

Guia de preparação AIB-C01

Tudo o que a prova cobra, organizado por domínio: conceitos-chave, comparações que confundem, pegadinhas reais, gatilhos do enunciado, questões comentadas e um plano para chegar pronto ao exame.

4 capítulos por domínio	**28** questões comentadas	**70%** nota de corte
6 provas completas no curso		

ANTES DE COMEÇAR

Como usar este guia

Este guia foi feito para ser estudado junto com as 6 provas completas do curso AIB-C01: o guia explica o que cai e onde estão as armadilhas; as provas treinam o ritmo e mostram o que você ainda não domina.

Quando	O que fazer
Semana 1	Faça o teste diagnóstico (página seguinte) e leia os capítulos 1 e 2. No fim da semana, faça a Prova 1 do curso no tempo do exame.
Semana 2	Leia os demais capítulos, começando pelos domínios em que você errou mais no diagnóstico. Faça as provas 2 e 3.
Semana 3	Releia os alertas de pegadinha e as tabelas de comparação. Faça as provas 4 e 5 e revise cada erro pela explicação.
Semana 4	Estude a folha de cola, faça a Prova 6 e refaça as provas abaixo de 80%. Agende o exame quando estiver acima de 80% em todas (a nota de corte é 70%).

Mapa do exame

Cada prova do curso tem 85 questões e 170 minutos, no formato do exame. Referência de aprovação: 70%.

Domínio (capítulo)	Peso	Neste guia
1. Estratégia de IA e Criação de Valor	28%	18 conceitos · 14 pegadinhas
2. Fundamentos e Letramento em IA	24%	18 conceitos · 14 pegadinhas
3. Governança e Liderança em IA Responsável	24%	18 conceitos · 14 pegadinhas
4. Prontidão, Liderança e Transformação com IA	24%	18 conceitos · 14 pegadinhas

Como ler os capítulos: os **tópicos-chave** são o que mais cai; a **tabela de memorização** separa o que a prova costuma confundir; os **alertas de prova** mostram as armadilhas das alternativas erradas; os **gatilhos** ligam palavras do enunciado à resposta; e o **checklist** diz se você está pronto para o domínio.

Estude um pouco todo dia, de graça

Instagram @strecht.simulados

Um post de prova por dia: questão comentada, pegadinhas e o "não confunda" de cada certificação.

Seguir no Instagram →

YouTube Strecht Simulados

Shorts diários com pegadinhas de prova e, aos sábados, um simulado comentado.

Inscrever-se no canal →

DIAGNÓSTICO

Você já sabe isso?

Responda às 12 questões sem consultar nada. As respostas, com a explicação e o capítulo para revisar, estão no fim do guia. Comece o estudo pelos capítulos em que errou.

D1 · Estratégia de IA e Criação de Valor

Um assistente de IA no site resolve dúvidas antes que o cliente abra um chamado. Qual KPI mostra diretamente o efeito dessa solução?

A. Dúvidas resolvidas sem abrir chamado
B. Quantidade de perguntas cadastradas na base
C. Número de visitas à página inicial do site
D. Tempo médio de carregamento das páginas

D2 · Estratégia de IA e Criação de Valor

Uma empresa planeja o portfólio de IA para os próximos dois anos. Qual composição tende a equilibrar resultados e inovação?

A. Só pequenas melhorias, sem nenhuma iniciativa de maior impacto
B. Ganhos rápidos e poucas apostas estratégicas maiores
C. Só apostas de longo prazo, sem entregas nos primeiros dois anos
D. Dezenas de iniciativas iniciadas ao mesmo tempo, sem priorização

D3 · Estratégia de IA e Criação de Valor

Uma empresa terá uso previsível e constante de recursos de computação para modelos de ML ao longo do próximo ano. Qual opção tende a reduzir o custo?

A. Contratar um Savings Plan com compromisso de uso
B. Usar somente preços sob demanda, sem compromisso
C. Aumentar o tamanho das instâncias para ganhar desempenho
D. Adiar a análise de custos para o fim do contrato

D4 · Fundamentos e Letramento em IA

Qual é a diferença entre treinar um modelo do zero e fazer ajuste fino (fine-tuning)?

A. O ajuste fino adapta um modelo já treinado com dados específicos
B. O treinamento do zero é sempre mais barato que o ajuste fino
C. Os dois termos significam a mesma coisa em projetos de IA
D. O ajuste fino cria um modelo novo sem usar nenhum dado de exemplo

D5 · Fundamentos e Letramento em IA

Um analista pede em um único prompt que o modelo leia um relatório, compare com o ano anterior, aponte riscos e redija um e-mail. O resultado vem confuso. Qual prática ajuda?

A. Repetir o mesmo prompt até que a resposta fique mais organizada
B. Dividir a tarefa em etapas, com instruções claras para cada uma
C. Pedir que o modelo responda em outro idioma para ganhar precisão
D. Retirar o relatório do prompt para deixar a instrução mais curta

D6 · Fundamentos e Letramento em IA

Uma empresa quer que um assistente criado no Amazon Bedrock responda com base nos documentos internos dela, usando RAG. Qual recurso do Bedrock atende a essa necessidade?

A. AWS Cost Explorer
B. Amazon Bedrock Guardrails
C. AWS Pricing Calculator
D. Amazon Bedrock Knowledge Bases

D7 · Governança e Liderança em IA Responsável

Uma empresa quer usar dados de clientes brasileiros em uma solução de IA generativa. Qual lei deve orientar as decisões sobre esses dados pessoais?

A. SOX
B. LGPD
C. HIPAA
D. PCI DSS

D8 · Governança e Liderança em IA Responsável

Um assistente de IA guarda os registros completos das conversas, com dados pessoais, sem prazo de exclusão. Qual medida é mais adequada?

A. Manter tudo guardado, pois os registros podem ser úteis no futuro
B. Enviar os registros aos usuários e apagar a cópia da empresa
C. Apagar os registros no mesmo dia, mesmo que sejam necessários à auditoria
D. Definir uma política de retenção com prazo e finalidade para os registros

D9 · Governança e Liderança em IA Responsável

Uma empresa treina um modelo próprio no Amazon SageMaker AI. Pelo modelo de responsabilidade compartilhada, quem deve avaliar se os dados de treinamento têm viés?

A. Os usuários finais que recebem as previsões
B. A AWS, que opera a infraestrutura do serviço
C. A empresa cliente, dona dos dados e do uso
D. O fornecedor do hardware usado no treinamento

D10 · Prontidão, Liderança e Transformação com IA

Um caso de uso de IA precisa de dados de vendas, mas 30% dos registros não têm o código do cliente. Qual aspecto da prontidão de dados está comprometido?

A. Propriedade dos dados
B. Qualidade dos dados
C. Volume dos dados
D. Acessibilidade dos dados

D11 · Prontidão, Liderança e Transformação com IA

Qual conjunto de métricas acompanha melhor uma iniciativa de IA ao longo do tempo?

A. Número de reuniões e de apresentações feitas
B. Adoção, qualidade, valor de negócio e custo
C. Páginas de documentação escritas pelo time
D. Quantidade de modelos testados pela equipe

D12 · Prontidão, Liderança e Transformação com IA

Quem deve participar de uma avaliação de prontidão da empresa para adotar IA?

A. A diretoria financeira, que aprova o orçamento
B. Líderes de negócio, tecnologia, dados, jurídico e pessoas
C. Um consultor externo, sem ouvir as áreas internas
D. A equipe de tecnologia, que conhece os sistemas

01

CAPÍTULO 1 · 28% DA PROVA

Estratégia de IA e Criação de Valor

É o domínio de maior peso (28%), então preciso dominar como transformar IA em resultado de negócio e provar isso: ROI, prazo de retorno, linha de base, KPIs, indicadores antecedentes e de resultado, benefícios tangíveis e intangíveis. Também caem construir, comprar ou firmar parceria, priorização por valor, viabilidade, sustentabilidade e alinhamento estratégico (escalar, pausar ou encerrar), quando a IA não é a solução, transição de processos e de plataformas, estruturas de preço da AWS, opções de preço do Amazon Bedrock, Pricing Calculator x Cost Explorer, AWS Marketplace, vantagem competitiva e transformação do modelo de negócio. As contas aparecem com distratoras que reproduzem erros típicos: dividir pelo benefício em vez do custo, esquecer custos recorrentes, não descontar o custo da IA. Meu maior erro é aceitar como KPI uma métrica de atividade (prompts, logins, sugestões geradas) em vez do efeito no processo.

Neste capítulo
18 tópicos-chave · 14 pegadinhas · 8 macetes · 14 gatilhos · 4 questões comentadas

Tópicos-chave

ROI de IA

ROI = (benefícios totais - custos totais) / custos totais. Benefícios: economia de tempo, redução de custos, crescimento de receita e ganhos de produtividade. Custos: licença ou consumo, inferência, integração, capacitação, governança e operação contínua. Em projetos com custo alto no primeiro ano e benefícios crescentes, calcule em vários anos incluindo os custos recorrentes.

Prazo de retorno (payback)

Tempo para os benefícios acumulados igualarem o investimento: investimento dividido pelo benefício líquido mensal. É complementar ao ROI e deve ser analisado junto com os custos recorrentes e os riscos do projeto.

Economia de tempo como valor

Calcula-se por horas economizadas x número de pessoas x custo por hora x período. Mas horas liberadas só viram ganho financeiro real quando o tempo é usado em atividades de valor ou reduz custos reais; confirmação do fornecedor ou registro no ponto não provam valor.

Linha de base

Medir os indicadores atuais antes da implantação para atribuir o efeito à IA, por exemplo tempo médio de atendimento e resolução no primeiro contato, ou tempo de análise, taxa de erro e custo por sinistro. Se a empresa lançou sem linha de base, reconstrói com histórico anterior ou áreas que ainda não usam IA. Comparar com um grupo de controle é a forma robusta de atribuir o ganho.

KPI de resultado x métrica de atividade

Bons KPIs medem o efeito direto da solução no processo e combinam eficiência e qualidade: dúvidas resolvidas sem abrir chamado, queda nas reatribuições e no tempo até a resolução, percentual de faturas processadas sem intervenção manual, tempo até a autorização com taxa de erro sob controle, conversão contra a linha de base. Número de prompts, logins, sugestões geradas, tokens, reuniões ou horas de treinamento medem atividade ou custo, não valor.

Benefícios tangíveis e intangíveis

Tangíveis: redução de custo, aumento de receita, redução de horas extras pagas. Intangíveis: satisfação de clientes, satisfação e produtividade dos funcionários, imagem da marca, confiança dos clientes. Intangíveis também se medem, com pesquisa antes e depois pelo mesmo método, e podem ser ligados a resultados como retenção e recompra.

Indicadores antecedentes (leading) e de resultado (lagging)

Antecedentes preveem o sucesso e permitem corrigir o rumo cedo: taxa de adoção, uso semanal, satisfação inicial dos usuários, tempo até os primeiros usuários obterem valor, participação nos treinamentos. De resultado confirmam o valor depois: aumento de receita no trimestre, ROI acumulado, lucro anual, participação de mercado. Uso semanal em queda após um pico antecipa risco de não entregar o valor previsto.

Critérios de priorização

O guia cita valor para o negócio, viabilidade, sustentabilidade e alinhamento estratégico, decidindo escalar, pausar ou encerrar. Na viabilidade, disponibilidade e qualidade dos dados costumam pesar mais. Sustentabilidade inclui pessoas, custos e capacidade de manter (só um especialista que vai sair é sinal fraco). Numa matriz valor x viabilidade, comece pelos de alto valor e alta viabilidade.

Escalar, pausar ou encerrar

Escalar quando há metas atingidas, boa adoção, custo previsível e patrocínio. Pausar a expansão diante de grande migração de dados ou regulação em consulta pública, mantendo o que já funciona. Encerrar e registrar os aprendizados quando o piloto não atinge metas e o problema perdeu prioridade, ou quando o processo apoiado será descontinuado, mesmo com metas técnicas atingidas. Critérios de sucesso e de encerramento devem ser combinados com o patrocinador antes do piloto.

Quando a IA não é a solução

Regras claras e estáveis (reembolso dentro do limite do cargo, relatório fixo com três indicadores, crédito com regras objetivas exigidas pela regulação) pedem automação tradicional. Se a demora vem do processo (três assinaturas para valores baixos), revise o processo antes. Sem dados relevantes, baixo volume com muita variedade ou sem base legal para os dados, a IA não se justifica; em decisões raras de altíssimo impacto, a IA apoia a análise e pessoas decidem.

Construir, comprar ou firmar parceria

Fatores citados no guia: orçamento, prazos, capacidades, propostas de fornecedores e requisitos regulatórios. Construir faz sentido quando o caso é diferencial competitivo e há dados e equipe para manter. Comprar é melhor para necessidades comuns, prazo curto e pouca equipe. Parceria com especialista serve a domínios regulados sem capacidade interna, mantendo a governança interna e com transferência de conhecimento para não criar dependência.

Avaliação de propostas de fornecedores

Critérios: aderência ao caso de uso, segurança e evidências de resultado, com teste em documentos reais e critérios de sucesso definidos. Compare pelo custo total de propriedade (licença ou consumo, integração, operação, suporte e treinamento), não só pela licença. Cláusula para o fornecedor treinar modelos com os dados da empresa se negocia ou se troca de opção, e em setor regulado o país de processamento dos dados deve ser conhecido e permitido.

Estruturas de preço de IA na AWS

Baseada em consumo: paga pelo volume processado, como tokens; varia com o uso e favorece demanda incerta. Baseada em instância: paga pelo tempo em que os recursos dedicados ficam ativos, como um endpoint do Amazon SageMaker AI, então recurso ocioso gera custo. Por assento: valor fixo por usuário ao mês, previsível e comum em ferramentas de produtividade.

Opções de preço do Amazon Bedrock e Savings Plans

Sob demanda: pago por token, flexível, sem reservar capacidade. Processamento em lote: para grandes volumes cujo resultado pode esperar (usado no dia seguinte), pode ter preço menor que o sob demanda. Throughput provisionado: capacidade reservada para volume alto e constante. Para uso de computação previsível e constante ao longo do ano, Savings Plans com compromisso de uso reduzem o custo frente ao sob demanda.

Controles de custo

Estimar antes com a AWS Pricing Calculator (projetando volume de uso e tokens) e acompanhar o gasto real com o AWS Cost Explorer. Identificar custos por área com etiquetas para ratear, definir limites de uso e alertas, desligar recursos ociosos e usar métricas unitárias, como custo por atendimento resolvido. Em IA generativa, as alavancas são o volume de tokens de entrada e saída e a escolha de um modelo adequado à tarefa.

Transição de processos e de plataformas

O guia cita continuidade do negócio, implicações de custo, prontidão dos dados e impacto no desempenho. Rode a IA em paralelo ao processo atual e compare, com plano de transição e contingência, e verifique antes se há histórico suficiente e com qualidade. Ao trocar de modelo de fundação, o custo subestimado é reescrever e testar de novo os prompts e as avaliações; para evitar dependência, garanta que dados, prompts e testes possam ser levados a outra solução.

Vantagem competitiva sustentável

Vem de dados proprietários, integração da IA aos processos e ciclos em que o uso gera dados que melhoram o produto, algo difícil de copiar. Usar o mesmo modelo de fundação dos concorrentes, comprar licenças ou lançar um chatbot genérico primeiro dá vantagem passageira. IA integrada ao sistema que as pessoas já usam aparece no momento da tarefa e aumenta o uso real.

Nível de investimento conforme maturidade do setor

Setor em que a IA é nova: investir de forma seletiva em casos que gerem diferencial; ser pioneiro dá aprendizado e dados mais cedo, mas o risco é investir alto em algo que o mercado pode não valorizar; seguir os pioneiros reduz a incerteza. Setor em que os concorrentes já usam IA amplamente: investir para alcançar a paridade e buscar diferenciais, porque ficar parado custa competitividade em preço, prazo e experiência. Nicho que valoriza atendimento humano: investimento moderado em eficiência interna e apoio aos atendentes.

Macetes para memorizar

💡 ROI

Lucro sobre o que gastei: (benefício - custo) / custo. Se a conta não subtraiu o custo ou dividiu pelo benefício, é distratora.

💡 Payback

Investimento dividido pelo ganho líquido de cada mês = meses para se pagar.

💡 Priorização

V-V-S-A decide E-P-E: Valor, Viabilidade, Sustentabilidade e Alinhamento decidem Escalar, Pausar ou Encerrar.

