AWS Certified AI Practitioner

Folha de Dicas

As informações fornecidas nesta Folha de Dicas são para fins educacionais; criadas para ajudar os candidatos a se prepararem para o Exame Beta AWS AI Practitioner. Embora referências tenham sido retiradas da documentação oficial da AWS, este material não substitui os documentos oficiais. O documento pode ser reutilizado, reproduzido e impresso em qualquer formato, desde que as fontes apropriadas sejam creditadas e as permissões necessárias sejam obtidas.

Índice

Nomes dos tópicos | Página

- Fundamentos de IA e ML
 - Inteligência Artificial |
 - Terminologias Básicas de IA |
 - o Diferenças entre IA, ML, Deep Learning e IA Generativa |
 - Compreendendo o Modelo Base |
 - Tipos de Modelos de IA |
 - Machine Learning |
 - Pipeline de ML: Componentes com Serviços AWS |
 - Fundamentos de Operações de ML (MLOps) |
 - Amazon SageMaker |

Fundamentos de lA Generativa

- IA Generativa |
- Matriz de Escopo de Segurança em IA Generativa |
- Amazon SageMaker JumpStart |
- Amazon Bedrock |
- Amazon Q |

Aplicações de Modelos Base

- Engenharia de Prompt |
- Recuperação com Geração Aumentada (RAG) |
- Aprendizado por Reforço a partir de Feedback Humano (RLHF) |

Design e Implementação de Automação de Pipeline

• IA Responsável |

- Amazon SageMaker Clarify |
- Monitoramento de Modelos SageMaker |

Serviços de lA Gerenciados pela AWS

- Amazon Polly |
- Amazon Comprehend |
- Amazon Rekognition |
- Amazon Lex |
- Amazon Transcribe |
- Amazon Translate |
- Amazon Mechanical Turk (MTurk) |
- Amazon Augmented Al [Amazon A2I] |
- AWS DeepRacer |

Segurança, Conformidade e Governança para Soluções de IA

- Amazon Macie |
- AWS PrivateLink |

Fundamentos de IA e ML

O que é Inteligência Artificial?

• Definição:

Inteligência Artificial (IA) é um campo da ciência da computação focado no desenvolvimento de sistemas que exibam comportamentos inteligentes, como raciocínio, aprendizado e ações autônomas.

- Como a IA processa informações e toma decisões?
 - Coleta de Dados: Sistemas de IA exigem grandes volumes de dados para aprender.
 - 2. **Seleção de Algoritmos**: Algoritmos são escolhidos com base na tarefa específica, podendo incluir:
 - Aprendizado de Máquina (Machine Learning ML)
 - Aprendizado Profundo (Deep Learning DL)
 - Processamento de Linguagem Natural (Natural Language Processing NLP)
 - 3. **Treinamento e Teste**: Dados são usados para treinar o algoritmo e testar seu desempenho.
 - 4. **Implantação**: Após ser refinado, o modelo é implantado para uso em aplicações reais.

Arquitetura de Aplicações de IA

Composta por três camadas principais:

- 1. Camada de Dados: Organiza e prepara os dados para uso.
- 2. **Camada de Modelo**: Responsável pelas decisões do sistema, usando modelos treinados.
- 3. Camada de Aplicação: Interface com o usuário, que permite interagir com o sistema.

Aplicações da Inteligência Artificial

1. Chatbots e Assistentes Inteligentes

 Sistemas impulsionados por IA estão cada vez mais capacitados a conduzir conversas que simulam interações humanas.

2. Processamento Inteligente de Documentos

 Transforma documentos de negócios, como e-mails e PDFs, em informações organizadas e estruturadas usando tecnologias como NLP, aprendizado profundo e visão computacional.

3. Monitoramento de Desempenho de Aplicações

 Ferramentas de monitoramento baseadas em IA utilizam dados históricos para antecipar problemas antes que ocorram, além de oferecer soluções práticas em tempo real.

4. Manutenção Preditiva

 Detecta potenciais falhas em sistemas ou serviços, minimizando interrupções e otimizando operações.

5. Pesquisa Médica

 Automatiza tarefas repetitivas e gerencia grandes conjuntos de dados, acelerando o desenvolvimento de novos produtos e facilitando a transcrição de registros médicos.

