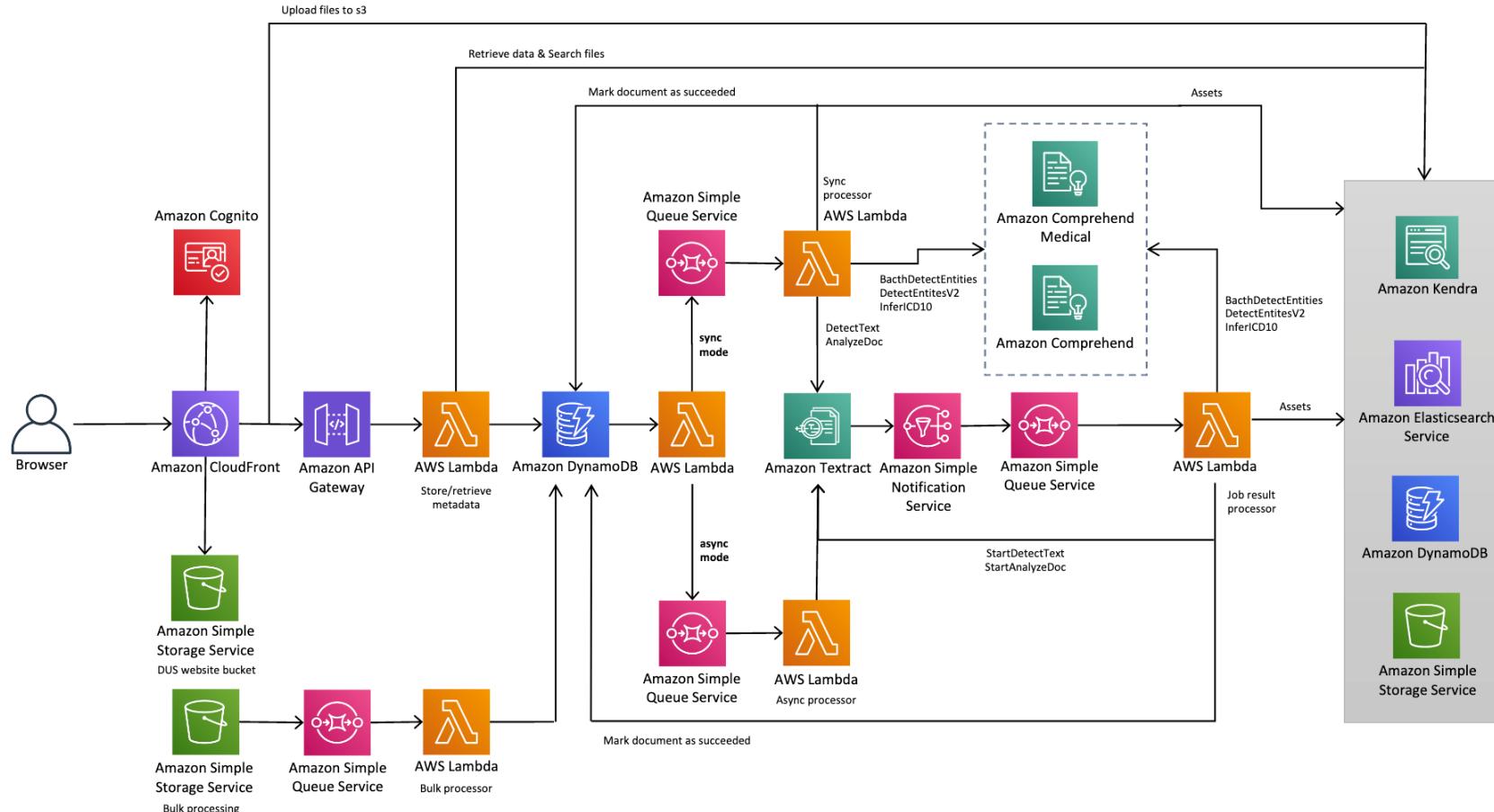
Arquiteturas de Voz/Texto

- Tradutor simultâneo



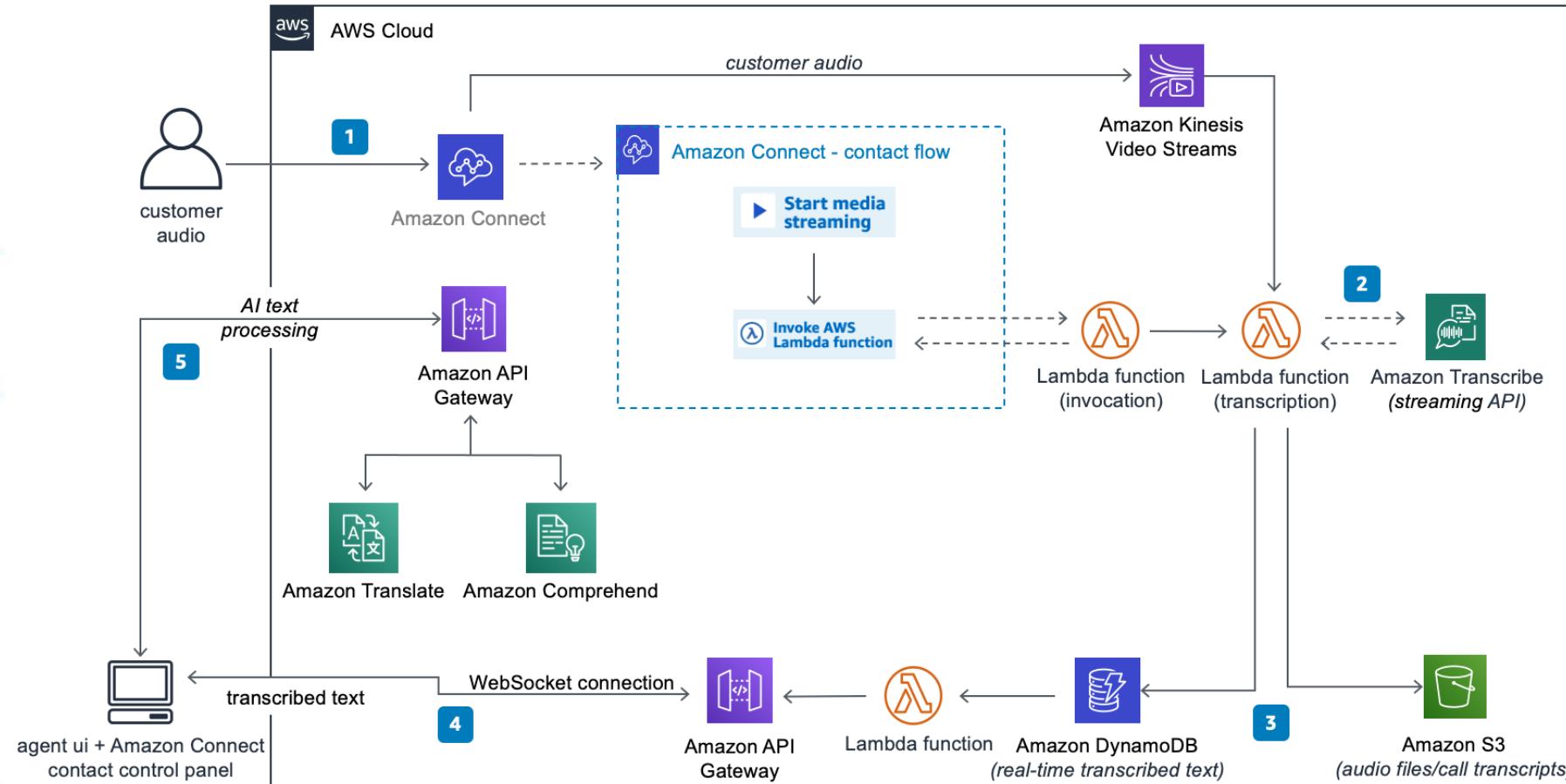
Arquiteturas de Visão Computacional

- Processamento de documentos



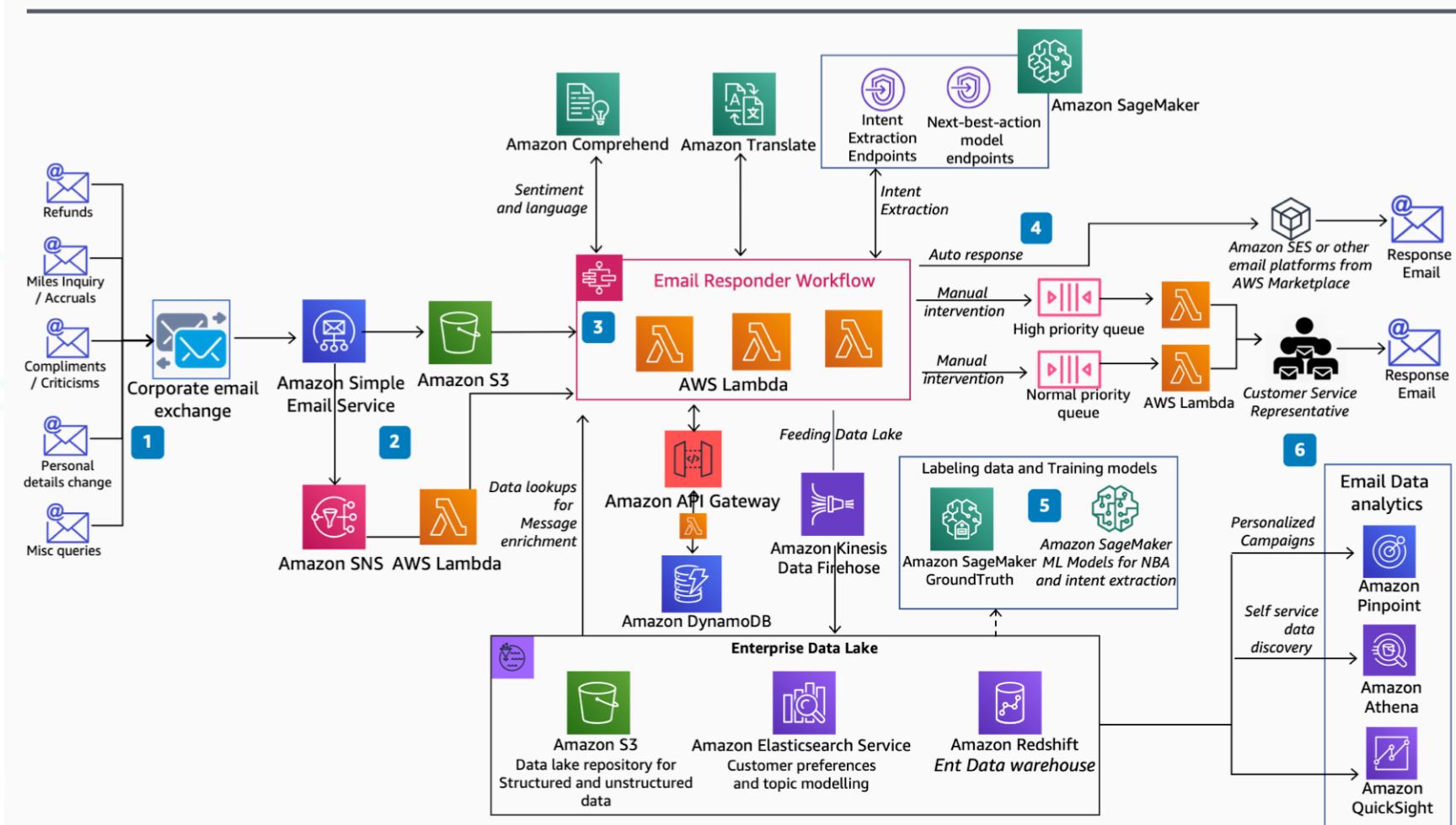
Arquiteturas de Voz/Texto

- Processamento de ligações



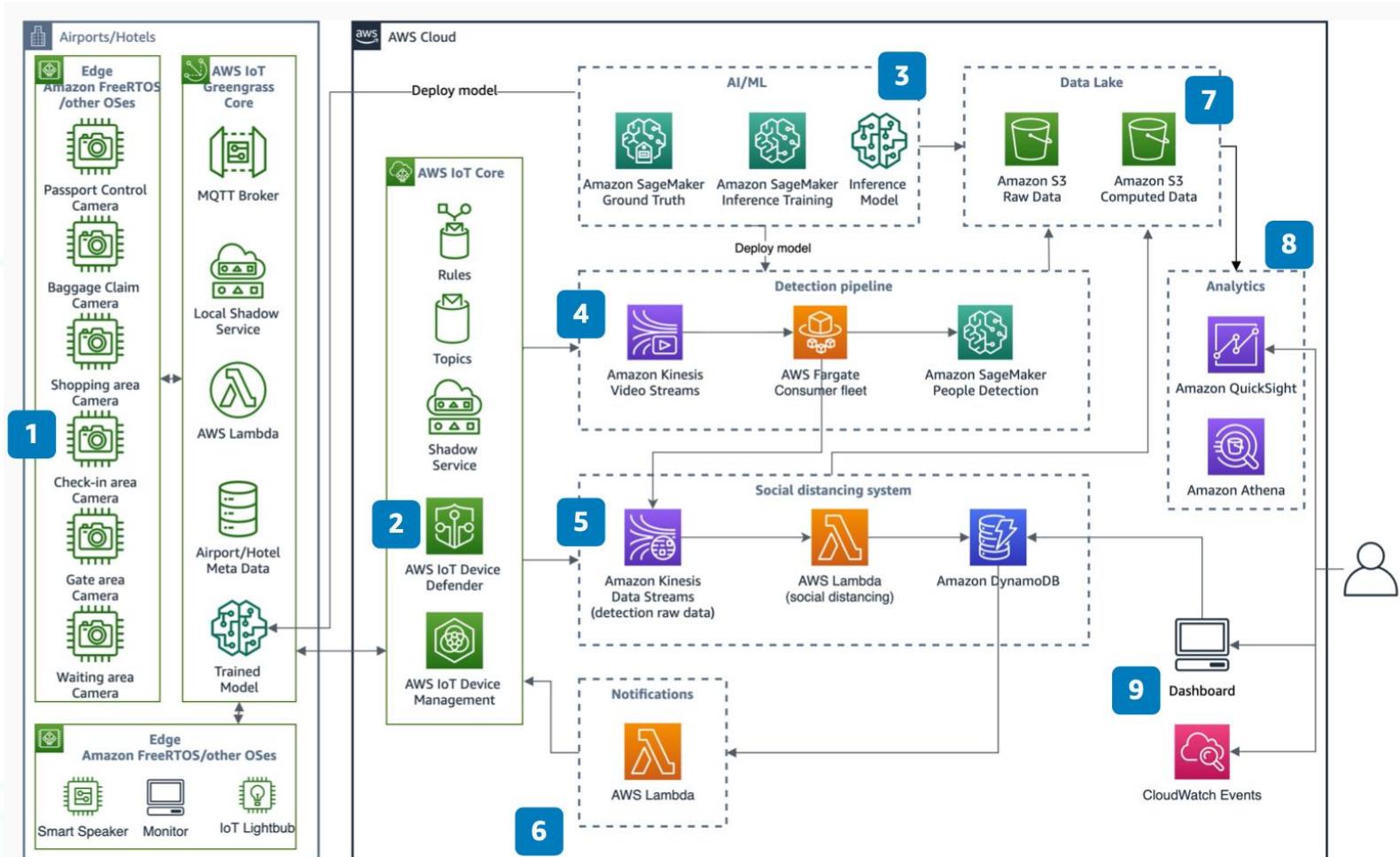
Arquiteturas de Texto

- Respondedor de e-mail



Arquiteturas com IoT

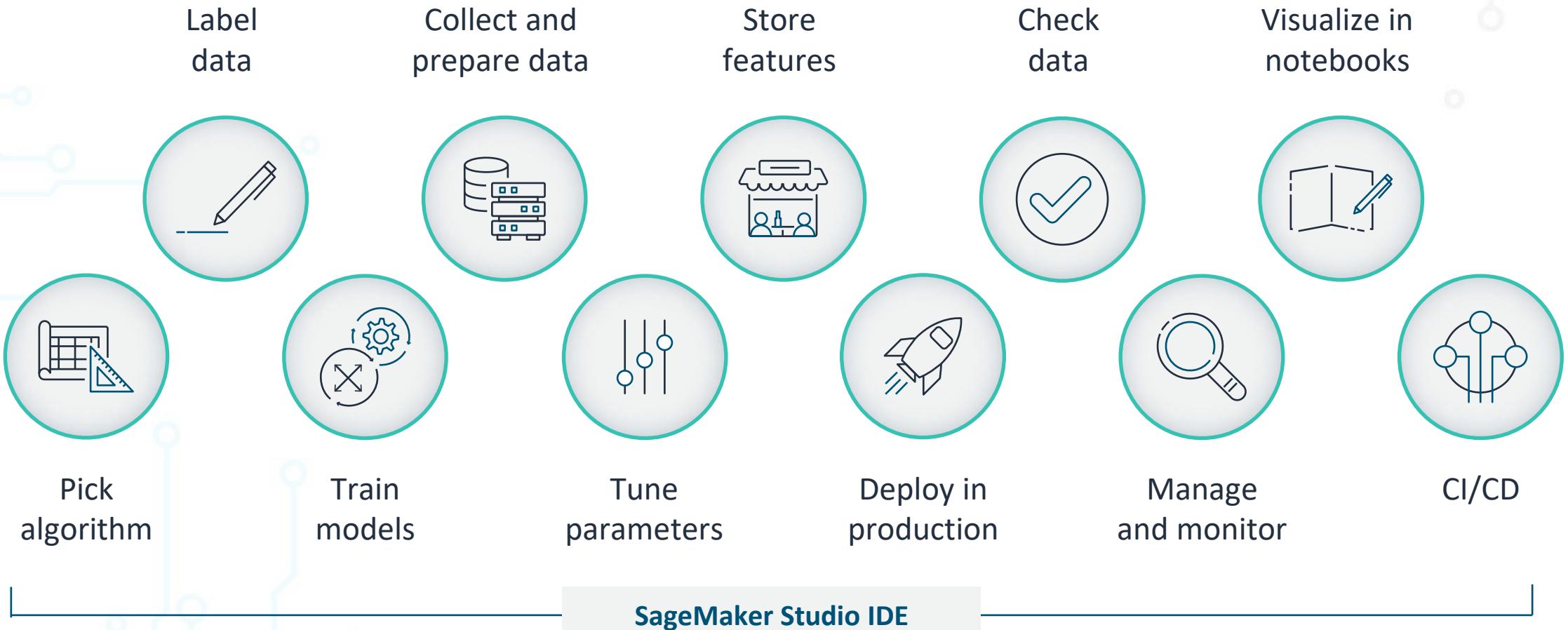