💡 Construir, comprar ou parceria

Constrói só o que diferencia e consegue manter; o resto compra; o que é regulado e desconhecido faz em parceria.

💡 Ferramentas de custo

Calculadora calcula o futuro; Explorer explora o passado.

💡 Estruturas de preço

Consumo = conta que varia com o uso; Assento = mensalidade por pessoa; Instância = taxímetro ligado, cobra mesmo parado.

💡 Opções do Bedrock

Pode esperar? Lote. Volume alto e constante? Provisionado. Incerto? Sob demanda.

💡 KPI

KPI mede o efeito, não o esforço: se conta cliques, prompts ou reuniões, está errado.

Tabela de memorização: não confunda

Quando é...	...e quando é
AWS Pricing Calculator Estimar custos futuros antes do uso, por exemplo no caso de negócio de um assistente que ainda não existe ou ao projetar o piloto para 10 mil usuários.	**AWS Cost Explorer** Analisar o gasto já realizado: quais serviços mais pesaram, evolução mês a mês, recursos ociosos e desvio frente à estimativa.
Preço por consumo Uso variável e difícil de prever, sem pagar por capacidade parada; atenção, com uso intenso pode sair mais caro que por usuário.	**Preço por assento** Muitos usuários diários e o financeiro quer custo mensal previsível por pessoa.
Processamento em lote no Amazon Bedrock Grandes volumes cujo resultado só é usado depois, como resumir milhões de documentos para o dia seguinte; pode ter preço menor que o sob demanda.	**Throughput provisionado no Amazon Bedrock** Volume alto e constante que exige capacidade garantida e resposta imediata.
Indicador antecedente (leading) Prever cedo e corrigir o rumo: adoção, uso semanal, satisfação inicial, tempo até o primeiro valor.	**Indicador de resultado (lagging)** Confirmar o valor gerado depois: receita do trimestre, ROI acumulado, lucro anual, participação de mercado.

Construir Caso de uso é o núcleo do diferencial competitivo e a empresa tem dados e equipe técnica para construir e manter.	**Comprar ou firmar parceria** Necessidade comum no mercado, prazo curto ou pouca equipe (comprar); domínio especializado e regulado sem capacidade interna (parceria com especialista, mantendo a governança).
Melhoria operacional A IA torna a operação mais eficiente sem mudar a forma de gerar receita: leitura de notas fiscais, otimização de rotas, processamento de pedidos 30% mais rápido.	**Transformação do modelo de negócio** A IA muda a forma de cobrar ou cria nova oferta: preço por comportamento ao volante, cobrança por tarefa concluída por agentes, manutenção preditiva vendida por assinatura, tutoria personalizada como novo serviço.
Benefício tangível Medido diretamente em valores: redução do custo operacional, aumento da receita de vendas, redução de horas extras pagas.	**Benefício intangível** Satisfação de clientes e funcionários, imagem da marca, confiança dos clientes; medido com pesquisa antes e depois e ligado a retenção e recompra.

Números que caem

Item	Valor
Peso do domínio na prova	**28% (o maior)**
Fórmula do ROI	**(benefícios totais - custos totais) / custos totais**
ROI com R$ 600 mil de benefício e R$ 400 mil de custo	**50%**
ROI com R$ 500 mil de benefício e R$ 200 mil de custo por ano	**150%**
ROI em 2 anos: R$ 500 mil de implantação, R$ 100 mil/ano de operação e R$ 400 mil/ano de benefício	**cerca de 14% ((800 - 700) / 700)**
Payback de R$ 1,2 milhão com benefício líquido de R$ 100 mil por mês	**12 meses**
Payback de R$ 240 mil com economia líquida de R$ 40 mil por mês	**6 meses**
Economia: 100 mil contatos, 40% para a IA, R$ 8 humano x R$ 2 IA	**R$ 240 mil por mês**
Economia: 10 mil erros/mês a R$ 50 cada, redução de 60%	**R$ 300 mil por mês**
Benefício anual: 2 h/semana x 500 pessoas x R$ 50/h x 48 semanas	**R$ 2,4 milhões**
Horas liberadas: atendimento de 20 para 12 minutos em 30 mil atendimentos/mês	**4.000 horas por mês**

Alertas de prova (pegadinhas)

⚠ **Alerta de prova:** Calcular o ROI como benefício dividido pelo custo (600 / 400 = 150%).
Como não cair: Isso é a razão benefício/custo. ROI desconta o custo antes de dividir: (600 - 400) / 400 = 50%.

⚠ **Alerta de prova:** Dividir o ganho líquido pelo benefício (200 / 600 = 33%; 300 / 500 = 60%).
Como não cair: O denominador do ROI é sempre o custo total, nunca o benefício.

⚠ **Alerta de prova:** No ROI de dois anos, ignorar os custos de operação (800 - 500) / 500 = 60%.
Como não cair: Some implantação e operação do período: (800 - 700) / 700 = cerca de 14%. Esquecer custos recorrentes é erro comum.

⚠ **Alerta de prova:** Na economia de atendimento, usar o custo humano cheio (R$ 320 mil) ou supor que todos os contatos migram (R$ 600 mil).
Como não cair: Economia = volume que realmente migra x diferença de custo por contato: 40 mil x (R$ 8 - R$ 2) = R$ 240 mil.

⚠ **Alerta de prova:** Considerar toda hora liberada como economia direta.
Como não cair: Só há ganho financeiro real se o tempo for usado em atividades de valor ou reduzir custos reais.

⚠ **Alerta de prova:** Escolher como KPI o número de prompts, sugestões geradas, logins, tokens ou licenças compradas.
Como não cair: Isso mede atividade, acesso ou custo. O KPI certo mede o efeito direto no processo, de preferência eficiência com qualidade.

⚠ **Alerta de prova:** Supor que cobrança por consumo sempre sai mais barata que por usuário.
Como não cair: Com uso intenso, consumo pode custar mais. Estime o uso real e compare o custo nos dois modelos antes da troca.

⚠ **Alerta de prova:** Estimar o custo em escala multiplicando o custo do piloto por cem ou usando o Cost Explorer para um assistente que ainda não existe.
Como não cair: Projete o volume de uso revendo as premissas e simule na AWS Pricing Calculator; o Cost Explorer só analisa gastos realizados.

⚠ **Alerta de prova:** Escalar porque a adoção é alta, mesmo com custo por uso maior que o valor gerado.
Como não cair: Escalar multiplica o prejuízo. Otimize custo e valor antes de expandir: a economia unitária precisa fechar.

⚠ **Alerta de prova:** Manter ou ampliar um piloto porque já houve muito investimento.
Como não cair: Investimento passado é custo afundado, não critério. Decida por valor, viabilidade, sustentabilidade e alinhamento estratégico.

⚠ **Alerta de prova:** Escalar para produção porque as metas técnicas foram atingidas, embora o processo vá ser descontinuado.
Como não cair: Alinhamento estratégico pesa tanto quanto o desempenho técnico. Encerre e registre o aprendizado; pausar também é errado, porque a descontinuação não é temporária.

⚠ **Alerta de prova:** Começar pelo caso de maior valor mesmo sem dados, porque o valor potencial é sempre o critério principal.
Como não cair: Sem dados o caso não é viável agora. Comece pelo caso viável e prepare os dados do outro em paralelo.

⚠ **Alerta de prova:** Achar que comprar pelo AWS Marketplace isenta custos de integração, transfere o código-fonte ou dispensa avaliação de riscos.
Como não cair: O benefício típico é compra e cobrança integradas à fatura da AWS e facilidade para encontrar e comparar soluções prontas; integração e avaliação de riscos continuam com a empresa.

⚠ **Alerta de prova:** Aceitar a demonstração com documentos de exemplo do fornecedor ou a lista de clientes famosos como prova dos 99% de acerto.
Como não cair: Peça um teste com documentos reais da empresa e critérios de sucesso definidos.

Como cai na prova

📝 **Como cai:** Conta de ROI, payback, economia mensal ou horas liberadas, com quatro valores numéricos.
Como resolver: Escreva a fórmula, some todos os custos do período e desconte o custo da IA. Confira se a sua resposta não coincide com um erro típico: razão benefício/custo, divisão pelo benefício, custo total sem a redução, casa decimal trocada.

📝 **Como cai:** Qual KPI mostra melhor o valor de uma solução descrita.
Como resolver: Pergunte o que a solução muda no processo e escolha o indicador desse efeito (tempo, conversão, resolução, erro), comparado à linha de base. Descarte volume de uso, quantidade gerada e esforço de preparação.

📝 **Como cai:** Cenário com prazo, equipe, diferencial e regulação, pedindo construir, comprar ou firmar parceria.
Como resolver: Diferencial + dados + equipe = construir. Comum + prazo curto + pouca equipe = comprar. Regulado + sem especialistas = parceria mantendo a governança. Adiar e construir do zero sem capacidade quase sempre são distratoras.

📝 **Como cai:** Situação de um piloto ou iniciativa e pergunta: escalar, pausar ou encerrar.
Como resolver: Metas + adoção + custo previsível + patrocínio = escalar de forma planejada. Incerteza temporária (migração de dados, regulação em consulta) = pausar a expansão. Sem valor ou sem alinhamento estratégico = encerrar e registrar aprendizados.

📝 **Como cai:** Perfil de uso e pergunta pela estrutura de preço, opção do Bedrock ou ferramenta de custo.
Como resolver: Previsibilidade por pessoa = assento; variável = consumo; recurso dedicado = instância; constante por um ano = Savings Plans; resultado no dia seguinte = lote; capacidade garantida = provisionado; antes = Pricing Calculator; depois = Cost Explorer.

📝 **Como cai:** Proposta de usar IA em um problema e pergunta qual a melhor recomendação.
Como resolver: Antes de aceitar IA, procure sinais de que ela não serve: regra fixa, causa no processo, falta de dados relevantes, baixo volume, sem base legal, decisão rara e de altíssimo impacto.

Gatilhos do enunciado

Se o enunciado trouxer...	Pense em
Estimar o custo mensal antes de construir, com base no volume esperado	**AWS Pricing Calculator**
Quais serviços mais pesaram na fatura e como o gasto evoluiu mês a mês	**AWS Cost Explorer**
Uso de computação previsível e constante ao longo do próximo ano	**Savings Plans com compromisso de uso**
Milhões de documentos por mês, resultado usado só no dia seguinte	**Processamento em lote no Amazon Bedrock**
Volume alto e constante no Bedrock, precisa de capacidade garantida	**Throughput provisionado**
Endpoint dedicado ativo 24 horas com poucas requisições / fatura alta após o fim do piloto	**Preço baseado em instância; desligar recursos ociosos**
Encontrar e comparar soluções prontas de terceiros que funcionam na AWS	**AWS Marketplace**

Política simples, regras fixas, relatório com indicadores fixos	**Regras de negócio ou automação tradicional; IA não é a solução**
Caso comum no mercado, prazo curto, pouca equipe técnica	**Comprar solução pronta (ou modelo de fundação gerenciado no Amazon Bedrock)**
Regulação rígida, sem especialistas internos	**Parceria com fornecedor especializado, mantendo a governança interna**
Assistente pronto que responda com base nos manuais da empresa	**Modelo pronto com RAG sobre os manuais**
Provar o ganho de produtividade de uma implantação futura	**Medir a linha de base antes**
Cobrança por tarefa concluída, serviço por assinatura, nova linha de receita com IA	**Transformação do modelo de negócio**
Proposta com licença barata x proposta que inclui integração, suporte e treinamento	**Custo total de propriedade**

🎯 Se você lembrar de só uma coisa deste capítulo...

- ROI divide pelo custo, e o custo inclui integração, capacitação, governança e operação contínua.
- Sem linha de base não há como provar valor; grupo de controle ou histórico reconstroem a comparação.
- KPI mede efeito no processo; adoção e satisfação inicial são antecedentes; receita e ROI são de resultado.
- Pricing Calculator estima antes, Cost Explorer analisa depois, Marketplace compra soluções de terceiros.
- Regras fixas, processo mal desenhado, falta de dados ou de base legal: a IA não é a resposta.

Essenciais para a prova (checklist)

☐ Sei calcular ROI, payback, economia por contato e horas liberadas sem cair nas distratoras numéricas.

☐ Sei escolher o KPI que mede o efeito direto de uma solução e descartar métricas de atividade.

☐ Sei separar indicadores antecedentes de indicadores de resultado e benefícios tangíveis de intangíveis.

☐ Sei aplicar os critérios valor, viabilidade, sustentabilidade e alinhamento estratégico para escalar, pausar ou encerrar.

☐ Sei reconhecer quando a IA não é a solução adequada.

☐ Sei decidir entre construir, comprar ou firmar parceria pelos fatores orçamento, prazos, capacidades, propostas e regulação.

☐ Sei associar consumo, instância e assento, Savings Plans e as opções sob demanda, lote e throughput provisionado do Bedrock ao perfil de uso.

☐ Sei quando usar AWS Pricing Calculator, AWS Cost Explorer e AWS Marketplace.

☐ Sei diferenciar melhoria operacional de transformação do modelo de negócio e de onde vem vantagem competitiva sustentável.

☐ Sei definir o nível de investimento conforme a maturidade do setor e a dinâmica competitiva.

Questões de revisão do capítulo

R1.1 · Estratégia de IA e Criação de Valor

Um projeto de IA vai atender duas áreas com volumes diferentes. Como ratear o custo de forma justa e controlar o gasto de cada uma?

A. Dividir o custo total em partes iguais, sem considerar o uso de cada área
B. Cobrar o custo total da área que pediu o projeto primeiro
C. Identificar os custos por área com etiquetas e acompanhar os relatórios
D. Deixar o custo fora do orçamento das áreas, como despesa geral

R1.2 · Estratégia de IA e Criação de Valor

Um gerente quer usar IA para aprovar reembolsos de viagem que seguem uma política simples: valor até o limite do cargo e comprovante válido. Qual é a melhor recomendação?

A. Treinar um modelo de ML com os reembolsos aprovados nos últimos anos
B. Implantar um agente de IA que negocie exceções com os funcionários
C. Usar regras de negócio, pois o problema não exige aprendizado com dados
D. Usar IA generativa para interpretar cada comprovante com mais flexibilidade

R1.3 · Estratégia de IA e Criação de Valor

A área de TI passou a usar IA para direcionar automaticamente os chamados à equipe certa. Qual KPI mostra melhor o valor dessa mudança?

A. Número de chamados abertos pelos usuários no mês
B. Horas de treinamento da equipe de suporte
C. Quantidade de categorias criadas no sistema
D. Queda nas reatribuições e no tempo até a resolução

R1.4 · Estratégia de IA e Criação de Valor

Antes de migrar a triagem de chamados para um modelo de IA, qual verificação sobre os dados é mais importante?

A. Se há histórico de chamados suficiente, com qualidade e categorias corretas
B. Se a equipe de triagem aceita trabalhar com o novo sistema desde o primeiro dia
C. Se o volume de chamados é pequeno o bastante para ser revisado em planilha
D. Se os chamados antigos podem ser apagados depois que o modelo for treinado

Respostas comentadas

R1.1. C) Identificar os custos por área com etiquetas e acompanhar os relatórios

Por quê: O guia do exame pede controles básicos de custo em implantações de IA. Saber quem consome o quê é a base para otimizar e responsabilizar.

✗ A) Incorreta. Divisão igual ignora que os volumes de uso são diferentes.
✗ B) Incorreta. Cobrar só de uma área distorce os incentivos e o controle.
✗ D) Incorreta. Sem atribuição, nenhuma área se responsabiliza pelo gasto.

> **Dica de prova:** Custo por área: identifique e acompanhe.

R1.2. C) Usar regras de negócio, pois o problema não exige aprendizado com dados

Por quê: O guia do exame inclui a habilidade de determinar quando a IA não é a solução adequada. Problemas com regras claras e estáveis são mais bem atendidos por automação tradicional.

✗ A) Incorreta. Treinar um modelo para reproduzir uma regra conhecida gera custo e risco sem ganho.
✗ B) Incorreta. Um agente que negocia exceções aumenta a complexidade e o risco de decisões fora da política.
✗ D) Incorreta. Flexibilidade de interpretação não é desejável em uma política objetiva e pode gerar decisões inconsistentes.

> **Dica de prova:** Nem todo problema precisa de IA. Regras claras pedem automação simples.

R1.3. D) Queda nas reatribuições e no tempo até a resolução

Por quê: O guia do exame pede definir KPIs que mostrem o efeito da IA no processo. Em operações de TI, precisão do direcionamento e tempo de resolução são indicadores centrais.