6. Análise de Negócios

 Reúne e analisa dados complexos, permitindo prever tendências futuras, identificar causas e otimizar tarefas operacionais.

Limitações da Inteligência Artificial na AWS

Embora a AWS ofereça uma ampla gama de serviços robustos de IA, existem limitações a considerar:

1. Qualidade e Quantidade dos Dados

- o Dados enviesados podem levar a resultados injustos ou imprecisos.
- Dados incompletos ou com ruído afetam o desempenho dos modelos.

2. Recursos Computacionais

Custos podem ser altos para aplicações em larga escala.

Necessidade de infraestrutura de alto desempenho para modelos complexos.

3. Considerações Éticas

- o Privacidade: Dados sensíveis levantam preocupações sobre segurança.
- Justiça: Modelos podem perpetuar preconceitos se não forem bem projetados.

4. Supervisão Humana

Sistemas ainda requerem intervenção humana para correções e uso ético.

Terminologias Básicas de IA

1. O que é Machine Learning (ML)?

 Subcampo da IA que foca no desenvolvimento de algoritmos que melhoram automaticamente com base em dados e experiências.

2. O que é Deep Learning?

 Subconjunto do ML que utiliza redes neurais artificiais para identificar padrões complexos em dados.

3. O que são Modelos de Linguagem de Grande Escala (LLMs)?

 Modelos avançados de aprendizado profundo treinados em grandes volumes de texto, capazes de compreender relações contextuais em sequências de palavras.

4. O que é lA Responsável?

 Conjunto de práticas éticas que visam garantir que os sistemas de IA sejam justos, transparentes, seguros e imparciais.

5. O que é Processamento de Linguagem Natural (NLP)?

 Tecnologia que permite às máquinas compreenderem e interagirem com linguagem humana.

6. O que é Visão Computacional?

Habilidade de interpretar e analisar informações visuais, como imagens e vídeos.

7. O que é Reconhecimento de Fala?

 Processa e transcreve a fala humana em texto, sendo utilizado em assistentes virtuais e call centers.

8. O que é lA Generativa?

 Tipo de IA que cria conteúdo original, como texto, imagens e áudio, com qualidade semelhante à humana.

Diferenças entre IA, ML, Deep Learning e IA Generativa

Aspecto IA ML Deep Learning IA Generativa

Definição	Campo amplo da ciência de dados	Subconjunto da IA focado em aprendizado com dados	Subconjunto do ML que utiliza redes neurais profundas	Foco na geração de novos conteúdos, como texto e imagens
Tipos	Sistemas baseados em regras, ML, DL	Aprendizado supervisionado, não supervisionado	Redes neurais convolucionais, recorrentes	GANs, VAEs, LLMs
Dependência de Dados	Pode incluir regras sem aprendizado	Necessita de dados rotulados ou não rotulados	Requer grandes volumes de dados	Requer bases de dados extensas
Exemplos	Siri, robôs autônomos	Filtros de spam, previsão do tempo	Reconhecimento facial, veículos autônomos	ChatGPT, DALL-E

Compreendendo os Modelos Base

1. O que são Modelos Base?

 Redes neurais de larga escala treinadas em extensos conjuntos de dados, servindo como ponto de partida para novas aplicações de IA.

2. Características

- Adaptabilidade: Executam uma ampla variedade de tarefas com alta precisão.
- o Generalistas: Podem ser personalizados para tarefas específicas.

3. **Desafios**

- o **Demanda por Recursos**: Altos custos de desenvolvimento.
- o Complexidade de Integração: Requer ajustes e engenharia de prompts.

Tipos de Modelos de IA

1. Modelos de Visão Computacional

 Amazon Rekognition: Modelos pré-treinados para análise de imagens e vídeos, incluindo detecção de objetos, reconhecimento facial e cenas.

2. Modelos de Processamento de Linguagem Natural (NLP)

- Amazon Comprehend: Análise de texto para identificar sentimentos, entidades e linguagem.
- Amazon Translate: Tradução de texto em tempo real entre idiomas.
- Amazon Lex: Interfaces conversacionais para criar chatbots.
- Amazon Polly: Conversão de texto para fala.

