

Guia de preparação AB-731

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.

3 capítulos por domínio	**21** 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 AB-731: 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 50 questões e 45 minutos, no formato do exame. Referência de aprovação: 70%.

Domínio (capítulo)	Peso	Neste guia
1. Identificar benefícios, recursos e oportunidades dos aplicativos e serviços de IA da Microsoft	35-40%	19 conceitos · 14 pegadinhas
2. Identificar o valor de negócio das soluções de IA generativa	35-40%	18 conceitos · 14 pegadinhas
3. Identificar uma estratégia de implementação e adoção dos aplicativos e serviços de IA da Microsoft	20-25%	19 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 9 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 · Identificar benefícios, recursos e oportunidades dos aplicativos e serviços de IA da Microsoft

A varejista Pão & Cia já centraliza seus dados e análises no Microsoft Fabric e quer treinar um modelo de previsão de demanda com os próprios dados. Segundo o Cloud Adoption Framework, qual solução é a indicada?

A. Copilot Studio, com um agente que aprende sozinho a prever a demanda
B. Microsoft 365 Copilot, que treina modelos de previsão dentro do Excel
C. Azure AI Content Safety, que detecta padrões nos dados de vendas
D. Microsoft Fabric, pois a empresa já usa a plataforma para seus dados e suas análises

D2 · Identificar benefícios, recursos e oportunidades dos aplicativos e serviços de IA da Microsoft

O CTO do Banco Digital Nuvem Azul construiu um aplicativo sobre uma versão específica de modelo no Microsoft Foundry. O que a liderança deve considerar sobre o ciclo de vida desse modelo?

A. Trocar de modelo toda semana, para usar a novidade mais recente do catálogo
B. Acompanhar as datas de aposentadoria e planejar a migração para versões novas
C. Congelar o app e proibir atualizações para evitar mudanças no comportamento do modelo
D. Considerar que a versão escolhida ficará disponível por tempo indeterminado na plataforma

D3 · Identificar benefícios, recursos e oportunidades dos aplicativos e serviços de IA da Microsoft

Na rede de concessionárias Auto Prime, todos os vendedores têm licença do Microsoft 365 Copilot e usam no Teams um agente criado no Copilot Studio, fundamentado em dados do locatário pelo Microsoft Graph. O CFO pergunta se esse uso consome Copilot Credits. Qual resposta está correta?

A. Não; agentes do Copilot Studio são gratuitos em qualquer canal, inclusive sites públicos
B. Não; para licenciados, agentes no Copilot, Teams ou SharePoint com grounding do Graph não consomem créditos
C. Sim; cada conversa consome créditos, mesmo para usuários com licença do Copilot
D. Sim; agentes funcionam só com pacotes pré-pagos comprados por usuário

D4 · Identificar o valor de negócio das soluções de IA generativa

O comitê executivo da Indústria Gráfica Prelo quer cortar investimentos em governança de dados e segurança dos projetos de IA para acelerar entregas. Segundo a Microsoft, quais consequências práticas fracas de dados e segurança podem trazer?

A. Viés nas saídas, vazamento de dados e falhas de compliance
B. Respostas mais rápidas e menor custo de armazenamento
C. Licenças mais baratas e menos treinamentos obrigatórios
D. Maior adoção, já que menos controles facilitam o uso no dia a dia

D5 · Identificar o valor de negócio das soluções de IA generativa

O CISO do Banco Cooperativo Raiz quer estar preparado para incidentes de segurança envolvendo os sistemas de IA, como comprometimento de modelos ou vazamento de dados por respostas. Segundo o Cloud Adoption Framework, qual prática é a recomendada?

A. Desligar os logs dos modelos para não gerar alertas desnecessários
B. Usar o mesmo plano genérico de TI, sem considerar as ameaças próprias de IA
C. Esperar o primeiro incidente para então decidir quem será acionado
D. Criar e testar planos de resposta para incidentes de IA, com escalonamento claro

D6 · Identificar o valor de negócio das soluções de IA generativa

A diretoria de inovação do Grupo Segurança Total quer priorizar onde investir primeiro em grounding das soluções de IA generativa. Segundo a Microsoft, em que situações o grounding é especialmente importante?

A. Textos de ficção e brainstorming criativo sem compromisso com fatos
B. Jogos internos de integração, em que a precisão não faz diferença
C. Rascunhos pessoais descartáveis, que ninguém além do autor vai ler ou revisar depois
D. Setores regulados, cenários de compliance e aplicações voltadas a clientes

D7 · Identificar uma estratégia de implementação e adoção dos aplicativos e serviços de IA da Microsoft

A montadora Rodas do Sul vai usar IA para ajustar automaticamente a linha de montagem robotizada e quer que o sistema se comporte de forma segura mesmo em situações inesperadas. Qual prática está alinhada ao princípio de confiabilidade e segurança?