✗ A) Incorreta. O número de chamados abertos não depende do direcionamento.
✗ B) Incorreta. Horas de treinamento são esforço, não resultado.
✗ C) Incorreta. Mais categorias não indicam melhor direcionamento.

> **Dica de prova:** Roteamento com IA: menos reatribuições, resolução mais rápida.

R1.4. A) Se há histórico de chamados suficiente, com qualidade e categorias corretas

Por quê: O guia do exame cita continuidade do negócio, custos, prontidão dos dados e impacto no desempenho como considerações na transição de processos para IA. Os dados vêm primeiro, pois sustentam o modelo.

✗ B) Incorreta. A adesão da equipe importa, mas é uma questão de mudança organizacional, não de dados.

✗ C) Incorreta. Um volume pequeno a ponto de caber em planilha indicaria pouca necessidade de IA.

✗ D) Incorreta. Apagar o histórico prejudica auditoria, retreinamento e comparação de resultados.

Dica de prova: Migrar processo para IA: confira antes se os dados estão prontos.

02

CAPÍTULO 2 · 24% DA PROVA

Fundamentos e Letramento em IA

Aqui preciso falar a língua da IA como gestor, sem programar: algoritmo, modelo, treinamento, inferência e previsão; IA x ML x IA generativa; dados estruturados x não estruturados; qualidade e rótulos; quando regras bastam e quando a IA se justifica; agentes e suas capacidades; drift e monitoramento; shadow AI e classificação de ferramentas; engenharia de prompt, tokens, janela de contexto, RAG e ajuste fino; Amazon Bedrock, Amazon SageMaker AI e Amazon Quick em nível estratégico; e as normas ISO/IEC 23053 e 42001. A prova traz cenários curtos de negócio que pedem o conceito certo ou a técnica mais adequada. Onde mais erro: trocar RAG por ajuste fino, achar que temperatura ou modelo maior resolve qualquer problema, acreditar que o modelo se corrige sozinho e usar IA onde uma regra fixa resolve.

Neste capítulo

18 tópicos-chave · 14 pegadinhas · 8 macetes · 14 gatilhos · 4 questões comentadas

Tópicos-chave

Algoritmo, modelo, treinamento, inferência e previsão

Algoritmo é o método que o computador segue para aprender padrões a partir de dados. Treinamento é a fase em que o algoritmo ajusta o modelo com exemplos históricos; o modelo é o resultado desse treinamento; inferência é o uso do modelo com dados novos para gerar previsões. Em produção, a inferência é a maior parte do custo recorrente, e o treinamento só se repete em ciclos de atualização.

IA, machine learning e IA generativa

IA é o campo mais amplo, de sistemas que executam tarefas que exigiriam inteligência humana. ML é o subconjunto que aprende padrões com dados; regras escritas uma a uma por especialistas não são ML. IA generativa usa modelos de ML treinados em grandes volumes para criar conteúdo novo (texto, resumo, imagem, código), enquanto o ML tradicional prevê, classifica ou estima valores.

Tipos de problema de ML

Classificação prevê categorias (fraude ou legítima, vai ou não cancelar) e exige exemplos rotulados. Regressão prevê um valor numérico (quanto o cliente vai gastar em reais). Agrupamento (clustering) forma grupos sem rótulos prévios, como segmentar clientes com hábitos parecidos. Recomendação sugere itens com base no histórico de uso ou de compras.

Dados estruturados e não estruturados

Estruturados seguem um esquema de campos definido, como tabelas e planilhas, e são a base de muitos modelos preditivos. Não estruturados são textos livres, e-mails, imagens, áudios e vídeos, e exigem processamento de linguagem natural, visão computacional ou transcrição. O local de armazenamento, o formato do arquivo, ter data e hora ou ter sido digitalizado não tornam um dado estruturado; com a IA generativa, contratos e e-mails ganharam valor porque os modelos leem, resumem e extraem informações deles.

Categorias de ferramentas de IA

O perfil do candidato inclui reconhecer mecanismos de recomendação, processamento de linguagem natural (classificar e-mails por assunto, extrair nomes e datas, análise de sentimento), visão computacional (fotos de danos, defeitos, rostos), extração de dados de documentos (renda e datas em comprovantes PDF) e IA para operações com clientes. Para analisar ligações das quais só existem áudios, o primeiro passo é transcrição de fala para texto. Modelo que recebe foto e texto e responde em texto é multimodal.

Qualidade dos dados

Dados incompletos, duplicados, inconsistentes ou enviesados levam a previsões ruins (garbage in, garbage out); mais dados ruins não corrigem e nenhum modelo compensa totalmente. Os dados precisam ser precisos e consistentes nos campos importantes e representar os casos reais de uso: um modelo treinado só com lojas de capitais prevê pior em cidades pequenas, sem avisar. Qualidade se constrói na origem, com responsáveis e regras de validação nos sistemas.

Dados rotulados e avaliação do modelo

Rótulos são exemplos que já trazem a resposta certa, como e-mails marcados como spam; um modelo supervisionado não aprende sem eles, então se nada foi classificado, rotula-se uma amostra com especialistas. Critérios diferentes de rotulagem entre equipes ensinam sinais contraditórios. Parte dos dados é separada para medir o desempenho com dados que o modelo não viu; acertar quase tudo no treino e errar muito com dados novos é sobreajuste: o modelo decorou e não generaliza.

Drift e monitoramento contínuo

Treinar com dados históricos pressupõe que o futuro se parece com o passado. Quando o comportamento real muda (novo meio de pagamento instantâneo, mudança de hábito dos consumidores), os dados atuais diferem dos de treino e o desempenho cai: é drift, corrigido monitorando e retreinando com dados recentes. Modelos não se retreinam sozinhos, não expiram e não se corrigem com o uso, então deve haver um responsável definido com processo de monitoramento e atualização; no SageMaker AI, o Model Monitor ajuda nisso.

Previsões são probabilísticas

Uma probabilidade de 80% de compra é uma estimativa que pode errar e deve apoiar a decisão junto com outros critérios. Todo modelo erra de dois jeitos: deixa passar o que deveria pegar e pega o que deveria deixar passar. Definir o ponto de equilíbrio, por exemplo entre custo da fraude e custo de bloquear bons clientes, é decisão do negócio, não da equipe técnica nem do fornecedor.

Automação por regras x IA

Regras são adequadas quando a lógica é conhecida e estável e o resultado precisa ser determinístico e auditável: alíquota definida em lei, frete por tabela de peso, multa contratual, limite de valor. IA é indicada para padrões complexos ou que mudam, texto livre em grande volume e imagens ou áudios. Muitas soluções combinam as duas: regras nos casos claros e IA com revisão nos ambíguos.

Agentes de IA

Capacidades centrais citadas no guia: autonomia (planejar e executar passos para um objetivo com pouca intervenção humana), uso de ferramentas (consultar estoque e registrar pedido no ERP), comunicação entre agentes (agente de vendas pede análise ao agente de crédito) e orquestração (um orquestrador coordena, distribui tarefas e combina resultados). O que o agente pode fazer é definido pelas ferramentas e permissões configuradas. Agir nos sistemas exige controles mais rígidos, e em cadeias de agentes os erros se propagam e ficam difíceis de rastrear.

Modelo de fundação

Pré-treinado em grandes volumes de dados e capaz de servir a muitas tarefas, ao contrário do modelo tradicional treinado para uma única tarefa com dados rotulados. A maioria das empresas usa modelos prontos porque treinar um do zero exige volumes enormes de dados e muita computação. Adapta-se por prompts, RAG ou ajuste fino, e pode ser acessado como serviço gerenciado, como no Amazon Bedrock.

Tokens e janela de contexto

Token é uma parte do texto, como um trecho de palavra, lida pelo modelo; no preço por consumo, mais tokens de entrada e saída significam mais custo. Janela de contexto é a quantidade máxima de tokens que o modelo considera de uma vez; ultrapassada, o modelo desconsidera cláusulas do início de documentos longos ou instruções do começo de conversas longas. Mitiga-se enviando só os trechos relevantes (RAG), resumindo o histórico, repetindo as instruções essenciais, dividindo o texto ou escolhendo modelo com janela maior.

Engenharia de prompt

Princípios básicos: definir papel e público, dar contexto e as informações necessárias, ser específico no objetivo, indicar formato e tamanho de saída (com exemplo quando outro sistema vai ler), incluir poucos exemplos (few-shot), dividir tarefas complexas em etapas e pedir resolução passo a passo. Delimitar as fontes e mandar avisar quando faltar informação reduz, mas não elimina, respostas inventadas. É a primeira abordagem a tentar, por ter menor custo e esforço.

RAG (geração aumentada por recuperação)

A solução busca trechos relevantes nas fontes da empresa e os envia ao modelo junto com a pergunta, sem retreinar. Dá respostas com informações atualizadas e permite citar as fontes; reduz alucinações, mas não as elimina, e não dispensa controle de acesso aos documentos. Se a busca traz trechos pouco relacionados, a melhoria está na qualidade e organização das fontes e da recuperação; no Bedrock, Knowledge Bases é o RAG gerenciado, e embeddings permitem buscar por significado.

Ajuste fino (fine-tuning)

Adapta um modelo já treinado com dados de exemplo específicos, mudando comportamento, estilo, tom e formato. Indicado quando o modelo precisa seguir um estilo muito específico de forma consistente e há muitos exemplos de qualidade; sem exemplos não há ajuste fino. É ruim para informação que muda com frequência (preços semanais, políticas mensais), porque o modelo fica desatualizado entre um ajuste e outro.

Alucinação e natureza probabilística

Alucinação é uma resposta plausível, mas falsa ou sem base; não é invasão de servidor, falha que impede a resposta nem resposta correta inesperada. Modelos generativos escolhem palavras de forma probabilística, por isso o mesmo prompt gera respostas um pouco diferentes, sem defeito. Por isso não servem para calcular sozinhos valores contábeis exatos, e cada troca de versão do modelo deve ser testada com casos reais e comparada à atual.

Shadow AI e classificação de ferramentas

Shadow AI é o uso de ferramentas de IA não aprovadas, com risco de vazamento de dados confidenciais e de uso de conteúdo sem controle de propriedade intelectual. O guia pede classificação transparente: aprovadas, bloqueadas e em avaliação; em avaliação significa que a ferramenta ainda não está liberada para uso com dados da empresa. A lista é mantida pela governança de IA, com segurança, jurídico e negócio, em revisões periódicas, e funciona melhor com alternativas aprovadas úteis e uma política que diga quais ferramentas e quais dados podem ser usados.

Macetes para memorizar

Escada de adaptação

Prompt, depois RAG, depois ajuste fino; treinar do zero quase nunca. Sobe um degrau só quando o anterior não resolve.

RAG x ajuste fino

RAG muda O QUE o modelo sabe (conhecimento atual, com fonte); ajuste fino muda COMO ele responde (tom, estilo, formato).

Conceitos básicos

Algoritmo aprende, modelo é o aprendido, treinamento é a aula, inferência é a prova de todo dia (e o custo contínuo).

Capacidades de agentes

A-F-C-O: Autonomia, Ferramentas, Comunicação entre agentes, Orquestração. Mais autonomia, mais controle.

Regras ou IA

Estável, explícito, exato e auditável vai para regra; complexo, mutável, texto ou imagem vai para IA; misturado, use os dois.

Normas ISO

42001 é de gestão e certifica; 23053 descreve sistemas de ML e dá vocabulário.

Classificação de ferramentas

Semáforo: verde aprovada, vermelho bloqueada, amarelo em avaliação (ainda não usa com dados da empresa).

Janela de contexto

Recortar, resumir, repetir: só os trechos relevantes, histórico condensado e instruções essenciais de novo.

Tabela de memorização: não confunda

Quando é...	...e quando é
RAG Conhecimento que muda com frequência, informação recente (lei do mês passado), necessidade de citar as fontes, respostas com base em documentos internos.	**Ajuste fino** Mudar tom de voz, estilo ou formato fixo de saída de forma consistente, quando há muitos exemplos de qualidade e o conhecimento já está no prompt.
Automação por regras Lógica conhecida, estável, determinística e auditável; passos sempre iguais.	**Machine learning ou IA generativa** Padrões complexos ou que mudam com frequência, texto livre em grande volume, imagens e áudios, categorias que mudam.
Agente de IA Tarefas com caminhos variáveis que exigem planejar etapas, escolher e usar ferramentas e agir em sistemas com autonomia.	**Chatbot, assistente ou fluxo fixo** Responder perguntas uma de cada vez, sugerir um texto para uma pessoa revisar ou executar passos sempre iguais, como um relatório diário.
Amazon Bedrock Aplicações de IA generativa com modelos de fundação de vários fornecedores via serviço gerenciado, sem gerenciar infraestrutura, escolhendo por qualidade, custo e latência.	**Amazon SageMaker AI** Construir, treinar e implantar modelos de ML personalizados, com cientistas de dados e dados proprietários, quando nenhum modelo pronto atende.

Amazon Quick Usuários de negócio sem conhecimento técnico que querem assistentes prontos para pesquisar documentos, analisar dados, automatizar tarefas e perguntar em linguagem natural sobre dados e painéis.	**Amazon Bedrock** Times que constroem aplicações de IA generativa escolhendo entre vários modelos de fundação, com Guardrails e Knowledge Bases.
ISO/IEC 42001 Padrão de sistema de gestão de IA, com políticas, responsabilidades, riscos e melhoria contínua; a empresa pode se certificar para dar evidências a clientes.	**ISO/IEC 23053** Framework para descrever sistemas de IA que usam machine learning, com foco em conceitos e componentes; ajuda a ter vocabulário comum com times técnicos, fornecedores e auditores.
Drift O modelo funcionava bem e os erros aumentaram com o tempo porque os dados atuais mudaram em relação aos de treino.	**Sobreajuste (overfitting)** O modelo acerta quase tudo nos dados de treino, mas erra muito com dados novos desde o início: decorou os exemplos.

Números que caem

Item	Valor
Peso do domínio na prova	**24%**
Probabilidade de 80% de compra	**em média, 2 de cada 10 casos assim não compram**
Classificações de ferramentas de IA citadas no guia	**3: aprovada, bloqueada e em avaliação**
Capacidades centrais de agentes citadas no guia	**4: autonomia, uso de ferramentas, comunicação entre agentes e orquestração**

Alertas de prova (pegadinhas)

⚠ **Alerta de prova:** Resolver informação nova ou que muda toda semana com ajuste fino, treinamento do zero ou reiniciando o serviço.
Como não cair: O modelo nunca conheceu a informação, e ajustar periodicamente deixa-o desatualizado. Informação recente ou dinâmica pede RAG.

⚠ **Alerta de prova:** Usar RAG para fazer o modelo responder no tom da marca e no formato do laudo.
Como não cair: RAG traz conhecimento externo. Se o conhecimento já está no prompt e o problema é estilo e formato, a técnica é ajuste fino (ou exemplos no prompt).

⚠ **Alerta de prova:** Aumentar a temperatura ou a criatividade para respostas mais completas, mais organizadas ou com informação interna.
Como não cair: Temperatura controla a variabilidade. Não dá acesso a informação, não resolve perda de contexto nem formato, e pode aumentar respostas inventadas.

⚠ **Alerta de prova:** Atribuir aos Guardrails o sumiço das cláusulas do início de um contrato de centenas de páginas.
Como não cair: Guardrails filtram conteúdo por política, não pela posição no texto. Perder o começo de documento longo é janela de contexto excedida.

⚠ **Alerta de prova:** Explicar a queda de desempenho com o tempo como overfitting, velocidade de inferência ou janela de contexto.
Como não cair: Mudança no comportamento real depois de um período bom é drift. Overfitting aparece já na comparação treino x dados novos, e janela de contexto é limite de modelos de linguagem.

⚠ **Alerta de prova:** Achar que o modelo se retreina sozinho a cada conversa, expira depois de alguns meses ou que o provedor de nuvem cuida do desempenho do caso de uso.
Como não cair: O mundo muda, o modelo não se corrige sozinho. Precisa de responsável definido, monitoramento e atualização pela empresa.

⚠ **Alerta de prova:** Classificar planilhas como não estruturadas ou fotos e gravações como estruturadas por terem formato fixo, data e hora ou estarem no sistema da empresa.
Como não cair: O que define é seguir um esquema de campos. Tabelas são estruturadas; textos livres, imagens e áudios são não estruturados.