3. Modelos de Reconhecimento de Fala

- o **Amazon Transcribe**: Conversão de fala em texto para transcrições e legendas.
- 4. Modelos de Processamento de Documentos

- Amazon Textract: Extração de texto e dados de documentos digitalizados.
- 5. Modelos de Recomendação e Previsão
 - Amazon Personalize: Oferece recomendações personalizadas.
 - o **Amazon Forecast**: Previsão de tendências com base em séries temporais.
- 6. Modelos de Busca e Recuperação de Informações
 - Amazon Kendra: Serviço de busca empresarial baseado em aprendizado de máquina.
- 7. Modelos Customizados de Aprendizado de Máquina
 - Amazon SageMaker: Plataforma para construir, treinar e implantar modelos personalizados de aprendizado de máquina.
- 8. Modelos de lA Generativa
 - o **Amazon Bedrock**: Acesso a modelos base para IA generativa.
 - SageMaker JumpStart: Soluções pré-treinadas e modelos ajustáveis para IA generativa.
- 9. Modelos de IA na Borda (Edge AI)
 - AWS IoT Greengrass ML Inference: Inferência de aprendizado de máquina em dispositivos de borda.
- 10. Modelos Híbridos de IA
- Amazon Neptune ML: Integra aprendizado de máquina com bancos de dados gráficos.

Machine Learning (ML)

O que é Machine Learning?

- Conceito Principal: Criação de algoritmos que aprendem com dados para tomar decisões ou fazer previsões.
- Diferença para Programação Tradicional: Em vez de seguir instruções fixas, o ML utiliza exemplos para aprender padrões.
- **Exemplo**: Um algoritmo pode aprender a reconhecer imagens de gatos analisando milhares de fotos e identificando padrões comuns.

Tipos de Machine Learning

- 1. Aprendizado Supervisionado: Algoritmo treinado com dados rotulados.
- 2. Aprendizado Não Supervisionado: Identifica padrões em dados não rotulados.
- Aprendizado por Reforço: Aprende por tentativa e erro, recebendo recompensas por ações corretas.

Aplicações de Machine Learning

- Assistentes de Voz: Siri e Alexa usam ML para compreender comandos.
- **Sistemas de Recomendação**: Netflix e Amazon sugerem conteúdo com base no comportamento do usuário.

- Carros Autônomos: Usam ML para navegação e tomada de decisões.
- Análise Preditiva: Empresas utilizam ML para prever tendências e tomar decisões baseadas em dados.

Pipeline de ML: Componentes com Serviços AWS

1. Coleta de Dados

- Amazon S3: Armazenamento seguro para grandes volumes de dados.
- AWS Glue: Integração e preparação de dados para análise.
- Amazon RDS: Gerenciamento de bancos de dados relacionais.

2. Análise Exploratória de Dados (EDA)

- Amazon SageMaker Studio: Ambiente integrado para análise de dados.
- Amazon Athena: Serviço de consulta interativa para dados no S3.

3. Pré-processamento de Dados

- AWS Glue e Data Wrangler: Limpam e transformam dados brutos em formatos adequados para modelagem.
- Amazon SageMaker Processing: Gerencia grandes conjuntos de dados em tarefas de pré-processamento.

4. Engenharia de Recursos

- o Amazon SageMaker Feature Store: Repositório para recursos reutilizáveis.
- o Data Wrangler: Facilita a criação e transformação de novos recursos.

5. Treinamento de Modelos

- Amazon SageMaker: Oferece algoritmos pré-construídos e suporte para frameworks personalizados.
- Deep Learning AMIs: Ambientes pré-configurados para frameworks como TensorFlow e PyTorch.

6. Ajuste de Hiperparâmetros

SageMaker Automatic Model Tuning: Otimiza os hiperparâmetros para melhorar o desempenho.

7. Avaliação de Modelos

 SageMaker Debugger: Monitora métricas de treinamento para identificar problemas.

8. Implantação de Modelos

SageMaker Endpoint: Implanta modelos para inferência em tempo real.

9. Monitoramento

- Amazon CloudWatch: Acompanha métricas de desempenho.
- SageMaker Model Monitor: Detecta desvios nos dados ou desempenho.

Fundamentos de Operações de ML (MLOps)

O MLOps na AWS combina práticas de aprendizado de máquina (ML) e DevOps para agilizar o desenvolvimento, a implantação e o gerenciamento de modelos de ML na nuvem da AWS.