A. Liberar o sistema e corrigir os problemas depois que acidentes acontecerem
B. Testar com rigor, inclusive em cenários imprevistos, e definir quando passar o controle a uma pessoa
C. Garantir que o sistema nunca transfira o controle a humanos, para evitar atrasos na linha de montagem
D. Testar o sistema em laboratório, com dados ideais e sem variações

D8 · Identificar uma estratégia de implementação e adoção dos aplicativos e serviços de IA da Microsoft

A Associação Comercial Litoral Norte tem assinaturas do Microsoft 365 Business Standard e quer oferecer a todos os colaboradores um chat de IA seguro, sem custo adicional nesta primeira fase. Qual opção atende a esse objetivo?

A. Copilot Studio, com um pacote pré-pago de Copilot Credits para cada funcionário
B. Copilot Chat, incluído sem custo extra para usuários com assinatura elegível
C. Microsoft 365 Copilot como add-on, contratado por usuário para a associação
D. Microsoft Foundry, com cobrança por tokens em uma assinatura do Azure

D9 · Identificar uma estratégia de implementação e adoção dos aplicativos e serviços de IA da Microsoft

Em 2016, a Microsoft lançou um chatbot chamado Tay, que aprendia com conversas públicas e, em menos de 24 horas, passou a repetir conteúdo de ódio. Que lição a diretora de inovação da agência Criativa Ponto deve aplicar ao lançar um assistente para o público?

A. Antecipar o uso indevido e projetar salvaguardas desde o primeiro dia
B. Evitar que chatbots aprendam com dados, usando um conjunto fixo de regras
C. Lançar rápido e corrigir depois, já que filtros atrasam a inovação da agência
D. Restringir a IA a públicos internos, onde não existe comportamento malicioso

01

CAPÍTULO 1 · 35-40% DA PROVA

Identificar benefícios, recursos e oportunidades dos aplicativos e serviços de IA da Microsoft

Este domínio é um grande exercício de casamento: dado o cenário, qual produto, experiência ou agente da Microsoft resolve. Preciso ter na ponta da língua o que o Copilot Chat sem licença faz e o que a licença paga acrescenta, o Copilot em cada app (Word, Excel, PowerPoint, Outlook, Teams), Researcher x Analyst, Pages x Notebooks x Search, o Microsoft Graph e os Copilot connectors. Depois vem o lado Azure: comprar, estender ou construir, agente declarativo x engine personalizada, Copilot Studio (SaaS) x Microsoft Foundry (PaaS) x infraestrutura, e qual Foundry Tool atende cada fluxo (Language, Translator, Speech, Vision, Document Intelligence, Content Understanding, AI Search, Content Safety). Onde mais erro: alternativas que parecem fazer algo parecido, como Translator para extrair campos ou Graph Data Connect no lugar de Copilot connectors. Regra que salva muitas: o Copilot sempre respeita as permissões existentes e os dados do cliente não treinam modelos de base.

Neste capítulo

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

Tópicos-chave

Microsoft Copilot Chat (sem licença paga)

Experiência de chat de IA segura incluída sem custo extra para quem entra com conta corporativa do Microsoft Entra e tem assinatura qualificada do Microsoft 365. É fundamentado principalmente na web; dados da empresa entram por upload de arquivo, colagem de texto ou agentes. No Outlook, usa e-mails, agenda, reuniões e chats do próprio usuário, sempre dentro das permissões. Está no app Copilot (web, desktop e celular), no Teams, no Outlook e no Edge; o modo Auto usa um roteador que escolhe o modelo conforme o prompt.

Licença paga do Microsoft 365 Copilot (agora Microsoft Copilot)

Acrescenta grounding amplo nos dados do Microsoft 365 via Graph, Work IQ e indexação semântica, acesso prioritário a modelos de raciocínio avançado e agentes mais avançados (como Researcher e Analyst). Inclui Copilot Chat, Copilot nos apps, capacidades de agentes do Copilot Studio quando usadas no Microsoft Copilot, Teams ou SharePoint e Copilot Analytics. Não dá acesso a conteúdo sem permissão nem isenta de retenção e rótulos. A mudança de nome não alterou segurança, compliance nem privacidade.

Copilot nos apps do Microsoft 365

Mapeamento pelo tipo de trabalho: documentos no Word (primeiro rascunho a partir de modelos), dados no Excel (previsão, orçamento, cenários), apresentações no PowerPoint, e-mail no Outlook (resumir conversas e sugerir respostas) e reuniões no Teams (resumir discussões, destacar decisões, acompanhar itens de ação). No PowerPoint: partir do arquivo do Word, usar estilos de título, manter imagens, começar pelo modelo da empresa e refinar o esboço. A disponibilidade nos apps varia conforme licença e configuração do locatário.