⚠ **Alerta de prova:** Usar um agente autônomo ou vários agentes para um cadastro com cinco passos fixos.
Como não cair: Autonomia e complexidade devem ser proporcionais à necessidade. Passos fixos sem exceções pedem um fluxo de automação.

⚠ **Alerta de prova:** Usar IA generativa ou um modelo de ML para calcular multa contratual, imposto ou valores contábeis exatos.
Como não cair: IA generativa pode gerar valores plausíveis, mas errados, e o ML aprenderia aproximações de uma regra já conhecida. Cálculo exato e auditável pede regras.

⚠ **Alerta de prova:** Acreditar que mais dados compensam rótulos inconsistentes ou que registros duplicados tornam o modelo mais preciso.
Como não cair: Rótulos conflitantes se contradizem e duplicados distorcem o aprendizado. Defina critérios únicos de rotulagem e corrija a qualidade na origem.

⚠ **Alerta de prova:** Escolher sempre o maior modelo disponível.
Como não cair: Para tarefas simples e de grande volume, um modelo menor tende a ter menor custo e latência com qualidade suficiente. Modelos maiores também alucinam.

⚠ **Alerta de prova:** Combater shadow AI proibindo toda IA, punindo quem usa ou fazendo o funcionário assumir o risco por escrito.
Como não cair: Proibir empurra o uso para a clandestinidade e o risco com dados de clientes é da empresa. Classifique as ferramentas e ofereça alternativas aprovadas.

⚠ **Alerta de prova:** Tratar ferramenta em avaliação como liberada para testes com dados reais ou para uso avisando o gestor.
Como não cair: Em avaliação significa ainda não liberada para uso com dados da empresa. Proibição definitiva é a classificação bloqueada.

⚠ **Alerta de prova:** Interpretar 80% de probabilidade como garantia ou deixar a equipe técnica escolher o ponto de corte pela maior precisão estatística.
Como não cair: Previsão é estimativa que pode errar. O equilíbrio entre os dois tipos de erro considera custos de negócio e é decisão do negócio.

Como cai na prova

📝 **Como cai:** Assistente de IA generativa com problema (não sabe algo recente, inventa regras, ignora o começo do texto, resposta genérica, custo alto) e pergunta a causa ou a técnica.
Como resolver: Classifique o problema: falta de conhecimento = RAG; estilo e formato = ajuste fino ou exemplos; instrução vaga = prompt com público, objetivo, formato e tamanho; texto grande demais = janela de contexto; custo = tokens e modelo. Descarte temperatura e troca de idioma.

📝 **Como cai:** Qual tipo de solução é mais adequada para um processo descrito.
Como resolver: Lógica fixa e auditável = regras. Prever número ou categoria com dados tabulares = ML preditivo. Criar ou resumir conteúdo aceitando variação revisada = IA generativa. Caminhos variáveis com uso de ferramentas = agente.

📝 **Como cai:** Modelo que piorou ou que vai mal em outro contexto.
Como resolver: Bom e depois piorou com mudança no mundo = drift; ótimo no treino e ruim com dado novo = sobreajuste; treinado com um público e usado em outro = falta de representatividade. Resposta: monitorar, avaliar com dados reais e retreinar, com responsável definido.

📝 **Como cai:** Qual serviço ou recurso da AWS atende, em nível estratégico.
Como resolver: Perfil de quem usa decide: negócio sem técnico = Amazon Quick; aplicação de IA generativa com vários modelos = Amazon Bedrock (RAG = Knowledge Bases); ML personalizado com cientistas de dados = SageMaker AI. Pricing Calculator, Cost Explorer e Marketplace são distratoras de custo e compra.

📝 **Como cai:** Escolha DUAS sobre características (RAG, ajuste fino, agentes, dados não estruturados, fatores de custo).
Como resolver: Avalie cada alternativa sozinha contra o critério do enunciado, marque as duas que o atendem e descarte absolutos como garantir respostas sempre corretas, dispensar governança ou funcionar sem dados.

📝 **Como cai:** Cenário de dado ou tarefa e pergunta pela categoria de IA.
Como resolver: Texto = processamento de linguagem natural; foto ou vídeo = visão computacional; áudio = transcrição; campos de PDFs = extração de documentos; histórico de uso ou de compras para sugerir itens = recomendação.

Gatilhos do enunciado

Se o enunciado trouxer...	Pense em
Políticas que mudam todo mês, informação recente, citar a fonte de cada resposta	**RAG; no Bedrock, Knowledge Bases**
Tom de voz da marca, formato padrão de laudo, muitos exemplos de qualidade	**Ajuste fino**
Três reclamações já classificadas incluídas no prompt	**Prompt com poucos exemplos (few-shot)**
Atue como analista de crédito sênior	**Definição de papel no prompt**
Desconsidera cláusulas do início ou esquece instruções dadas no começo da conversa	**Janela de contexto excedida; enviar só o relevante, resumir histórico, repetir instruções**
Custo subiu depois que usuários colaram documentos inteiros	**Custo por consumo cresce com tokens**
Erros aumentaram depois de mudança no comportamento real ou novo meio de pagamento	**Drift; retreinar com dados recentes**
Acerta quase tudo no treino e erra muito com dados novos	**Sobreajuste: decorou e não generaliza**
Tabela fixa, regra definida em lei ou contrato, precisa ser auditável	**Automação baseada em regras**
Consulta estoque e registra pedido no ERP / um agente pede análise a outro / agentes coordenados entre si	**Uso de ferramentas / comunicação entre agentes / orquestração**

Dividir clientes em grupos sem defini-los antes / prever valor em reais	**Agrupamento (não supervisionado) / regressão**
Buscar produtos por significado, sem as palavras exatas do catálogo	**Embeddings e busca semântica**
Gestor sem conhecimento técnico quer assistente para pesquisar documentos, analisar dados e perguntar sobre painéis	**Amazon Quick**
Funcionários usam ferramentas gratuitas da internet com dados de clientes	**Shadow AI; classificar ferramentas em aprovadas, bloqueadas ou em avaliação**

🎯 **Se você lembrar de só uma coisa deste capítulo...**

- RAG traz conhecimento atualizado com fonte; ajuste fino muda estilo e formato; prompt é o primeiro degrau.
- Modelo não se corrige sozinho: drift exige monitoramento, retreino e um responsável.
- Regra fixa, exata e auditável não precisa de IA; texto livre, imagem e padrão mutável precisam.
- Janela de contexto e tokens explicam esquecimento de partes do texto e aumento de custo.
- Em avaliação não é liberado: shadow AI se combate com classificação clara e alternativas aprovadas.

Essenciais para a prova (checklist)

☐ Sei definir algoritmo, modelo, treinamento, inferência e previsão e explicar por que a inferência gera custo contínuo.

☐ Sei explicar a relação entre IA, ML e IA generativa e dar exemplos de ML preditivo e de IA generativa.

☐ Sei classificar dados como estruturados ou não estruturados e associar NLP, visão computacional, transcrição e extração de documentos.

☐ Sei explicar por que qualidade, representatividade e rótulos consistentes determinam o resultado do modelo.

☐ Sei diferenciar drift de sobreajuste e dizer quem deve monitorar e atualizar o modelo.

☐ Sei decidir entre automação por regras, ML, IA generativa e agente.

☐ Sei as quatro capacidades centrais de agentes e por que agir em sistemas exige mais controle.

☐ Sei aplicar os princípios de prompt e reconhecer few-shot, definição de papel e divisão em etapas.

☐ Sei quando usar RAG e quando usar ajuste fino, e o que a janela de contexto e os tokens afetam.

☐ Sei o papel estratégico de Amazon Bedrock, SageMaker AI e Quick e a diferença entre ISO/IEC 42001 e 23053.

Questões de revisão do capítulo

R2.1 · Fundamentos e Letramento em IA

Uma empresa quer que o modelo responda no tom de voz da marca e em um formato padrão de laudo técnico. O conhecimento necessário já vem no próprio prompt. Qual técnica é mais indicada?

A. Aumento da janela de contexto do modelo usado hoje
B. Ajuste fino com exemplos no tom e no formato desejados
C. RAG com uma base de documentos sobre a história da marca
D. Classificação de ferramentas de IA como aprovadas

R2.2 · Fundamentos e Letramento em IA

Um assistente de IA generativa resume contratos, mas começa a desconsiderar cláusulas do início quando o documento tem centenas de páginas. Qual é a causa mais provável?

A. O documento excede a janela de contexto do modelo
B. Os Guardrails bloquearam as cláusulas mais antigas
C. A temperatura do modelo está configurada muito baixa
D. O modelo sofreu drift por falta de novos dados

R2.3 · Fundamentos e Letramento em IA

Em qual situação machine learning tende a ser mais adequado do que regras fixas?

A. Aplicar uma alíquota de imposto definida em lei para cada produto
B. Calcular o frete a partir de uma tabela fixa de faixas de peso
C. Detectar fraudes com padrões complexos que mudam com frequência
D. Bloquear pedidos acima de um valor-limite definido pela diretoria

R2.4 · Fundamentos e Letramento em IA

Um time quer melhorar as respostas de um modelo de fundação para um novo caso de uso. Qual abordagem costuma ser tentada primeiro, por ter menor custo e esforço?

A. Troca de provedor de nuvem
B. Treinamento do zero
C. Engenharia de prompt
D. Ajuste fino do modelo

Respostas comentadas

R2.1. B) Ajuste fino com exemplos no tom e no formato desejados

Por quê: Técnicas de adaptação de modelos incluem RAG, que traz conhecimento atualizado, e ajuste fino, que adapta o comportamento, o estilo e o formato. Muitas vezes, bons prompts com exemplos já resolvem, antes de partir para o ajuste fino.

✗ A) Incorreta. Uma janela de contexto maior não muda o tom nem o formato das respostas.
✗ C) Incorreta. RAG traz conhecimento externo. O problema aqui é estilo e formato, e o conhecimento já está no prompt.
✗ D) Incorreta. Classificar ferramentas é medida de governança contra shadow AI, não técnica de adaptação.

> **Dica de prova:** Conhecimento: RAG. Comportamento, tom e formato: ajuste fino ou prompts com exemplos.

R2.2. A) O documento excede a janela de contexto do modelo

Por quê: Modelos de IA generativa processam texto em tokens e têm uma janela de contexto máxima. Para documentos grandes, alternativas incluem dividir o texto em partes, resumir por seções, usar RAG ou escolher um modelo com janela maior.

✗ B) Incorreta. Guardrails filtram conteúdo indesejado por política, não por posição do trecho no documento.
✗ C) Incorreta. Temperatura baixa torna as respostas mais previsíveis, mas não faz o modelo perder partes do texto.
✗ D) Incorreta. Drift é a queda de desempenho ao longo do tempo por mudança nos dados, não a perda de trechos de um documento longo.

> **Dica de prova:** Resposta que perde partes de textos muito longos: suspeite da janela de contexto e do limite de tokens.

R2.3. C) Detectar fraudes com padrões complexos que mudam com frequência

Por quê: O guia do exame pede decidir quando usar automação baseada em regras e quando usar IA. Regras funcionam bem para lógica estável e explícita. IA é indicada quando os padrões são complexos, variáveis ou difíceis de descrever.

✗ A) Incorreta. Alíquotas definidas em lei são regras claras, melhor tratadas por automação.
✗ B) Incorreta. Uma tabela fixa de frete é resolvida com regras simples.
✗ D) Incorreta. Um limite de valor é uma regra objetiva e não exige aprendizado.

> **Dica de prova:** Padrão complexo e mutável: ML. Regra clara e estável: automação.

R2.4. C) Engenharia de prompt

Por quê: Em geral, as abordagens seguem uma escada de esforço: engenharia de prompt, depois RAG para trazer conhecimento e, se necessário, ajuste fino para mudar o comportamento.

✗ A) Incorreta. Trocar de provedor não melhora, por si só, as respostas para o caso de uso.
✗ B) Incorreta. Treinar do zero é a opção mais cara e demorada.
✗ D) Incorreta. O ajuste fino exige dados de exemplo, tempo e custo maiores.

> **Dica de prova:** Primeiro o prompt, depois RAG e, se necessário, ajuste fino.

03

CAPÍTULO 3 · 24% DA PROVA

Governança e Liderança em IA Responsável

Aqui preciso aplicar IA responsável a decisões de negócio: reconhecer a dimensão em jogo (equidade, explicabilidade, privacidade, segurança, transparência, robustez, controlabilidade), decidir trade-offs sem abrir mão de princípios, saber quando exigir supervisão humana e quais salvaguardas usar. Também caem estrutura de governança (comitê multifuncional, dono de negócio responsável, conselho, inventário, classificação de risco), compliance (LGPD, AI Act, setores regulados, residência de dados), modelo de responsabilidade compartilhada, controles de acesso, Amazon Bedrock Guardrails, viés ao longo do ciclo de vida, propriedade intelectual e riscos de confiabilidade como alucinação, degradação de dados, drift, envenenamento e injeção de prompt. A prova traz um cenário e pede a dimensão, o controle ou a conduta mais adequada. Onde mais erro: aceitar um aviso como solução, transferir a responsabilidade ao fornecedor e aplicar o mesmo controle a tudo em vez de controle proporcional ao risco.

Neste capítulo

18 tópicos-chave · 14 pegadinhas · 8 macetes · 14 gatilhos · 4 questões comentadas

Tópicos-chave

Dimensões da IA responsável

A AWS descreve equidade, explicabilidade, privacidade e segurança, segurança contra danos (safety), controlabilidade, veracidade e robustez, governança e transparência. Na prova, o cenário descreve um sintoma e pede a dimensão afetada. Custo e portabilidade não são dimensões da IA responsável.

Equidade

Resultados justos entre grupos de pessoas: crédito aprovado menos para uma região com perfis financeiros parecidos, reconhecimento de voz que erra mais com certos sotaques, imagens geradas que repetem o mesmo perfil em cargos de liderança. Variáveis que funcionam como substitutas de características sensíveis, como o CEP, geram viés mesmo sem intenção; a decisão coerente é avaliar o impacto e ajustar ou retirar a variável, mesmo perdendo precisão.

Explicabilidade

Entender e explicar como o sistema chega ao resultado: por que o crédito foi negado, por que um fornecedor recebeu nota alta, quais fatores de risco levam um cliente a cancelar (o que orienta a ação do gestor). Em decisões de alto impacto e contexto regulado, a explicabilidade costuma pesar mais que um ganho de precisão de um modelo mais complexo.

Transparência

Informar as partes interessadas sobre o uso da IA e suas limitações: avisar que o cliente fala com um assistente, indicar conteúdo gerado por IA conforme as regras aplicáveis, publicar os critérios e oferecer canal para contestar uma prioridade definida por modelo. Omitir o uso de IA, esconder que a decisão é tomada por um modelo ou transferir ao público a tarefa de descobrir não são transparência. Transparência não exige revelar tudo: com segredo comercial, explique os principais fatores e critérios, sem revelar detalhes proprietários.

Privacidade

Usar dados pessoais exige base legal, finalidade clara, minimização, mascaramento, controle de acesso e retenção limitada com prazo e finalidade. Dados coletados para uma finalidade (gravações de atendimento) não podem ser reaproveitados livremente para outra (treinar modelo), e as regras valem também em pilotos. Modelos podem memorizar dados de treino, por isso se removem dados pessoais antes do ajuste fino e se testam vazamentos; em ambientes de teste, usam-se dados mascarados ou sintéticos.

Segurança (safety), robustez e controlabilidade

Segurança trata de evitar resultados que causem dano: doses de medicamentos, resumos de prontuário que omitem alergias, procedimentos de manutenção de redes elétricas; quanto maior o potencial de dano, mais rígidos os controles, as fontes aprovadas e a revisão de especialistas. Robustez é funcionar de forma confiável em condições variadas (pouca luz, ângulos diferentes, câmeras novas exigem novos testes). Controlabilidade é conseguir monitorar, interromper e corrigir o comportamento, com peso maior em agentes que executam ações.

Governança desde a concepção

Práticas de IA responsável entram no planejamento, como parte do desenho da solução, quando mudanças são mais baratas: avaliar riscos e impactos na concepção, definir métricas de equidade antes do desenvolvimento e manter um cartão do modelo (model card) com finalidade, dados, limitações e testes. Para verificar se uma diretriz é seguida, usam-se checkpoints de revisão nas etapas do projeto com evidências registradas, não autoavaliação nem lembrete anual.