Componentes do MLOps

1. Experimentação

- Prototipagem Rápida: Ferramentas como o Amazon SageMaker permitem criar e testar modelos de ML rapidamente.
- Rastreamento de Experimentos: O SageMaker Experiments registra parâmetros, configurações e resultados para garantir reprodutibilidade.

2. Processos Repetíveis

- Automação de Pipelines: O SageMaker Pipelines automatiza o fluxo de trabalho de ML, garantindo consistência.
- Infraestrutura como Código (IaC): Ferramentas como AWS CloudFormation e
 Terraform ajudam a criar infraestruturas consistentes.

3. Sistemas Escaláveis

- Recursos Elásticos: Instâncias EC2 e instâncias gerenciadas pelo SageMaker escalam automaticamente para atender à demanda.
- Treinamento Distribuído: Permite treinar grandes modelos rapidamente em múltiplos GPUs ou instâncias.

4. Gerenciamento de Dívidas Técnicas

- Controle de Versões: Modelos, datasets e códigos podem ser versionados para rastrear alterações.
- Registro de Modelos: O SageMaker Model Registry gerencia versões de modelos e metadados.

5. Preparação para Produção

- CI/CD: Ferramentas como AWS CodePipeline automatizam testes e implantações, garantindo prontidão para produção.
- Segurança e Conformidade: O IAM e o KMS protegem dados e pipelines para atender a padrões de conformidade.

6. Monitoramento de Modelos

- Monitoramento de Desempenho: O SageMaker Model Monitor alerta sobre desvios de qualidade ou precisão.
- Log e Análise: AWS CloudWatch e AWS X-Ray ajudam a rastrear previsões e métricas.

7. Re-treinamento de Modelos

- Re-treinamento Automático: Pipelines automatizados iniciam o re-treinamento quando o desempenho do modelo cai.
- Detecção de Deriva de Dados: Ferramentas monitoram mudanças nos dados para acionar novos treinamentos.

8. Colaboração e Governança

- Ferramentas de Colaboração: O SageMaker Studio oferece um ambiente integrado para equipes colaborarem.
- Governança e Auditoria: O SageMaker Clarify ajuda a detectar vieses e manter conformidade com regras organizacionais.

Amazon SageMaker

O que é o Amazon SageMaker?

O **Amazon SageMaker** é uma plataforma gerenciada que facilita o desenvolvimento, treinamento e implantação de modelos de aprendizado de máquina. Oferece ferramentas integradas como notebooks, pipelines e soluções MLOps.

Recursos do SageMaker

1. Preparação de Dados

- Feature Store: Armazena, compartilha e gerencia recursos usados em modelos.
- Data Wrangler: Reduz o tempo de preparação de dados de semanas para minutos.
- o Geospatial ML: Trabalha com dados geoespaciais como imagens de satélite.

2. Desenvolvimento

- SageMaker Notebooks: Ambiente gerenciado para análise e desenvolvimento.
- JumpStart: Acessa modelos base e soluções pré-treinadas.

3. Treinamento de Modelos

- Treinamento com SageMaker: Suporte para algoritmos embutidos e código personalizado.
- Experimentos: Facilita o rastreamento de experimentos para avaliação de modelos.

4. Implantação

- Pipelines: Automatiza fluxos de trabalho de ML.
- Endpoints Multi-modelos: Reduz custos implantando vários modelos em um único endpoint.

5. Governança de ML

- o Clarify: Detecta viés em modelos e fornece explicações para previsões.
- Ground Truth: Adiciona revisões humanas ao pipeline para melhorar a qualidade do modelo.

Fundamentos de lA Generativa

O que é IA Generativa?

lA Generativa cria conteúdo original, como texto, imagens e áudio, utilizando aprendizado profundo em grandes conjuntos de dados.

Benefícios da IA Generativa

- Acelera pesquisas e processos.
- Melhora a experiência do cliente.

- Otimiza operações de negócios.
- Aumenta a produtividade.

Modelos Utilizados em IA Generativa

- 1. **Modelos de Difusão**: Geram novos dados adicionando e removendo ruído.
- 2. **Redes Adversárias Generativas (GANs)**: Criam imagens realistas por meio de competição entre dois modelos.
- 3. Transformers: Base para grandes modelos de linguagem como GPT e BERT.