Benefícios centrais do Copilot

Aumentar a produtividade automatizando tarefas repetitivas, melhorar a qualidade das decisões com insights rápidos de dados e fortalecer a governança com compliance de nível empresarial. Embutir o Copilot nos apps acelera fluxos, eleva a qualidade e mantém mensagens e formatação consistentes. O Copilot é parceiro, não substituto: use para primeiros rascunhos e revise antes de compartilhar externamente.

Agentes Researcher e Analyst

Researcher faz pesquisas de várias etapas, sintetizando fontes da web e do trabalho (arquivos, e-mails, reuniões, chats) num relatório estruturado com citações; demora mais por design e faz perguntas de esclarecimento para orientar a pesquisa. Analyst transforma dados brutos (CSV, planilhas) em insights, calculando estatísticas, tendências e outliers e raciocinando em etapas como um cientista de dados. Para resumos rápidos, use o Copilot padrão.

Copilot Pages, Notebooks e Search

Pages transforma uma resposta do Copilot em página editável e compartilhável que a equipe coedita em tempo real. Notebooks reúne conversas, arquivos, páginas, notas de reunião e links de um projeto e responde com base nesse conteúdo (armazenamento no SharePoint ou OneDrive). Copilot Search é a busca universal com IA em linguagem natural sobre o conteúdo da organização, alcança sistemas de terceiros via Copilot connectors, respeita permissões e permite seguir para o chat.

Microsoft Graph e Work IQ

O Graph é o gateway para dados e inteligência de serviços como Microsoft 365 e Microsoft Entra; liga pessoas, documentos, conversas e agendas de forma segura e sensível a permissões. Já faz parte do Microsoft 365: não exige projeto de integração nem licença separada para o Copilot. Só chamadores autorizados acessam os dados. O Work IQ é a camada de inteligência que usa os dados do Graph para entender contexto e padrões de trabalho.

Graph API, Copilot connectors e Graph Data Connect

Graph API dá acesso a dados e insights; Copilot connectors trazem dados externos (Jira, ERP, CRM) para o Graph, para o Copilot e o Microsoft Search; Graph Data Connect leva dados do Graph em escala para o Azure (sentido inverso). Conectores sincronizados rastreiam e indexam o conteúdo no Graph; conectores federados consultam a fonte em tempo real via MCP, sem indexar. Ambos respeitam as permissões da origem.

Comprar, estender ou construir

Comprar: experiências prontas do Copilot para tarefas comuns, implantação rápida. Estender: conectar o Copilot a sistemas de negócio por agentes, Copilot connectors e ações. Construir: controle total para processos proprietários ou altamente regulados, com mais custo, prazo e habilidades especializadas. Ordem recomendada: comprar para tarefas comuns e depois estender ou construir; abordagem híbrida é comum.

Agente declarativo x agente de engine personalizada

Declarativo usa o orquestrador e os modelos do Copilot, herda compliance e segurança do Microsoft 365, não precisa de hospedagem extra e é projetado para uso individual; combina instruções, conhecimento e ações (API plugin para dados em tempo real). Cria-se com Agent Builder (low-code) ou pro-code (Agents Toolkit). Engine personalizada dá orquestração e modelos próprios, colaboração em grupo no Teams, mensagens proativas e uso fora do Microsoft 365, exigindo hospedagem, compliance e IA responsável próprios.

Microsoft Copilot Studio

Plataforma SaaS low-code para criar agentes e workflows: pode-se descrever o agente em linguagem natural e testá-lo antes de publicar. Conecta ERP, CRM e sistemas por conectores e ações e publica no Teams, no Microsoft Copilot, em sites e apps móveis. Agent flows combinam automação previsível com revisão humana; com conta própria, o agente age de forma proativa. Em produção, oferece inventário, segurança, acesso por função e gestão de custos; a cobrança é por Copilot Credits conforme a complexidade de cada resposta ou ação.

Árvore de decisão do Cloud Adoption Framework

Escolha a solução que atende com a menor responsabilidade operacional: SaaS (Copilot Studio) primeiro, PaaS (Foundry) quando o SaaS não dá controle suficiente e infraestrutura do Azure (VMs, AKS) só para runtime ou controle especial, nunca como padrão. Produtividade individual nos apps: Microsoft 365 Copilot. Função específica: Copilot por função (Security Copilot). Copilot dentro de produto (Power BI) vem com o acesso ao produto. Treinar ML com dados próprios: Fabric se a empresa já usa, Azure Machine Learning nos demais casos. Caso de trabalho individual vai para o Copilot; automação de negócio integrada a sistemas vai para soluções integradas.