Supervisão humana e critérios de escalonamento

Revisão humana antes de cada decisão é exigida quando a decisão é difícil de reverter ou tem efeito significativo sobre direitos, saúde ou oportunidades de pessoas; sugestões opcionais e rascunhos revisados podem ser só monitorados. A supervisão precisa ser efetiva: o revisor pode alterar ou rejeitar, tem informação e tempo suficientes e vê a confiança do modelo; aprovar 99% em segundos indica viés de automação. Escale para uma pessoa quando a confiança é baixa, o valor ou impacto é alto, o cliente contesta ou está em situação vulnerável.

Estrutura de governança e responsabilidade

Comitê multifuncional com áreas de negócio, tecnologia, jurídico e compliance, com papéis definidos e integrado às estruturas existentes de segurança e de dados, sem processos paralelos. A responsabilidade pelos resultados é de um dono de negócio definido, apoiado pela governança, nunca do fornecedor, do cientista de dados ou do algoritmo. O conselho define o apetite a risco e acompanha os riscos mais relevantes; em alto risco, validação por equipe independente e auditoria externa reforçam a confiança sem transferir a responsabilidade.

Classificação de risco e inventário

O nível de risco pesa sobretudo o impacto potencial sobre pessoas e o grau de autonomia da IA; ações automáticas sem revisão humana e decisões sobre direitos ou oportunidades elevam o risco. Os controles e a frequência de revisão são proporcionais ao risco (triagem de pronto-socorro x chatbot de receitas), e mudança de finalidade dispara reclassificação imediata. O inventário de sistemas de IA registra finalidade, nível de risco e responsável de negócio; exceções a ferramentas bloqueadas passam por processo formal.

LGPD e decisões automatizadas

Para dados pessoais de clientes brasileiros, a referência é a LGPD. Biometria, como reconhecimento facial, é dado pessoal sensível, com cuidados reforçados. O titular pode pedir revisão de decisões tomadas unicamente com base em tratamento automatizado que afetem seus interesses, então o processo deve prever canal de revisão e informações sobre os critérios; pedidos de exclusão valem também para dados usados em modelos, excluindo-os das bases e dos treinos futuros.

AI Act e compliance multinacional

A Lei de IA da União Europeia organiza obrigações por nível de risco: seleção de candidatos e avaliação de crédito de pessoas físicas tendem a ser alto risco, com gestão de riscos, supervisão humana e requisitos adicionais, sem serem proibidos. Em vários países, mapeie os requisitos de cada jurisdição e atribua responsáveis, sem seguir só a lei da matriz. Requisitos setoriais, como guardar registros das interações por cinco anos, entram no desenho desde o início.

Modelo de responsabilidade compartilhada

A AWS cuida da segurança da nuvem: infraestrutura global, hardware dos serviços gerenciados e segurança física dos data centers. O cliente cuida do que coloca e faz na nuvem: dados, acessos, configurações, região onde os dados ficam (residência de dados) e uso responsável das saídas, inclusive avaliar viés nos dados de treinamento. O Amazon Bedrock não usa prompts e respostas para treinar os modelos de base nem os repassa aos fornecedores dos modelos.

Controles de acesso e segurança de dados

Em assistentes com RAG, a recuperação deve respeitar as permissões que cada usuário já tem na origem; não dá para treinar o modelo para esquecer documentos. Agentes recebem só as permissões necessárias (só leitura para quem gera relatórios), integração com gestão de identidades revoga acessos de quem sai, e acessos são registrados e revisados periodicamente, sem senhas compartilhadas. Registrar prompts e respostas, respeitando a privacidade, serve para auditar, apurar incidentes e melhorar a solução.

Amazon Bedrock Guardrails

Aplica políticas de segurança às interações com o modelo: filtros de conteúdo nocivo nas entradas e saídas, temas negados (por exemplo, recomendação personalizada de investimento, encaminhada a especialista), filtros de informações sensíveis que mascaram dados pessoais, verificações de fundamentação e filtros de ataques de prompt. Os controles devem ser proporcionais ao público, mais rígidos para crianças, e podem ser aplicados a diferentes modelos. Não treina modelos, não estima custos e não fornece RAG (isso é Knowledge Bases).

Riscos de confiabilidade

O guia cita alucinações, degradação da qualidade dos dados e drift do modelo. Alucinação se reduz com respostas baseadas em fontes verificadas, verificação de fundamentação, citação de fontes, revisão humana por amostragem e checagem de citações na fonte oficial. Degradação aparece quando um sistema de origem muda o formato ou a base de conhecimento fica desatualizada (política de trocas antiga); mudanças de versão do modelo são planejadas e testadas com antecedência, com plano de reversão testado e monitoramento com alertas e responsáveis.

Viés ao longo do ciclo de vida

O viés pode surgir na coleta e seleção dos dados (contratações de dez anos com um só perfil), no treinamento e depois da implantação, com mudanças nos dados (bias drift). Antes do lançamento, teste o desempenho separadamente por grupo e revise a representatividade dos dados; depois, compare resultados entre grupos periodicamente e monitore o perfil dos dados que chegam. A precisão média geral pode esconder erros concentrados em um grupo, e uma validação inicial não garante ausência de viés no futuro.

Propriedade intelectual e conteúdo sintético

Antes de treinar ou ajustar com textos comprados, confirme se a licença permite usá-los para treinar modelos. Na contratação, verifique uso de dados e titularidade do conteúdo gerado; em código gerado, verifique trechos com licenças incompatíveis e mantenha revisão por pares; oriente a não inserir conteúdo protegido e revise o que for publicado. Recriar voz ou imagem de pessoa real sem autorização viola direitos de imagem e voz, e voz sintética usada em fraude se combate confirmando pedidos sensíveis por outro canal e com aprovações formais.

Macetes para memorizar

💡 Sintoma para dimensão

Grupo tratado diferente = Equidade; 'por quê?' = Explicabilidade; 'é IA?' = Transparência; dano a pessoa = Segurança; condição diferente = Robustez; parar e corrigir = Controlabilidade; dado pessoal = Privacidade.

💡 Supervisão humana

Alto impacto + difícil de reverter + direitos de pessoas = humano antes de agir. Baixo impacto e reversível = monitorar.

💡 Controles proporcionais

Receita culinária não é triagem de pronto-socorro: controle e revisão crescem com o risco.

💡 Responsabilidade compartilhada

AWS cuida DA nuvem; cliente cuida do que põe NA nuvem (dados, acessos, região, uso).

💡 Guardrails x Knowledge Bases

Guardrails filtra e bloqueia; Knowledge Bases informa com documentos.

💡 Momento da IA responsável

Governança nasce no desenho, não na véspera do lançamento nem depois da reclamação.

💡 Agentes que agem

Menor privilégio, limite de valor, registro de tudo e escalonamento acima do limite.

💡 Trade-off negócio x princípio

Nem abrir mão do princípio, nem cancelar: achar outro caminho (consentimento, escopo menor, explicar fatores principais).

Tabela de memorização: não confunda

Quando é...	...e quando é
Equidade Há diferença injustificada de resultado entre grupos de pessoas (região, sotaque, perfil).	**Explicabilidade** As pessoas não entendem por que o resultado saiu assim e precisam conhecer os fatores; ajuda a investigar um problema de equidade, mas não é a dimensão violada nele.
Transparência Informar que há IA envolvida, que o conteúdo foi gerado por IA e quais são as limitações.	**Explicabilidade** Mostrar os fatores e critérios que levaram a uma decisão específica, como um crédito negado.
Segurança (safety) O risco é dano a pessoas: doses, alergias omitidas, procedimentos de segurança.	**Robustez** O sistema falha em condições variadas ou inesperadas: pouca luz, ângulos diferentes, câmeras novas.
Controlabilidade Garantir que a empresa consiga monitorar, interromper e corrigir o comportamento do sistema, sobretudo agentes.	**Transparência** Comunicar às partes interessadas o uso da IA e seus limites.

Amazon Bedrock Guardrails Bloquear temas proibidos, filtrar conteúdo nocivo, mascarar dados sensíveis, verificar fundamentação e barrar ataques de prompt.	**Amazon Bedrock Knowledge Bases** Conectar o modelo aos documentos da empresa para responder com RAG gerenciado.
Drift Mudança natural no perfil dos dados ao longo do tempo que derruba o desempenho.	**Envenenamento dos dados de treinamento** Manipulação intencional, como inserir registros falsos para favorecer uma empresa; combatida com controle de acesso, rastreabilidade e validação dos dados.
Responsabilidade da AWS Segurança da nuvem: infraestrutura de rede global, hardware dos serviços gerenciados, segurança física dos data centers.	**Responsabilidade do cliente** Segurança na nuvem: dados, acessos, configurações, região dos dados, viés nos dados de treino e uso responsável das respostas.

Números que caem

Item	Valor
Peso do domínio na prova	**24%**
Aprovação de revisores que sinaliza viés de automação no exemplo clássico	**99% das sugestões aprovadas em poucos segundos**

Alertas de prova (pegadinhas)

⚠ **Alerta de prova:** O modelo foi validado sem viés no lançamento, então nenhuma ação é necessária.
Como não cair: O viés pode surgir com o tempo (bias drift). A governança deve garantir monitoramento contínuo de viés.

⚠ **Alerta de prova:** Acompanhar só a precisão média geral do modelo para detectar viés.
Como não cair: A média pode esconder erros concentrados em um grupo. Teste e compare o desempenho por grupo.

⚠ **Alerta de prova:** Retirar o nome do bairro e manter o CEP, ou remover o nome e manter o histórico completo de localização.
Como não cair: O CEP continua indicando o bairro e a localização completa identifica a pessoa. O efeito permanece.

⚠ **Alerta de prova:** Resolver com um aviso: recomendar investimento com aviso no fim, avisar que o modelo pode revelar dados, informar que o modelo usa o endereço.
Como não cair: Aviso não torna aceitável o que exige profissional habilitado, vazamento de dados ou tratamento injusto. É preciso tema negado, remoção de dados ou ajuste da variável.

⚠ **Alerta de prova:** Transferir a responsabilidade ao fornecedor do modelo, à AWS, aos auditores ou ao algoritmo.
Como não cair: A empresa responde pela solução que oferece e pelo uso que faz da IA. Sempre há um dono de negócio responsável.

⚠ **Alerta de prova:** Aplicar o mesmo controle rigoroso, ou a mesma frequência de revisão, a todos os casos de uso.
Como não cair: Controles proporcionais ao risco: mais rigor onde há impacto sobre pessoas e autonomia, menos onde o impacto é baixo.

⚠ **Alerta de prova:** Manter a classificação de risco porque o modelo e os dados continuam os mesmos, mesmo com nova finalidade.
Como não cair: O mesmo modelo pode ter risco diferente conforme a decisão que apoia. Mudança de finalidade dispara reclassificação imediata.

⚠ **Alerta de prova:** Achar que o AI Act proíbe IA em recrutamento ou crédito, ou que basta um aviso ao usuário.
Como não cair: São usos de alto risco, permitidos com requisitos adicionais, como gestão de riscos e supervisão humana.

⚠ **Alerta de prova:** Recusar o pedido de exclusão porque os dados já foram usados no treinamento, ou esperar o próximo retreinamento anual.
Como não cair: O direito do titular continua valendo e tem prazo. Exclua das bases e dos treinos futuros, com apoio do jurídico.

⚠ **Alerta de prova:** Treinar o modelo para esquecer documentos sensíveis ou pedir que os funcionários evitem temas sigilosos.
Como não cair: Não há como garantir que o modelo esqueça, e comportamento não é controle. O controle está no acesso às fontes, respeitando as permissões de cada usuário.

⚠ **Alerta de prova:** Reduzir alucinações trocando para um modelo maior, aumentando a temperatura ou pedindo citações mais detalhadas.
Como não cair: Modelos maiores também alucinam, temperatura alta aumenta a variação e detalhe deixa a citação falsa mais convincente. Use fontes verificadas, verificação de fundamentação e revisão humana.

⚠ **Alerta de prova:** Usar dados sem base legal só no piloto e pedir autorização quando a solução estiver em produção.
Como não cair: As regras de privacidade valem também em pilotos. Busque o objetivo por outro caminho, como pedir consentimento ou usar dados menos sensíveis.

⚠ **Alerta de prova:** Para cumprir o prazo, lançar sem testes de conteúdo nocivo, ou então cancelar a iniciativa.
Como não cair: Há meio-termo: lançar com escopo reduzido e controles mínimos já testados.

⚠ **Alerta de prova:** Acreditar que o Amazon Bedrock usa os prompts para treinar os modelos, a menos que haja pedido formal.
Como não cair: O Bedrock não usa prompts e respostas para treinar os modelos de base nem os compartilha com os provedores; ainda assim, o cliente decide quais dados envia e controla os acessos.

Como cai na prova

📝 **Como cai:** Cenário com um problema de um sistema de IA e pergunta qual dimensão da IA responsável está em risco.
Como resolver: Identifique o sintoma central e mapeie: diferença entre grupos = equidade; falta de entender o porquê = explicabilidade; não saber que é IA = transparência; dano físico ou à saúde = segurança; falha em condições variadas = robustez. Descarte dimensões que só ajudam a investigar.

📝 **Como cai:** Objetivo de negócio (prazo, precisão, vendas) em conflito com um princípio.
Como resolver: Escolha o caminho que mantém o princípio e ainda busca o objetivo: proteger dados mesmo atrasando, priorizar explicabilidade em crédito, escopo reduzido com controles testados. Elimine 'aprovar e revisar depois', 'deixar para o fornecedor' e cancelar sem alternativa.

📝 **Como cai:** Novo uso de dados ou expansão regulatória (LGPD, AI Act, vários países, setor regulado) e pergunta pelo primeiro cuidado.
Como resolver: O primeiro passo é legal: origem e base legal dos dados, finalidade informada, mapeamento das leis aplicáveis, região de processamento. Aspectos técnicos como formato, volume e qualidade de áudio vêm depois.

📝 **Como cai:** Agente ou assistente que age ou acessa dados sensíveis e pergunta pelo controle mais adequado.
Como resolver: Procure menor privilégio, permissões herdadas da origem, limites de valor e de ações, registro para auditoria e escalonamento a humanos. Descarte administrador, acesso amplo, autorrevisão do agente e revisão anual.

📝 **Como cai:** Nome do risco a partir de um cenário (dados adulterados, instruções escondidas, formato de entrada mudou, base desatualizada, resposta inventada).
Como resolver: Intencional na base de treino = envenenamento; instrução maliciosa no texto ou documento = injeção de prompt; entrada mudou = degradação da qualidade dos dados; fonte velha no RAG = base desatualizada; plausível e falso = alucinação.

📝 **Como cai:** Quem deve responder, compor ou revisar (comitê, dono, conselho, validação, lista de ferramentas).
Como resolver: Responsável é sempre alguém da empresa: dono de negócio apoiado pela governança, comitê multifuncional com papéis definidos, conselho no apetite a risco, validação independente de quem desenvolveu.

Gatilhos do enunciado

Se o enunciado trouxer...	Pense em
Explicar ao cliente por que o crédito foi negado	**Explicabilidade**
Modelo aprova menos uma região com perfis financeiros parecidos	**Equidade**
Clientes conversam sem saber que não é uma pessoa	**Transparência**
Monitorar, interromper e corrigir um agente sempre que necessário	**Controlabilidade**
Doses de medicamento, alergias omitidas, procedimentos em redes elétricas	**Segurança (safety)**
Falha com pouca luz, ângulos diferentes ou câmeras novas	**Robustez**
Revisores aprovam 99% das sugestões em poucos segundos	**Viés de automação; supervisão apenas formal**
Temas proibidos, conteúdo ofensivo, público infantil	**Amazon Bedrock Guardrails: filtros de conteúdo e temas negados**
Instruções escondidas nas mensagens ou em documento recuperado pelo RAG	**Injeção de prompt (direta ou indireta); guardrails com filtros de ataques de prompt e testes**
Registros falsos inseridos na base de treino para favorecer alguém	**Envenenamento dos dados de treinamento**
Dados pessoais devem permanecer no país	**Região da AWS onde os dados são processados e armazenados**

Boas práticas da AWS para projetar e operar cargas de IA de forma responsável	**AWS Well-Architected Framework, Lente de IA Responsável**
Agente com permissão de apagar registros, mas só gera relatórios	**Menor privilégio: retirar permissões e deixar só leitura**
Voz sintética do diretor financeiro pedindo transferência urgente	**Confirmar por outro canal e seguir aprovações formais**

🎯 **Se você lembrar de só uma coisa deste capítulo...**

- A responsabilidade nunca sai da empresa: fornecedor, AWS, auditor e algoritmo não respondem pelo seu uso da IA.
- Controles e revisões proporcionais ao risco; mudança de finalidade exige reclassificar.
- Viés e qualidade se monitoram continuamente; validação no lançamento não basta.
- Aviso não substitui controle: tema negado, mascaramento, menor privilégio e revisão humana resolvem.
- LGPD para dados pessoais no Brasil (biometria é sensível); AI Act trata crédito e recrutamento como alto risco, não como proibidos.