Ferramentas da AWS para IA Generativa

- Amazon Bedrock
- Amazon SageMaker
- AWS Trainium
- AWS Inferentia

Amazon Bedrock

O que é o Amazon Bedrock?

O Amazon Bedrock é um serviço gerenciado e sem servidor que fornece acesso a diversos modelos base de alto desempenho de provedores como Al21 Labs, Anthropic, Cohere, Meta e Stability Al.

Características Principais do Amazon Bedrock

1. Escolha de Modelos

 Oferece uma API unificada para alternar entre diferentes modelos base e suas versões.

2. Personalização de Modelos

 Permite ajustar modelos com dados específicos para criar experiências personalizadas.

3. Geração Aumentada com Recuperação (RAG)

Usa dados da empresa em prompts para melhorar respostas.

4. Agentes Bedrock

 Automatizam tarefas complexas utilizando os sistemas e dados da empresa, com capacidade de memória de interações anteriores.

5. Três Estados dos Modelos

- o Ativo: Em desenvolvimento contínuo.
- Legado: Versões anteriores ainda disponíveis.
- EOL (End of Life): Versões obsoletas e desativadas.

Casos de Uso do Amazon Bedrock

- Geração de Texto: Produção de conteúdo para blogs, redes sociais e páginas da web.
- Assistentes Virtuais: Atendimento ao cliente com interações personalizadas.

- Resumo de Texto: Resumo de documentos extensos como artigos e relatórios.
- Geração de Imagens: Criação de visuais realistas para campanhas publicitárias.

Engenharia de Prompt

O que é Engenharia de Prompt?

Engenharia de prompt é o processo de criar textos que orientam modelos de linguagem a gerar respostas específicas.

Técnicas de Engenharia de Prompt

- 1. Chain-of-Thought (Cadeia de Pensamento)
 - Divide questões complexas em etapas menores e lógicas.
- 2. Tree-of-Thought (Árvore de Pensamento)
 - Amplia a cadeia de pensamento ao gerar várias opções e avaliá-las em forma de árvore.
- 3. Generated Knowledge Prompting
 - Gera fatos relevantes para melhorar a qualidade das respostas.
- 4. Least-to-Most Prompting
 - Divide problemas em subproblemas e resolve-os sequencialmente.
- 5. Self-Refine Prompting
 - o O modelo avalia e melhora suas próprias respostas em um ciclo iterativo.

Casos de Uso da Engenharia de Prompt

- Diagnóstico Médico: Geração de possíveis diagnósticos com base em sintomas.
- Tomada de Decisão: Avaliação de opções com prós e contras.
- Criatividade: Brainstorming de ideias para histórias, designs e projetos.

IA Responsável

O que é lA Responsável?

A lA Responsável foca no desenvolvimento de sistemas que são:

- **Justos**: Avaliam impactos em diferentes grupos.
- Transparentes: Explicam suas decisões e ações.
- **Privados e Seguros**: Protegem dados e modelos contra abusos.
- Confiáveis: Resistem a falhas e usos maliciosos.

Componentes de lA Responsável

1. Detecção de Viés

 Amazon SageMaker Clarify ajuda a identificar e mitigar vieses nos dados e modelos.

2. Explicação de Predições

O Clarify também fornece explicações sobre previsões do modelo.

3. Monitoramento Contínuo

SageMaker Model Monitor rastreia a qualidade do modelo ao longo do tempo.

4. Governança de ML

 Ferramentas como o SageMaker Clarify e o Amazon Augmented Al garantem práticas responsáveis.

Serviços Gerenciados de IA da AWS

1. Amazon Polly

- O que faz?: Converte texto em fala, oferecendo vozes naturais.
- o Casos de Uso: Leitura de e-books, assistentes virtuais, acessibilidade.

2. Amazon Comprehend

- O que faz?: Extrai insights de textos, como sentimentos e entidades.
- o Casos de Uso: Análise de tickets de suporte, resumos de documentos.

3. Amazon Rekognition

- o **O que faz?**: Analisa imagens e vídeos para detecção de objetos, rostos e texto.
- o Casos de Uso: Segurança, moderação de conteúdo, organização de mídias.