- Arquitetura de IoT + ML



Serviços de Inteligência Artificial

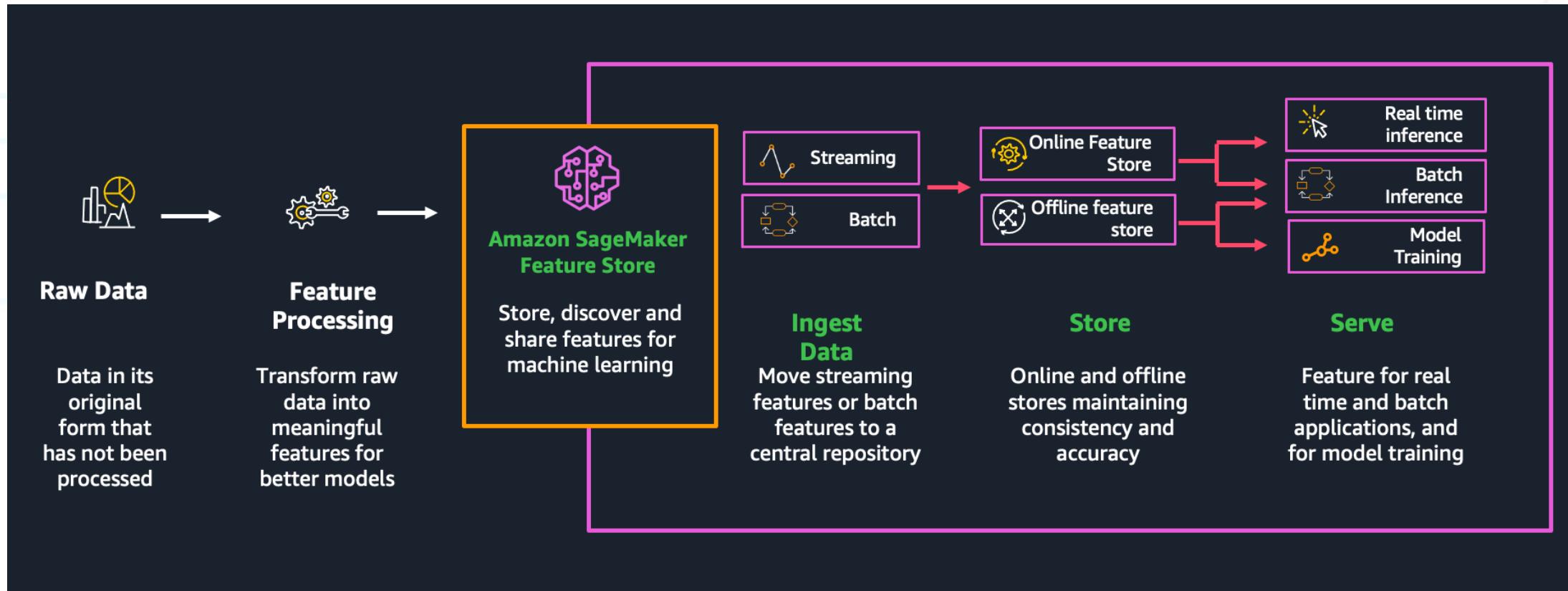
 <h3>Estudos analíticos avançados de texto</h3> <p>Use o processamento de linguagem natural para extraír insights e relacionamentos de texto não estruturado</p> <p>Amazon Comprehend »</p>	 <h3>Automatize as revisões de código</h3> <p>Automatize as revisões de código e identifique as linhas de código mais caras</p> <p>Amazon CodeGuru »</p>	 <h3>Chatbots</h3> <p>Crie agentes conversacionais com facilidade para melhorar o atendimento ao cliente e aumentar a eficiência da central de contatos</p> <p>Amazon Lex »</p>	 <h3>Tradução em tempo real</h3> <p>Amplie seu alcance com tradução eficiente e econômica para atender a públicos-alvo em vários idiomas</p> <p>Amazon Translate »</p>
 <h3>Previsão sob demanda</h3> <p>Crie modelos de previsão precisos com base na mesma tecnologia de previsão por machine learning usada pela Amazon.com</p> <p>Amazon Forecast »</p>	 <h3>Análise de documentos</h3> <p>Extraia texto e dados automaticamente de milhões de documentos em poucas horas, reduzindo o esforço manual</p> <p>Amazon Textract »</p>	 <h3>Pesquisa empresarial</h3> <p>Adicione recursos de pesquisa com linguagem natural às aplicações para que os usuários encontrem as informações que precisam com mais facilidade</p> <p>Amazon Kendra »</p>	 <h3>Texto em fala</h3> <p>Converta texto em falas que soam verdadeiras para dar voz a suas aplicações</p> <p>Amazon Polly »</p>
 <h3>Prevenção de fraudes</h3> <p>Identifique atividades online possivelmente fraudulentas com base na mesma tecnologia usada na Amazon.com</p> <p>Amazon Fraud Detector »</p>	 <h3>Análise de imagens e vídeos</h3> <p>Adicione análise de imagens e vídeos a suas aplicações para catalogar ativos, automatizar fluxos de trabalho de mídia e extrair significado</p> <p>Amazon Rekognition »</p>	 <h3>Recomendações personalizadas</h3> <p>Personalize experiências para os clientes usando a tecnologia de machine learning aperfeiçoada em anos de uso na Amazon.com</p> <p>Amazon Personalize »</p>	 <h3>Transcrição</h3> <p>Adicione facilmente recursos de conversão de fala em texto com alta qualidade a suas aplicações e fluxos de trabalho</p> <p>Amazon Transcribe »</p>