Essenciais para a prova (checklist)

☐ Sei reconhecer cada dimensão da IA responsável a partir de um sintoma.

☐ Sei resolver trade-offs entre objetivo de negócio e princípios sem abrir mão do princípio nem cancelar sem alternativa.

☐ Sei quando exigir revisão humana antes da decisão e o que torna a supervisão efetiva.

☐ Sei compor um comitê de governança e dizer quem é responsável pelos resultados de uma aplicação de IA.

☐ Sei aplicar classificação de risco, controles proporcionais e inventário de sistemas de IA.

☐ Sei os pontos de LGPD e AI Act que caem: dado sensível, revisão de decisão automatizada, exclusão, alto risco.

☐ Sei dividir responsabilidades entre AWS e cliente e o que o Bedrock faz com prompts.

☐ Sei os recursos do Amazon Bedrock Guardrails e quando usar Guardrails ou Knowledge Bases.

☐ Sei onde o viés surge no ciclo de vida e como detectá-lo antes e depois do lançamento.

☐ Sei nomear e mitigar alucinação, degradação de dados, drift, envenenamento, injeção de prompt e riscos de propriedade intelectual.

Questões de revisão do capítulo

R3.1 · Governança e Liderança em IA Responsável

Um caso de uso foi classificado como de baixo risco quando servia apenas para organizar avaliações internas. Agora ele passou a indicar quais funcionários serão promovidos. O que a governança deve fazer?

A. Aguardar a próxima auditoria anual para rever a classificação de risco
B. Encerrar o caso de uso, pois decisões sobre pessoas não podem usar IA
C. Reclassificar o risco e aplicar os controles exigidos pelo novo nível
D. Manter a classificação, pois o modelo e os dados continuam os mesmos

R3.2 · Governança e Liderança em IA Responsável

Pela LGPD, o titular pode pedir a revisão de decisões tomadas unicamente com base em tratamento automatizado que afetem seus interesses. O que um processo de crédito automatizado deve prever?

A. Uma cláusula em que o cliente renuncia ao direito de pedir revisão
B. Um prazo de cinco anos para responder aos pedidos de revisão
C. Uma resposta padrão dizendo que a decisão do modelo é definitiva
D. Um canal para pedir revisão e informações sobre os critérios usados

R3.3 · Governança e Liderança em IA Responsável

Ao classificar o nível de risco de um caso de uso de IA, qual critério deve ter mais peso?

A. O tempo que a equipe levou para desenvolver o primeiro protótipo
B. O impacto potencial sobre pessoas e o grau de autonomia da IA
C. A quantidade de linhas de código usadas para construir a solução
D. O número de fornecedores que oferecem soluções parecidas no mercado

R3.4 · Governança e Liderança em IA Responsável

Um assistente de IA generativa acessa documentos internos. Como evitar que um funcionário receba respostas baseadas em documentos que ele não pode ver?

A. Pedir aos funcionários que evitem fazer perguntas sobre temas sigilosos da empresa
B. Remover os documentos sigilosos apenas das respostas mais longas
C. Treinar o modelo para esquecer os documentos mais sensíveis
D. Aplicar controles de acesso que respeitem as permissões de cada usuário

Respostas comentadas

R3.1. C) Reclassificar o risco e aplicar os controles exigidos pelo novo nível

Por quê: O guia do exame pede aplicar frameworks de classificação de risco ao longo de todo o ciclo de vida da IA. Uma mudança de finalidade pode mudar o nível de risco e os controles necessários.

✗ A) Incorreta. A mudança de finalidade deve disparar a reavaliação imediatamente, não na próxima auditoria.
✗ B) Incorreta. O uso é possível com controles adequados, como supervisão humana e análise de viés.
✗ D) Incorreta. O mesmo modelo pode ter risco diferente conforme a decisão que apoia.

> **Dica de prova:** Mudou a finalidade, reavalie o risco.

R3.2. D) Um canal para pedir revisão e informações sobre os critérios usados

Por quê: O guia do exame pede identificar e tratar riscos de compliance em processos que usam IA. No Brasil, a LGPD traz direitos específicos sobre decisões automatizadas.

✗ A) Incorreta. Esse direito não pode ser afastado por cláusula contratual.
✗ B) Incorreta. Um prazo tão longo esvazia o direito do titular.
✗ C) Incorreta. Tratar a decisão como definitiva contraria o direito de revisão.

> **Dica de prova:** Decisão automatizada no Brasil: direito de revisão e de informação (LGPD).

R3.3. B) O impacto potencial sobre pessoas e o grau de autonomia da IA

Por quê: O guia do exame pede aplicar frameworks de classificação de risco para priorizar decisões de governança e compliance. O impacto sobre pessoas e a autonomia da IA são critérios centrais.

✗ A) Incorreta. O tempo de desenvolvimento não mede risco.
✗ C) Incorreta. O tamanho do código não indica o risco do uso.
✗ D) Incorreta. A oferta de mercado não muda o impacto do caso de uso.

> **Dica de prova:** Risco de IA: impacto sobre pessoas e autonomia.

R3.4. D) Aplicar controles de acesso que respeitem as permissões de cada usuário

Por quê: O guia do exame pede identificar controles de acesso e medidas de segurança de dados adequadas para sistemas de IA. Em soluções com RAG, a recuperação deve respeitar as permissões dos usuários.

✗ A) Incorreta. Depender do comportamento dos funcionários não é um controle confiável.
✗ B) Incorreta. O tamanho da resposta não tem relação com o controle de acesso.
✗ C) Incorreta. Não é possível garantir que um modelo esqueça conteúdo. O controle deve estar no acesso às fontes.

> **Dica de prova:** Assistente com dados internos: as respostas herdam as permissões do usuário.

04

CAPÍTULO 4 · 24% DA PROVA

Prontidão, Liderança e Transformação com IA

Aqui eu preciso diagnosticar onde a empresa está e conduzir a mudança: dimensões de prontidão, lacunas em pessoas, processos, tecnologia e governança, modelo de maturidade, prontidão de dados, AWS CAF, patrocínio executivo, campeões de IA, comunicação, barreiras culturais, letramento, CoE e a passagem de piloto para produção. Quase tudo cai como cenário curto em que o enunciado diz o que está bom e pede o que falta, ou descreve uma reação humana (medo, resistência, desconfiança) e pede a intervenção da liderança. Onde mais erro: escolher a alternativa radical (punir, proibir, centralizar tudo, esperar a plataforma perfeita, desligar o processo antigo no primeiro dia) em vez da gradual e mensurável. Também confundo qualidade com acessibilidade dos dados e as fases envisionar, experimentar, lançar e escalar. As seis perspectivas do AWS CAF caem sempre como 'qual perspectiva trata de X'.

Neste capítulo

18 tópicos-chave · 14 pegadinhas · 8 macetes · 14 gatilhos · 4 questões comentadas

Tópicos-chave

Dimensões de prontidão para IA

O guia do exame cita cinco: alinhamento da liderança, qualidade dos dados, preparação cultural, infraestrutura técnica e frameworks de governança. A avaliação deve envolver líderes de negócio, tecnologia, dados, jurídico e pessoas, e não só TI, financeiro ou um consultor sem ouvir as áreas. Idade dos funcionários, quantidade de servidores próprios e número de fornecedores de nuvem não são dimensões de prontidão.

Lacunas de capacidade (pessoas, processos, tecnologia, governança)

O cenário costuma dizer o que já está bom, e a lacuna é a dimensão que sobra. Bons dados e plataforma, mas poucos sabem aplicar IA: pessoas. Profissionais e tecnologia bons, mas cada projeto com fluxo próprio de aprovação, teste e implantação: processos. Sem política, sem avaliação de riscos ou sem quem aprove: governança. Pessoas e processos prontos, mas sem ambiente seguro para testar: tecnologia. Depois do diagnóstico, priorize investimentos nas lacunas mais críticas, e não em todas com o mesmo peso.

Modelo de maturidade em IA

Serve para saber onde a empresa está, da experimentação à implantação em escala corporativa, e definir o próximo passo realista. Pilotos isolados por área, sem padrões nem governança comum, indicam experimentação. Produção em várias áreas com governança comum e métricas de valor indica escala corporativa; maturidade alta é IA integrada à estratégia e aos processos, com medição. Em estágio inicial, o investimento prioritário são fundamentos: dados, governança básica e capacitação.

Trilhas de progressão

O guia pede priorizar investimentos e criar trilhas com base nos objetivos estratégicos e no nível atual de maturidade. Se o marketing já usa IA em produção e o jurídico nem começou, a resposta é trilhas por área conforme o nível, com padrões comuns. Muitos pilotos soltos pedem governança comum e levar os de maior valor para produção, não iniciar mais pilotos.

Prontidão de dados: qualidade, acessibilidade e silos

Qualidade é o dado estar correto e completo (30% dos registros sem código do cliente). Acessibilidade é o dado existir, mas não poder ser usado a tempo (pedido formal que leva semanas). Silos são sistemas separados por área, com definições diferentes de cliente, que geram dados fragmentados e inconsistentes e pioram os modelos. Documentos digitalizados sem padrão ainda exigem extração e organização: digitalizado não é pronto.

Estratégia, propriedade e compartilhamento de dados

A base da IA são donos dos dados definidos, regras de qualidade e acordos de compartilhamento. O curador de dados (data steward) cuida da qualidade e das regras de uso de um conjunto de dados no dia a dia. Acordos entre áreas definem formato, qualidade e mudanças; acordos com terceiros definem finalidade, proteção e regras de uso. A estratégia de dados torna os dados confiáveis e reutilizáveis entre projetos, em vez de cada projeto criar a sua base.

Instrumentos de gestão de dados

Catálogo de dados: mostra quais dados existem, onde estão e quem responde por eles. Linhagem de dados: registra origem e transformações, respondendo a auditores. Cadastro mestre com padrões e responsáveis: resolve códigos e nomes diferentes para o mesmo item antes de um projeto como previsão de demanda. Gestão central de consentimentos e finalidades: resolve quando ninguém sabe o que cada cliente autorizou. Indicadores de qualidade com metas e donos: detectam dados com erro a tempo.

Requisitos de infraestrutura para IA

Avaliados em nível de decisão de negócio. Com dados sensíveis, o essencial é acesso seguro e controlado às fontes. Para não comprar GPUs, serviços gerenciados em nuvem pagos conforme o uso; para picos como a Black Friday, capacidade de escalar com a demanda e custos sob controle. Para milhões de clientes, avaliar capacidade de escala, disponibilidade, segurança e custos. Infraestrutura duplicada por área pede plataforma comum com controles padrão, e testes rápidos com segurança pedem ambiente isolado com dados protegidos.

AWS Cloud Adoption Framework (AWS CAF)

Ajuda a avaliar a prontidão, identificar e priorizar capacidades e planejar a evolução, com orientação específica para IA. Organiza as capacidades em seis perspectivas: negócio, pessoas, governança, plataforma, segurança e operações. Não define algoritmos nem arquitetura de modelos, não calcula preços (isso é a AWS Pricing Calculator) e não emite certificações.

Patrocínio executivo e campeões de IA

O patrocinador executivo garante recursos, alinha prioridades e remove barreiras entre as áreas; se ele sai, defina logo um novo e realinhe a liderança. Campeões de IA são multiplicadores de adoção nas áreas: profissionais respeitados, interessados em IA e com tempo para o papel, que incentivam o uso e compartilham boas práticas. Campeões não aprovam projetos sozinhos, não substituem a governança e não assumem decisões executivas.

Times multifuncionais com responsabilidade clara

Iniciativas de IA envolvem negócio, tecnologia, jurídico e compliance, com padrões claros de quem decide o quê (por exemplo, uma matriz de responsabilidades). Sem isso, as decisões travam, e mais reuniões não resolvem. Jurídico e compliance entram desde o início para identificar requisitos e riscos cedo e evitar retrabalho, sem programar e sem assumir a responsabilidade, que continua com o dono de negócio.

Comunicação transparente sobre IA

Reduz a resistência explicando prazos, resultados esperados, efeitos nos papéis e planos de capacitação. Silêncio aumenta boatos, prometer que nenhum cargo será afetado destrói a confiança e deixar cada gestor decidir o que contar gera mensagens desencontradas. Expectativas irreais anunciadas por um executivo devem ser ajustadas com metas realistas, marcos e resultados medidos, nunca com métricas manipuladas.

Barreiras culturais e intervenções da liderança

O guia cita aversão a risco, resistência à mudança e medo de errar. As intervenções que funcionam: tratar experimentos como aprendizado e reconhecer quem testa, compartilhar os aprendizados de experimentos que falharam, envolver gerentes resistentes na escolha dos casos e no desenho do novo papel, comparar com o erro atual do processo e começar com uso assistido, e dar o exemplo usando as ferramentas. Punir, obrigar sem envolver ou esperar a resistência passar sozinha não funcionam.

Letramento em IA na empresa

O guia cita programas de prova de conceito, hackathons, treinamentos e treinamento em IA responsável. As trilhas variam por papel (executivos, gestores, times técnicos): executivos precisam de sessões curtas e práticas com casos do setor; usuários de um assistente, de uso responsável, limites da ferramenta e cuidados com dados, não de programação. Como as ferramentas mudam rápido, o letramento é um programa contínuo, com comunidade de prática e horas reservadas na rotina.

Transição de papéis humanos

A IA assume o repetitivo, e as pessoas passam para supervisão, tratamento de exceções e decisões que exigem julgamento, equilibrando pensamento crítico, empatia e criatividade com as capacidades da IA. No marketing, a IA gera variações e as pessoas definem a estratégia e aprovam as peças. A requalificação começa pelas funções mais afetadas, e supervisores de um assistente desenvolvem análise crítica das respostas, tratamento de exceções e empatia.

Abordagem iterativa: envisionar, experimentar, lançar, escalar

É a sequência citada como exemplo no guia do exame. Envisionar define visão, objetivos e casos de uso prioritários; experimentar testa hipóteses com protótipos de escopo limitado e mede; lançar coloca em produção com operação e monitoramento; escalar expande o que foi comprovado para mais áreas, com padrões comuns, suporte e medição de valor. Pular a experimentação aumenta o risco do investimento.

Centro de Excelência (CoE) em IA

Define padrões, compartilha práticas e apoia as áreas a escalar a IA; não executa sozinho todos os projetos, não aprova orçamento de tecnologia e não substitui o comitê de governança. Um modelo mais centralizado faz sentido no início, com pouca experiência e sem padrões; com áreas maduras, evolui para um modelo federado, com padrões centrais e execução nas áreas. Seu sucesso se mede pelo valor gerado e pela velocidade com que as áreas entregam IA.

Do piloto à produção e à escala

A produção exige governança, monitoramento, suporte e requisitos operacionais: documentar, treinar a operação, definir níveis de serviço e trocar cópias manuais por fontes automatizadas com donos. Uma POC vai para produção com valor comprovado contra a linha de base, riscos aceitáveis e dono definido. Na escala, expanda por ondas com plano de contingência, desligue o processo antigo aos poucos, com metas atingidas e plano de retorno, e acompanhe adoção, qualidade, valor de negócio e custo.

Macetes para memorizar

Dimensões de prontidão

L-D-C-I-G: Liderança alinhada, Dados de qualidade, Cultura preparada, Infraestrutura técnica, Governança. O cenário elimina as que estão boas; marque a que sobrou.

Lacunas de capacidade

P-P-T-G: Pessoas, Processos, Tecnologia, Governança. 'Ninguém sabe usar' = Pessoas; 'cada um faz de um jeito' = Processos; 'falta ambiente' = Tecnologia; 'ninguém aprova nem responde' = Governança.

Fases iterativas

Ver, Testar, Ligar, Espalhar: Envisionar (visão e objetivos), Experimentar (protótipo e medição), Lançar (produção com operação), Escalar (expandir com padrões).

Seis perspectivas do AWS CAF

Três de gestão e gente: Negócio, Pessoas, Governança. Três técnicas: Plataforma (base técnica), Segurança (proteção), Operações (dia a dia).

CoE

Nasce central, cresce federado. CoE habilita e padroniza; quem executa são as áreas.