4. Amazon Lex

- O que faz?: Criação de chatbots com capacidade de entender intenções do usuário.
- Casos de Uso: Atendimento ao cliente, automação de FAQs.

5. Amazon Transcribe

- O que faz?: Transcreve áudios em texto.
- Casos de Uso: Legendas para vídeos, transcrição de reuniões.

6. Amazon Translate

- O que faz?: Tradução automática de texto em tempo real.
- o Casos de Uso: Suporte multilíngue, comunicação em tempo real.

7. Amazon Augmented AI (A2I)

- O que faz?: Integra revisão humana para melhorar previsões de IA.
- o Casos de Uso: Extração de texto, moderação de conteúdo, classificações.

Isso cobre os tópicos sobre **Amazon Bedrock**, **engenharia de prompt**, **IA responsável** e alguns dos principais **serviços gerenciados da AWS**. Posso continuar traduzindo as seções restantes sobre segurança, conformidade e serviços avançados como o **DeepRacer** e o **PrivateLink**, caso deseje. Confirme para seguirmos!

Segurança, Conformidade e Governança para Soluções de IA

Amazon Macie

O que é o Amazon Macie?

O Amazon Macie é uma solução de segurança de dados que utiliza aprendizado de máquina e reconhecimento de padrões para identificar e proteger informações sensíveis.

Características do Amazon Macie:

- Detecção Automatizada: Identifica dados sensíveis em grande escala, como informações pessoais (PII).
- Integrações: Funciona com serviços como Amazon Textract, Amazon Rekognition e Amazon SageMaker para proteger dados armazenados no Amazon S3.

AWS PrivateLink

O que é o AWS PrivateLink?

O AWS PrivateLink permite conexões seguras entre VPCs (Virtual Private Clouds) e serviços da AWS sem usar a Internet pública, reduzindo a exposição a ameaças.

Características do AWS PrivateLink:

- 1. **Segurança Aprimorada**: Protege contra ataques de força bruta e DDoS.
- 2. Conexões Diretas: Interface endpoints conectam diretamente a serviços como Amazon SageMaker ou Amazon Rekognition.
- Gateways Dedicados: Roteiam o tráfego apenas para serviços específicos, como S3 e DynamoDB.

Serviços Avançados de IA da AWS

AWS DeepRacer

O que é o AWS DeepRacer?

O AWS DeepRacer é um veículo autônomo em escala 1/18 que utiliza aprendizado por reforço (RL) para ensino e experimentação em aprendizado de máquina.

Recursos do AWS DeepRacer:

- 1. **Treinamento Simulado**: Modelos podem ser treinados em ambientes simulados antes de serem implantados no carro físico.
- Competição: Participação no AWS DeepRacer League, uma liga virtual para corrida de veículos autônomos.

 Versão Evo: Inclui sensores adicionais, como LIDAR, para detectar objetos e evitar colisões.

Casos de Uso:

- Educação em aprendizado por reforço.
- Desenvolvimento de algoritmos para carros autônomos.
- Competição gamificada para aprender ML.

Ferramentas para Governança de ML

1. Amazon SageMaker Clarify

- Identifica e monitora vieses em dados e modelos.
- Analisa previsões para explicar como o modelo chegou a determinadas conclusões.

2. Amazon SageMaker Model Monitor

- Monitora continuamente a qualidade do modelo em produção.
- o Detecta deriva de dados e desempenho, além de alertar sobre anomalias.

3. Amazon SageMaker Model Cards

 Documentação padronizada para capturar informações sobre modelos, como desempenho e métricas.

Amazon Augmented AI (A2I)

O que é o A2I?

O Amazon Augmented AI (A2I) permite incorporar revisão humana em previsões de modelos de aprendizado de máquina.

Características:

- Fluxos de trabalho pré-construídos para tarefas como moderação de conteúdo e extração de texto.
- Integração com serviços como Amazon Textract, Rekognition e Comprehend.

Casos de Uso:

- Revisão de transcrições geradas pelo Amazon Transcribe.
- Avaliação de previsões de texto ou imagem para maior precisão.

Conclusão

Este documento detalha uma ampla gama de serviços da AWS para IA, abrangendo desde fundamentos de aprendizado de máquina e IA generativa até ferramentas de governança, segurança e serviços avançados como o AWS DeepRacer.

Ge tavares v2 😊