Amazon SageMaker: Built to make ML more accessible



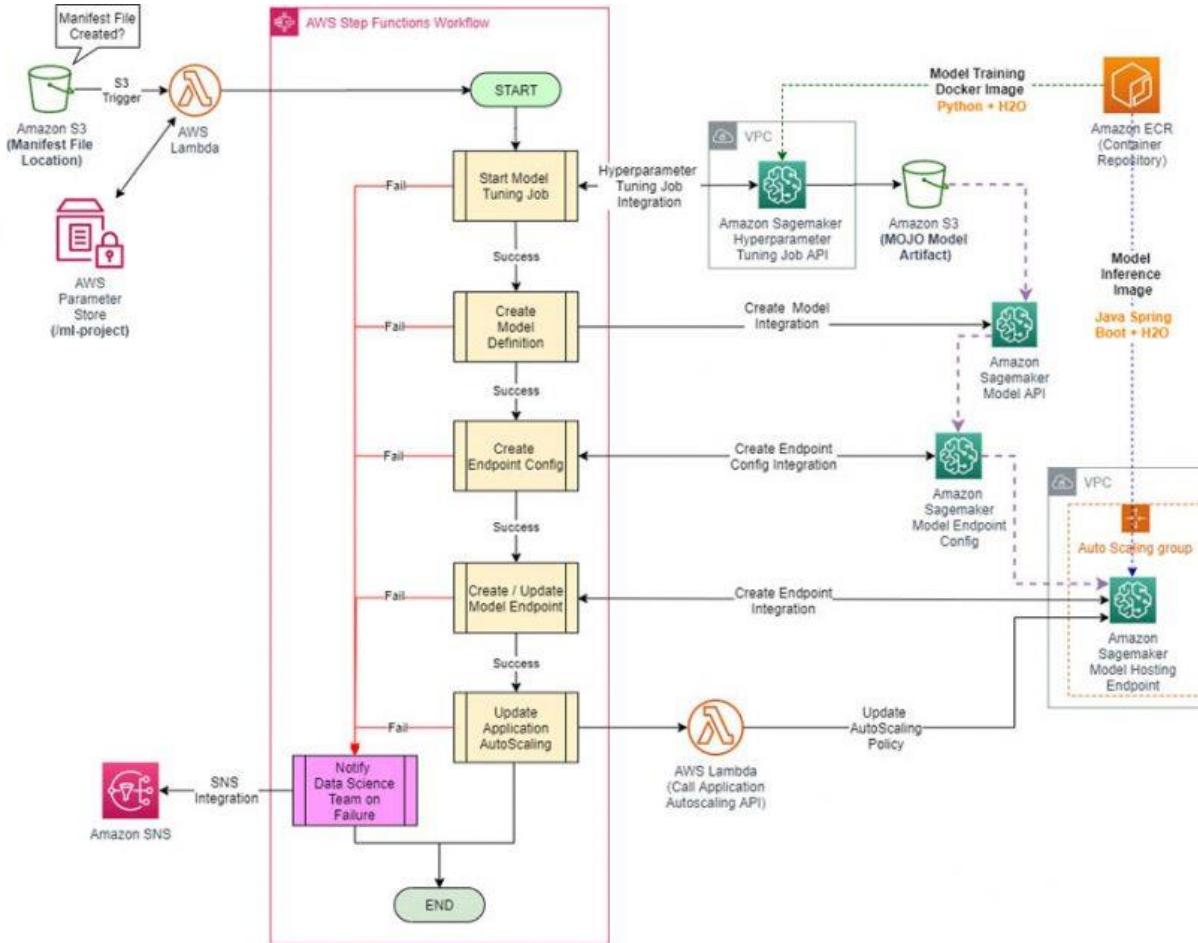
Construindo modelo customizado

- Feature Store



Construindo modelo customizado

- Sagemaker Endpoint + Pipeline



AWS ML Infraestrutura

Workflow Services



Amazon
SageMaker



Deep
Learning AMI



Deep Learning
Containers



AWS
Batch



AWS
ParallelCluster



Elastic
Kubernetes
Service



Elastic
Container
Service



Amazon
EMR

Frameworks



TensorFlow



PyTorch



mxnet



Keras



GLUON



Compute, Networking, and Storage



EC2 P4
instances



EC2 P3
instances



EC2 G4
instances



EC2 Inf1
instances



Elastic
Inference



AWS
Outposts



Elastic Fabric
Adapter



Amazon
S3



Amazon
EBS



Amazon
FSx



Amazon
EFS

The AWS ML Stack

Broadest and most complete set of Machine Learning capabilities

AI SERVICES



ML SERVICES



ML FRAMEWORKS & INFRASTRUCTURE



Broadest and most complete set of ML capabilities



Building your team's skills



Hands-on learning



AWS DeepRacer
reinforcement learning

AWS DeepComposer
Generative AI

AWS DeepLens
deep learning

AWS ML training
and certification

Partnerships
with MOOCs



Training and certification



AWS is uniquely equipped to help you achieve ML success

AI/ML challenges and needs

AWS advantage

Addressing business value



Proven use case and industry solutions to solve business problems

Reducing complexity



Amazon SageMaker helps data science teams improve productivity by 10x

Gaining experience and expertise



Amazon brings 20+ years of experience in ML and an extensive partner network to support 100K+ customers' ML journey

Lacking ML talent



Accelerate your journey with pre-trained models and access to AWS Training, AWS DeepRacer, and Amazon ML Solutions Lab

Implementing a data strategy



AWS can help you develop a clear data strategy to get the most out of your data

Addressing large compute and storage needs



Effectively address ML's storage and compute needs with AWS capabilities

Intrinsic Complexities with Machine Learning

- Understanding the business domain
- Selecting the appropriate Model
- Selecting the appropriate Features
- Fine tuning

Incidental Complexities with Machine Learning

- Integrating with Data Warehouse
- Scaling model training & serving
- Keeping consistency between: Prototyping vs Production,
Training vs Inference
- Keeping track of multiple models, versions, experiments
- Supporting iteration on ML models
 - ML models take on average 8 to 12 weeks to build
 - ML workflows tended to be slow, fragmented, and brittle

Thanks !



Vinicius Fernandes Caridá

vfcarida@gmail.com



@vinicius caridá



@vfcarida



@vinicius caridá



@vfcarida



@vinicius caridá



@vfcarida