Piloto para produção

Documenta, treina, SLA, monitora, dono. E desliga o antigo só aos poucos, com plano de retorno.

Comunicação com a força de trabalho

Quando, o que se espera e o que muda para mim, mais o plano de capacitação. Nada de silêncio, nada de promessa vazia.

Barreiras culturais

Medo, aversão e resistência se tratam com envolver, mostrar ganho real e tornar seguro errar; nunca com punir, obrigar ou esperar.

Tabela de memorização: não confunda

Quando é...	...e quando é
CoE centralizado Início da jornada, com pouca experiência e sem padrões: o CoE concentra o conhecimento e cria os padrões.	**CoE federado** Áreas maduras ou CoE virando gargalo: padrões centrais e execução nas áreas. Centralizar mais ou extinguir o CoE são as alternativas erradas.
Fase de envisionar Definir visão de longo prazo, objetivos de negócio e casos de uso prioritários.	**Fase de experimentar** Testar hipóteses com protótipos de escopo limitado e medir os resultados para validar a viabilidade.
Fase de lançar Colocar em produção o que foi validado, com operação, monitoramento e responsáveis definidos.	**Fase de escalar** Expandir o que já foi comprovado para mais áreas e unidades, com padrões comuns, suporte e medição de valor.
Qualidade dos dados Os dados estão disponíveis, mas incompletos, errados ou inconsistentes (campos faltando, duplicados).	**Acessibilidade dos dados** Os dados existem e são bons, mas não podem ser usados com agilidade (pedido formal que leva semanas).

Lacuna de processos Pessoas e tecnologia boas, mas cada projeto segue fluxo próprio de aprovação, teste e implantação, e as entregas atrasam.	**Lacuna de governança** Projetos sem política de uso, sem avaliação de riscos ou sem definição de quem aprova cada caso, ou que não saem do piloto por falta de responsáveis e critérios.
Patrocinador executivo Garantir recursos, alinhar prioridades e remover barreiras entre as áreas; quando sai, é preciso definir logo outro.	**Campeões de IA** Multiplicar a adoção nas áreas e compartilhar boas práticas, sem aprovar projetos sozinhos nem substituir a autoridade executiva ou a governança.
Perspectiva de plataforma (AWS CAF) Base técnica: arquitetura, dados e ambientes para construir e operar as soluções de IA.	**Perspectiva de operações (AWS CAF)** Manter as soluções funcionando no dia a dia: monitoramento, incidentes e níveis de serviço.

Números que caem

Item	Valor
Peso do domínio na prova	**24%**
Perspectivas do AWS CAF	**6: negócio, pessoas, governança, plataforma, segurança e operações**
Dimensões de prontidão citadas no guia do exame	**5: alinhamento da liderança, qualidade dos dados, preparação cultural, infraestrutura técnica e frameworks de governança**
Fases da abordagem iterativa citada no guia	**4: envisionar, experimentar, lançar e escalar**
Dimensões das lacunas de capacidade	**4: pessoas, processos, tecnologia e governança**

Alertas de prova (pegadinhas)

⚠ **Alerta de prova:** Iniciar mais pilotos em todas as áreas para aumentar as chances de sucesso.
Como não cair: Mais pilotos sem padrão reforçam o estágio de experimentação dispersa. O próximo passo é governança comum e levar os pilotos de maior valor para produção.

⚠ **Alerta de prova:** Esperar uma plataforma completa de IA antes do primeiro caso de uso ou antes de liberar novas ideias.
Como não cair: Esperar a plataforma completa atrasa valor e aprendizado. Comece com vitórias rápidas: dados prontos, baixo risco e ganho mensurável.

⚠ **Alerta de prova:** Resolver o gargalo do CoE contratando mais gente para ele executar tudo ou centralizando ainda mais.
Como não cair: Crescer o CoE para executar tudo mantém a dependência central. O ajuste é o CoE definir padrões e as áreas executarem; extinguir o CoE também é errado, porque elimina padrões e coordenação.

⚠ **Alerta de prova:** Dizer aos funcionários que nenhum cargo será afetado, para reduzir a ansiedade.
Como não cair: Prometer o que pode não se confirmar destrói a confiança. A resposta certa é informar com honestidade prazos, impactos e planos de requalificação.

⚠ **Alerta de prova:** Punir publicamente projetos que não atingiram a meta ou exigir garantia de sucesso antes de aprovar.
Como não cair: Isso reforça o medo de errar, barreira cultural citada no guia. A liderança deve tornar a experimentação segura e reconhecer e compartilhar os aprendizados.

⚠ **Alerta de prova:** Tornar o uso da IA obrigatório e medir quem deixa de usar, ou premiar quem usa mais vezes.
Como não cair: Obrigar sem envolver gera adesão só formal, e premiar volume estimula uso sem valor. Envolva as pessoas, mostre ganhos reais e ligue as metas a resultados.

⚠ **Alerta de prova:** Deixar os campeões de IA aprovarem sozinhos os projetos da área ou assumirem as decisões executivas quando o patrocinador sai.
Como não cair: Campeões incentivam a adoção e compartilham práticas. Aprovação continua com a governança, e a autoridade executiva exige um novo patrocinador.

⚠ **Alerta de prova:** Desligar o processo antigo no mesmo dia em que a IA entra em produção, para economizar logo.
Como não cair: Isso elimina a alternativa em caso de falha. Desative de forma gradual, após metas atingidas por um período, com plano de retorno e comunicação às equipes.

⚠ **Alerta de prova:** Piloto aprovado: remover controles, reduzir testes ou apenas liberar para mais usuários.
Como não cair: Produção exige mais controles e testes, não menos: governança, monitoramento, suporte, documentação, treinamento da operação e níveis de serviço.

⚠ **Alerta de prova:** Treinar todos os funcionários em programação ou em arquitetura dos modelos.
Como não cair: A maioria precisa de uso responsável, verificação das respostas, limites da ferramenta e cuidados com dados. As trilhas variam por papel.

⚠ **Alerta de prova:** Usar o AWS CAF para calcular o preço mensal, escolher algoritmos ou certificar a empresa.
Como não cair: O CAF é um guia de planejamento organizacional por perspectivas. Preço se estima na AWS Pricing Calculator, e certificação de gestão de IA vem de normas como a ISO/IEC 42001.

⚠ **Alerta de prova:** Achar que documentos digitalizados já estão prontos ou que a IA vai organizá-los sozinha, sem esforço.
Como não cair: Estar em formato digital não é estar pronto. A preparação exige planejamento, validação e responsáveis, e afeta prazo, custo e viabilidade.

⚠ **Alerta de prova:** Levar as soluções do hackathon direto para produção ou usar o evento para avaliar desempenho individual.
Como não cair: O hackathon serve para gerar e testar ideias rápido, unindo negócio e tecnologia. Protótipos ainda passam por avaliação, governança e preparação.

⚠ **Alerta de prova:** Copiar os dados para planilhas acessíveis a todos ou para a máquina de cada analista, para resolver disputas ou falta de acesso.
Como não cair: Cópias sem regras criam silos, dados desatualizados e riscos de segurança. A solução estruturada é acordo de compartilhamento com regras de uso e responsáveis, ou catálogo de dados.

Como cai na prova

📝 **Como cai:** Cenário diz o que está bom (diretoria alinhada, dados bons, plataforma existe) e pergunta qual dimensão ou lacuna precisa de atenção.
Como resolver: Risque cada alternativa que o enunciado declara como resolvida. Entre as que sobram, escolha a que explica o sintoma descrito: ninguém usa = cultura; ninguém sabe aplicar = pessoas; fluxos diferentes = processos; sem responsáveis = governança.

📝 **Como cai:** Descrição da situação da empresa e pergunta sobre o estágio de maturidade ou o próximo passo.
Como resolver: Pilotos isolados sem padrão = experimentação; produção em várias áreas com governança comum e métricas = escala corporativa. O próximo passo é sempre o realista e incremental: governança comum e produção dos melhores pilotos, ou fundamentos se o estágio é inicial.

📝 **Como cai:** Reação humana à IA (medo de demissão, resistência, desconfiança, falta de tempo, gerentes ameaçados) e pergunta pela ação da liderança.
Como resolver: Elimine punir, obrigar, silenciar, prometer que nada muda, substituir pessoas e esperar passar. Fique com a alternativa que envolve as pessoas, comunica com transparência e oferece capacitação ou tempo.

📝 **Como cai:** Pergunta direta: no AWS CAF, qual perspectiva trata de X.
Como resolver: Mapeie: estratégia e portfólio = negócio; cultura e habilidades = pessoas; riscos, benefícios e gastos = governança; base técnica = plataforma; proteção de dados e identidades = segurança; monitoramento, incidentes e níveis de serviço = operações.

📝 **Como cai:** Iniciativa em transição (piloto aprovado, expansão para filiais ou países, processo antigo a desligar) e pergunta pelo cuidado necessário.
Como resolver: Procure a alternativa gradual e controlada: ondas, contingência, métricas, documentação, níveis de serviço, plano de retorno. Em expansão internacional, avalie idiomas, leis locais, dados e suporte em cada país.

📝 **Como cai:** Problema de dados descrito por sintoma e pergunta pelo aspecto ou pela iniciativa que resolve.
Como resolver: Incompleto = qualidade; existe mas demora = acessibilidade; definições diferentes por área = silos e cadastro mestre; sem dono = definir donos e regras; não se sabe o que existe = catálogo; não se sabe de onde veio = linhagem.

Gatilhos do enunciado

Se o enunciado trouxer...	Pense em
Diretoria alinhada e dados bons, mas os funcionários desconfiam e evitam as ferramentas	**Preparação cultural**
Diretores com prioridades opostas disputando o mesmo orçamento	**Alinhamento da liderança; definir no nível executivo objetivos, prioridades e métricas comuns**
Servidores antigos sem integração, cada extração leva semanas	**Infraestrutura técnica e acesso aos dados**
Vários pilotos isolados por área, sem padrões nem governança comum	**Estágio de experimentação**
Casos em produção em várias áreas, com governança comum e métricas de valor	**Implantação em escala corporativa**
Ninguém sabe que dados existem, onde estão e quem responde por eles	**Catálogo de dados com descrições, localização e responsáveis**
Auditores perguntam de onde vieram os dados e por quais transformações passaram	**Linhagem de dados**
O modelo quebra sempre que outra área muda o formato dos dados	**Acordo entre áreas sobre formato, qualidade e mudanças**

AWS CAF: cultura, liderança e habilidades / riscos, benefícios e gastos / proteção de dados e identidades / estratégia e portfólio	**Pessoas / Governança / Segurança / Negócio**
Funcionários dizem que não têm tempo para aprender IA	**Reservar horas na rotina para aprendizado prático com casos reais**
Aprendizados de um time não chegam aos outros e os erros se repetem	**Comunidade de prática com registro de lições aprendidas; revisões ao fim de cada fase**
Mais pedidos de IA do que capacidade, cada diretor achando o seu o mais urgente	**Comitê de portfólio multifuncional com critérios comuns**
Todos os casos passam pelo mesmo processo longo, inclusive os de baixo risco	**Caminho mais simples para casos de baixo risco (classificação de risco)**
Diretoria quer ver todas as iniciativas de IA em um só lugar	**Painel de portfólio com valor, custo, risco e estágio**

🎯 **Se você lembrar de só uma coisa deste capítulo...**

- O enunciado entrega o que está bom: a resposta é a dimensão que sobra.
- Escala é gradual: vitórias rápidas, ondas, contingência e desligamento do antigo só com plano de retorno.
- CoE habilita e padroniza; campeões multiplicam a adoção; nenhum dos dois substitui governança ou patrocínio executivo.
- Medo e resistência se tratam com transparência, envolvimento e capacitação, nunca com punição, silêncio ou promessa vazia.
- AWS CAF tem seis perspectivas e serve para planejar capacidades, não para preço, algoritmo ou certificação.

Essenciais para a prova (checklist)

☐ Sei listar as cinco dimensões de prontidão e reconhecer cada uma pelo sintoma do cenário.

☐ Sei identificar a lacuna (pessoas, processos, tecnologia ou governança) quando o enunciado diz o que já está bom.

☐ Sei diferenciar experimentação de escala corporativa e apontar o próximo passo realista.

☐ Sei separar qualidade, acessibilidade e silos de dados e associar catálogo, linhagem, cadastro mestre e acordos de compartilhamento ao problema certo.

☐ Sei as seis perspectivas do AWS CAF e o que cada uma trata.

☐ Sei o papel do patrocinador executivo, dos campeões de IA e do CoE, e quando o CoE deve ser centralizado ou federado.

☐ Sei a ordem e o foco de envisionar, experimentar, lançar e escalar.

- ☐ Sei o que muda na passagem de piloto para produção e quais critérios levam uma POC à produção.
- ☐ Sei escolher a intervenção certa para medo de errar, aversão a risco e resistência à mudança.
- ☐ Sei desenhar letramento por papel: executivos, gestores, usuários e times técnicos.

Questões de revisão do capítulo

R4.1 · Prontidão, Liderança e Transformação com IA

A diretoria está alinhada e os dados são bons, mas os funcionários desconfiam da IA e evitam usar as ferramentas. Qual dimensão de prontidão precisa de atenção?

A. Preparação cultural
B. Infraestrutura técnica
C. Alinhamento da liderança
D. Qualidade dos dados

R4.2 · Prontidão, Liderança e Transformação com IA

Uma empresa quer usar IA generativa em escala com dados sensíveis de clientes. Qual requisito de infraestrutura é essencial?

A. Uma rede sem fio mais rápida nos escritórios da empresa
B. Computadores mais potentes para cada funcionário da empresa
C. Acesso seguro e controlado às fontes de dados
D. Monitores maiores para exibir melhor as respostas da IA

R4.3 · Prontidão, Liderança e Transformação com IA

Ao anunciar um programa de IA, qual comunicação ajuda a reduzir a resistência dos funcionários?

A. Detalhes técnicos sobre os modelos que serão usados
B. Metas de redução de quadro para cada área da empresa
C. Prazos, resultados esperados e efeitos nos papéis
D. Uma mensagem única dizendo que a IA vai aumentar a eficiência

R4.4 · Prontidão, Liderança e Transformação com IA

Por que envolver o jurídico e o compliance desde o início de um projeto de IA, e não só antes do lançamento?

A. Para atrasar o projeto até que todas as dúvidas desapareçam
B. Para que eles escrevam o código da solução junto com a equipe
C. Para identificar requisitos e riscos cedo e evitar retrabalho
D. Para transferir a eles a responsabilidade pelo resultado do projeto

Respostas comentadas

R4.1. A) Preparação cultural

Por quê: Dimensões de prontidão citadas no guia: alinhamento da liderança, qualidade dos dados, preparação cultural, infraestrutura técnica e frameworks de governança.

✗ B) Incorreta. O cenário não aponta problemas técnicos de infraestrutura.

✗ C) Incorreta. O cenário diz que a diretoria está alinhada.

✗ D) Incorreta. O cenário diz que os dados são bons.

> **Dica de prova:** Tudo pronto, mas ninguém usa: preparação cultural.

R4.2. C) Acesso seguro e controlado às fontes de dados

Por quê: O guia do exame pede avaliar os requisitos de tecnologia e infraestrutura para iniciativas de IA. Com dados sensíveis, integração segura e controle de acesso são prioridade.

✗ A) Incorreta. Rede sem fio não resolve segurança nem integração de dados.

✗ B) Incorreta. Serviços de IA na nuvem não dependem de computadores potentes para cada usuário.

✗ D) Incorreta. Monitores não são requisito de infraestrutura para IA.

> **Dica de prova:** IA com dados sensíveis: integração segura e acesso controlado.

R4.3. C) Prazos, resultados esperados e efeitos nos papéis

Por quê: Estratégias de comunicação transparentes tratam as preocupações da força de trabalho, explicando o que muda, quando e como as pessoas serão apoiadas.

✗ A) Incorreta. Detalhes técnicos não tratam as preocupações dos funcionários.

✗ B) Incorreta. Anunciar metas de corte sem contexto aumenta o medo e a resistência.

✗ D) Incorreta. Uma mensagem única e genérica não responde às preocupações.

> **Dica de prova:** Comunicar IA para funcionários: quando, o que se espera e o que muda para cada um.

R4.4. C) Para identificar requisitos e riscos cedo e evitar retrabalho

Por quê: O guia do exame pede montar times multifuncionais com negócio, tecnologia, jurídico e compliance, e com padrões claros de responsabilidade.

✗ A) Incorreta. O objetivo é orientar, não atrasar.

✗ B) Incorreta. O papel do jurídico e do compliance não é programar.

✗ D) Incorreta. A responsabilidade pelo resultado continua com o dono de negócio.

> **Dica de prova:** Jurídico e compliance desde o início: menos retrabalho.

DIA DA PROVA

Estratégia de prova

O AIB-C01 avalia um profissional de negócio que avalia, defende ou escala IA, trabalha ao lado de times técnicos e não constrói as soluções; a prova prioriza decisão estratégica sobre implementação técnica. Quase toda pergunta é um cenário corporativo curto que pede a ação mais adequada, a dimensão afetada, o KPI certo ou o serviço em nível estratégico. A resposta certa costuma ser equilibrada, mensurável, gradual e responsável, com dono definido e controles proporcionais. As distratoras são extremos (proibir tudo, liberar tudo, punir, esperar, cancelar), transferência de responsabilidade ao fornecedor, métricas de atividade e detalhes técnicos fora do escopo, como configurar serviços ou programar modelos. Nas contas de ROI e payback, as alternativas erradas reproduzem erros típicos de cálculo.

Como responder

✓ Planeje o tempo pelo formato atual: 85 questões em 170 minutos, cerca de 2 minutos por questão; marque as contas e os cenários longos para revisar no fim.

✓ Responda todas: questão sem resposta conta como errada e não há penalidade por chute.

✓ Lembre que a pontuação vai de 100 a 1000 com mínimo de 700 e é compensatória: não precisa passar em cada domínio, só no total; o domínio de Estratégia pesa 28% e os outros três, 24% cada.

✓ Em múltipla resposta há duas ou mais corretas entre cinco ou mais opções: marque exatamente a quantidade pedida (por exemplo, Escolha DUAS), avaliando cada alternativa sozinha contra o critério do enunciado; só pontua se marcar todas as corretas.

✓ Leia o que o enunciado declara como resolvido (diretoria alinhada, dados bons, tecnologia adequada) e risque as alternativas que repetem isso.

✓ Elimine primeiro extremos e transferências de responsabilidade: proibir, punir, esperar, cancelar sem análise, deixar a decisão com o fornecedor ou com a TI sozinha.

✓ Descarte respostas técnicas fora do escopo do candidato, como configurar servidores, escrever código ou ajustar hiperparâmetros; o exame cobra decisão de negócio.

✓ Em cálculos, escreva a fórmula (ROI divide pelo custo; payback é investimento sobre ganho líquido mensal) e confira se a sua resposta não é a distratora do erro clássico.

Erros que derrubam candidatos

✗ Dividir o ganho pelo benefício ou esquecer custos de operação ao calcular o ROI.

✗ Escolher como KPI uma métrica de atividade, como prompts, logins, sugestões geradas ou tokens.

✗ Trocar RAG por ajuste fino e achar que temperatura ou modelo maior resolve alucinação e falta de informação.

✗ Aceitar um aviso ao usuário como solução para um problema de privacidade, equidade ou segurança.

✗ Achar que a responsabilidade é do fornecedor, da AWS ou do algoritmo.

✗ Confundir AWS Pricing Calculator (estimar antes) com AWS Cost Explorer (analisar o gasto real).

✗ Escolher a alternativa radical em mudança organizacional: punir, obrigar, prometer que nada muda ou desligar o processo antigo no primeiro dia.

Na semana da prova

- ☐ Refaça as contas de ROI, payback, economia por contato e horas liberadas até sair de cabeça sem cair nas distratoras.
- ☐ Revise as tabelas de associação: sintoma para dimensão de IA responsável, perspectiva do AWS CAF, dimensão de prontidão e lacuna de capacidade.
- ☐ Revise as escolhas de serviço e custo: Bedrock, SageMaker AI, Quick, Guardrails, Knowledge Bases, Pricing Calculator, Cost Explorer, Marketplace, consumo, instância, assento, lote e throughput provisionado.
- ☐ Releia RAG x ajuste fino, regras x IA, agente x assistente e drift x sobreajuste, que se repetem em vários formatos.
- ☐ Faça uma prova completa cronometrada no ritmo de cerca de 2 minutos por questão e revise o porquê de cada erro, não só a resposta.
- ☐ Na véspera, só revisão leve das listas de lembrar e dos macetes; durma bem.

REVISÃO FINAL

Folha de cola

Leia na véspera e na manhã da prova: os gatilhos e as pegadinhas mais importantes de cada domínio.

1. Estratégia de IA e Criação de Valor

‣ Estimar o custo mensal antes de construir, com base no volume esperado → **AWS Pricing Calculator**

‣ Quais serviços mais pesaram na fatura e como o gasto evoluiu mês a mês → **AWS Cost Explorer**

‣ Uso de computação previsível e constante ao longo do próximo ano → **Savings Plans com compromisso de uso**

‣ Milhões de documentos por mês, resultado usado só no dia seguinte → **Processamento em lote no Amazon Bedrock**

‣ Volume alto e constante no Bedrock, precisa de capacidade garantida → **Throughput provisionado**

⚠ Isso é a razão benefício/custo. ROI desconta o custo antes de dividir: (600 - 400) / 400 = 50%.

⚠ O denominador do ROI é sempre o custo total, nunca o benefício.

⚠ Some implantação e operação do período: (800 - 700) / 700 = cerca de 14%. Esquecer custos recorrentes é erro comum.

2. Fundamentos e Letramento em IA

‣ Políticas que mudam todo mês, informação recente, citar a fonte de cada resposta → **RAG; no Bedrock, Knowledge Bases**

‣ Tom de voz da marca, formato padrão de laudo, muitos exemplos de qualidade → **Ajuste fino**

‣ Três reclamações já classificadas incluídas no prompt → **Prompt com poucos exemplos (few-shot)**

‣ Atue como analista de crédito sênior → **Definição de papel no prompt**

‣ Desconsidera cláusulas do início ou esquece instruções dadas no começo da conversa → **Janela de contexto excedida; enviar só o relevante, resumir histórico, repetir instruções**

⚠ O modelo nunca conheceu a informação, e ajustar periodicamente deixa-o desatualizado. Informação recente ou dinâmica pede RAG.

⚠ RAG traz conhecimento externo. Se o conhecimento já está no prompt e o problema é estilo e formato, a técnica é ajuste fino (ou exemplos no prompt).

⚠ Temperatura controla a variabilidade. Não dá acesso a informação, não resolve perda de contexto nem formato, e pode aumentar respostas inventadas.

3. Governança e Liderança em IA Responsável

‣ Explicar ao cliente por que o crédito foi negado → **Explicabilidade**

‣ Modelo aprova menos uma região com perfis financeiros parecidos → **Equidade**

‣ Clientes conversam sem saber que não é uma pessoa → **Transparência**

‣ Monitorar, interromper e corrigir um agente sempre que necessário → **Controlabilidade**

‣ Doses de medicamento, alergias omitidas, procedimentos em redes elétricas → **Segurança (safety)**

⚠ O viés pode surgir com o tempo (bias drift). A governança deve garantir monitoramento contínuo de viés.

⚠ A média pode esconder erros concentrados em um grupo. Teste e compare o desempenho por grupo.

⚠ O CEP continua indicando o bairro e a localização completa identifica a pessoa. O efeito permanece.

4. Prontidão, Liderança e Transformação com IA

‣ Diretoria alinhada e dados bons, mas os funcionários desconfiam e evitam as ferramentas → **Preparação cultural**

‣ Diretores com prioridades opostas disputando o mesmo orçamento → **Alinhamento da liderança; definir no nível executivo objetivos, prioridades e métricas comuns**

‣ Servidores antigos sem integração, cada extração leva semanas → **Infraestrutura técnica e acesso aos dados**

‣ Vários pilotos isolados por área, sem padrões nem governança comum → **Estágio de experimentação**

‣ Casos em produção em várias áreas, com governança comum e métricas de valor → **Implantação em escala corporativa**

⚠ Mais pilotos sem padrão reforçam o estágio de experimentação dispersa. O próximo passo é governança comum e levar os pilotos de maior valor para produção.

⚠ Esperar a plataforma completa atrasa valor e aprendizado. Comece com vitórias rápidas: dados prontos, baixo risco e ganho mensurável.

⚠ Crescer o CoE para executar tudo mantém a dependência central. O ajuste é o CoE definir padrões e as áreas executarem; extinguir o CoE também é errado, porque elimina padrões e coordenação.

GABARITO

Respostas do diagnóstico

D1. A) Dúvidas resolvidas sem abrir chamado → revise o capítulo 1

Por quê: O guia do exame pede KPIs para iniciativas de IA. Em autoatendimento, a taxa de contenção, combinada com a satisfação, mostra o valor gerado.

✗ B) Incorreta. O tamanho da base não mostra se as dúvidas foram resolvidas.

✗ C) Incorreta. Visitas ao site não mostram resolução de dúvidas.

✗ D) Incorreta. Tempo de carregamento é desempenho técnico do site.

> **Dica de prova:** Autoatendimento: meça quantas dúvidas são resolvidas sem chamado.

D2. B) Ganhos rápidos e poucas apostas estratégicas maiores → revise o capítulo 1

Por quê: O guia do exame pede priorizar iniciativas de IA e escalar a partir de vitórias de curto prazo, sem perder o alinhamento estratégico.

✗ A) Incorreta. Só pequenas melhorias limitam o potencial da IA.

✗ C) Incorreta. Sem entregas por dois anos, a iniciativa tende a perder apoio.

✗ D) Incorreta. Muitas iniciativas sem priorização dispersam recursos.

> **Dica de prova:** Portfólio: ganhos rápidos mais poucas apostas estratégicas.

D3. A) Contratar um Savings Plan com compromisso de uso → revise o capítulo 1

Por quê: Estratégias de otimização de custos incluem escolher o modelo de preço adequado ao padrão de uso. Para consumo estável, Savings Plans reduzem o custo em relação ao sob demanda.

✗ B) Incorreta. Sob demanda é flexível, mas costuma ser mais caro para uso constante e previsível.

✗ C) Incorreta. Instâncias maiores aumentam o custo sem necessidade.

✗ D) Incorreta. Adiar a análise impede otimizar os custos durante o ano.

> **Dica de prova:** Uso previsível: compromisso com desconto (Savings Plans). Uso incerto: sob demanda.

D4. A) O ajuste fino adapta um modelo já treinado com dados específicos → revise o capítulo 2

Por quê: O guia do exame pede reconhecer técnicas de adaptação de modelos, como RAG e ajuste fino. Aproveitar modelos pré-treinados reduz custo e prazo em relação a treinar do zero.

✗ B) Incorreta. Treinar do zero costuma exigir muito mais dados, tempo e custo.

✗ C) Incorreta. São abordagens diferentes, com custos e esforços diferentes.

✗ D) Incorreta. O ajuste fino depende de dados de exemplo.

> **Dica de prova:** Ajuste fino: partir de um modelo pronto e especializar.

D5. B) Dividir a tarefa em etapas, com instruções claras para cada uma → revise o capítulo 2

Por quê: Princípios básicos de engenharia de prompt incluem clareza, contexto e divisão de tarefas complexas em passos. O guia do exame pede aplicar esses princípios para obter os resultados desejados.

✗ A) Incorreta. Repetir o mesmo prompt não corrige a falta de clareza da instrução.
✗ C) Incorreta. Trocar o idioma não resolve uma instrução sobrecarregada.
✗ D) Incorreta. Sem o relatório, o modelo não tem o conteúdo necessário para a análise.

> **Dica de prova:** Tarefa complexa: quebre em etapas.

D6. D) Amazon Bedrock Knowledge Bases → revise o capítulo 2

Por quê: O guia do exame cita Guardrails e Knowledge Bases entre os recursos do Amazon Bedrock que o candidato deve conhecer em nível de negócio. Knowledge Bases é o caminho gerenciado para RAG no Bedrock.

✗ A) Incorreta. O Cost Explorer analisa gastos realizados.
✗ B) Incorreta. O Guardrails aplica controles de segurança e de conteúdo, mas não conecta o modelo aos documentos.
✗ C) Incorreta. A Pricing Calculator estima custos, não adiciona conhecimento ao assistente.

> **Dica de prova:** RAG no Bedrock: Knowledge Bases. Controles de conteúdo: Guardrails.

D7. B) LGPD → revise o capítulo 3

Por quê: O guia do exame pede identificar e tratar riscos de compliance regulatório em processos que usam IA. Para dados pessoais no Brasil, a referência é a LGPD.

✗ A) Incorreta. A SOX trata de controles financeiros de empresas listadas nos Estados Unidos.
✗ C) Incorreta. A HIPAA trata de dados de saúde nos Estados Unidos.
✗ D) Incorreta. O PCI DSS é um padrão para dados de cartão de pagamento, não uma lei geral de dados pessoais.

> **Dica de prova:** Dados pessoais no Brasil: LGPD.

D8. D) Definir uma política de retenção com prazo e finalidade para os registros → revise o capítulo 3

Por quê: Privacidade é uma dimensão da IA responsável. Registros de IA precisam de prazo, finalidade e proteção definidos.

✗ A) Incorreta. Guardar sem prazo contraria a minimização e aumenta o risco.
✗ B) Incorreta. Isso não é uma prática de retenção adequada e pode expor dados.
✗ C) Incorreta. Apagar no mesmo dia pode violar exigências de auditoria.

> **Dica de prova:** Registros com dados pessoais: prazo e finalidade definidos.

D9. C) A empresa cliente, dona dos dados e do uso → revise o capítulo 3

Por quê: No modelo de responsabilidade compartilhada, a AWS cuida da segurança da nuvem, e o cliente cuida do que coloca e faz na nuvem, incluindo dados, acessos e uso responsável dos modelos.

✗ A) Incorreta. Usuários finais não têm como avaliar os dados de treinamento.

✗ B) Incorreta. A AWS é responsável pela segurança da nuvem, não pelo conteúdo dos dados do cliente.

✗ D) Incorreta. O fornecedor de hardware não tem relação com a qualidade dos dados.

> **Dica de prova:** Dados e uso do modelo: responsabilidade do cliente.

D10. B) Qualidade dos dados → revise o capítulo 4

Por quê: Prontidão de dados envolve qualidade, acessibilidade e os efeitos dos silos. Dados incompletos reduzem a confiabilidade dos modelos.

✗ A) Incorreta. O cenário não trata de quem é dono dos dados.

✗ C) Incorreta. O problema não é falta de registros, e sim registros incompletos.

✗ D) Incorreta. Os dados estão disponíveis; o problema é que estão incompletos.

> **Dica de prova:** Campos faltando: qualidade dos dados.

D11. B) Adoção, qualidade, valor de negócio e custo → revise o capítulo 4

Por quê: O guia do exame pede estabelecer mecanismos contínuos de feedback e métricas de sucesso para acompanhar o progresso e o valor de longo prazo das iniciativas de IA.

✗ A) Incorreta. Reuniões e apresentações medem atividade, não resultado.

✗ C) Incorreta. Documentação é importante, mas não mede o valor da iniciativa.

✗ D) Incorreta. Testar muitos modelos não indica valor gerado.

> **Dica de prova:** Acompanhar IA: adoção, qualidade, valor e custo.

D12. B) Líderes de negócio, tecnologia, dados, jurídico e pessoas → revise o capítulo 4

Por quê: O guia do exame cita alinhamento da liderança, qualidade dos dados, preparo cultural, infraestrutura técnica e governança como dimensões de prontidão.

✗ A) Incorreta. O financeiro sozinho não avalia dados, cultura e tecnologia.

✗ C) Incorreta. Um consultor pode apoiar, mas precisa ouvir as áreas.

✗ D) Incorreta. A tecnologia sozinha não avalia cultura, governança e estratégia.

> **Dica de prova:** Avaliação de prontidão: várias áreas participando.

Agora, treine no formato real

Faça as 6 provas completas do curso AIB-C01: 510 questões inéditas, tempo do exame, explicação de cada alternativa e link da documentação para revisar.

Ver as 6 provas completas e o cupom do mês →

Até o dia da prova, uma revisão por dia no seu feed

Instagram @strecht.simulados

Um post de prova por dia: questão comentada, pegadinhas e o "não confunda" de cada certificação.

Seguir no Instagram →

YouTube Strecht Simulados

Shorts diários com pegadinhas de prova e, aos sábados, um simulado comentado.

Inscrever-se no canal →

Material independente de preparação, sem afiliação ou endosso de Amazon Web Services (AWS). As questões são inéditas e não reproduzem a prova real.