Microsoft Foundry

Plataforma Azure gerenciada (PaaS) para criar aplicações e agentes sob medida; reúne agentes, modelos e ferramentas sob um agrupamento com acesso, rede e políticas unificados e consolidou o antigo Azure AI Studio. Controles: identidade do Microsoft Entra, RBAC, filtros de conteúdo, isolamento de rede e Azure Policy, mais rastreamento, avaliações e monitoramento. A nuvem global permite começar pequeno e escalar sem redesenhar. Agentes vão do agente de prompt declarativo (hospedado pelo Foundry) ao hospedado com código, e a orquestração agêntica usa o Foundry Agent Service.

Catálogo do Foundry e escolha de modelo

Catálogo com milhares de modelos de vários provedores, com comparação lado a lado usando dados próprios antes de qualquer compromisso de capacidade. Modelos vendidos diretamente pelo Azure são hospedados pela Microsoft sob os Product Terms, com suporte e SLAs, e passam por revisões de IA responsável que não dispensam a avaliação do cliente; o catálogo também traz modelos de parceiros e da comunidade. Versões são aposentadas: acompanhe datas e planeje migração. Abordagens: começar pelo mais simples (ferramenta pré-construída), depois modelos de base, RAG, fine-tuning ou orquestração agêntica.

Foundry Tools por cenário

Azure Language: sentimento, extração e categorização de texto. Azure Translator: tradução em mais de cem idiomas. Azure Speech: fala em texto e texto em fala. Azure Vision: analisa imagens e vídeos e lê texto em imagens (defeitos, etiquetas). Document Intelligence: extrai campos de faturas e formulários. Content Understanding: transforma documentos, imagens, vídeo e áudio em campos estruturados. Azure AI Search: indexa e recupera conhecimento interno (base do RAG). Azure AI Content Safety: modera texto e imagem por categoria e severidade.

Azure AI Search

Indexa conteúdo e usa busca semântica para recuperar o documento certo; oferece enriquecimento (chunking, vetorização), busca híbrida (palavra-chave e vetores) com ranking semântico, filtros e controles de acesso. A busca clássica é previsível e de baixa latência; a recuperação agêntica usa um LLM para dividir perguntas complexas em subconsultas executadas em paralelo, útil quando os documentos usam termos diferentes da pergunta.

Proteção de dados corporativos (EDP)

Prompts e respostas seguem os mesmos compromissos do Data Protection Addendum e dos Product Terms aplicados a e-mails e arquivos, com a Microsoft como processadora; vale para o Microsoft Copilot e o Copilot Chat. Prompts, respostas e dados acessados pelo Graph não treinam modelos de base, os dados ficam no limite de serviço do Microsoft 365 e são criptografados em trânsito. Filtros de segurança sinalizam conteúdo inadequado, sem revisão humana da Microsoft em cada resposta.

Implantações de modelo no Foundry

Padrão: capacidade compartilhada para tráfego variável, desenvolvimento e testes. Prioridade: baixa latência sem compromisso longo. Provisioned throughput: capacidade dedicada medida em PTUs para produção crítica com tráfego alto e latência consistente, reservada mesmo sem requisições (ter cota não garante capacidade na região). Batch: processamento em massa sem requisito de latência, com resultados assíncronos.

Solução integrada e plataformas de apoio

Uma solução integrada da Microsoft aplica políticas de governança de forma consistente entre Microsoft 365 e Azure, integra o Microsoft Purview para proteção de ponta a ponta e aplica acesso baseado em identidade. Microsoft Agent 365 é o plano de controle de agentes, apoiado em Entra, Defender, Purview e Intune. Microsoft Fabric unifica dados no OneLake. Outras soluções generativas: Copilot no Dynamics 365 (CRM, ERP, atendimento, agentes autônomos como Supplier Communications Agent), Copilot no Power Platform (apps e fluxos low-code) e GitHub Copilot (código).

Macetes para memorizar

💡 Copilot por app

Word escreve, Excel calcula, PowerPoint apresenta, Outlook responde, Teams lembra (decisões e itens de ação).

💡 Researcher x Analyst

Researcher pesquisa e cita; Analyst analisa e calcula.

💡 Pages, Notebooks, Search

Pages = página para coeditar; Notebooks = caderno do projeto; Search = achar.

💡 Comprar, estender, construir

Primeiro compra (comum), depois estende (integra), por último constrói (proprietário e regulado).

💡 Árvore do CAF

SaaS antes de PaaS, PaaS antes de infraestrutura: menor responsabilidade operacional vence.

💡 Sentido dos dados no Graph

Connectors = entrada (de fora para o Graph). Data Connect = saída (do Graph para o Azure).

💡 Declarativo x engine personalizada

Declarativo herda tudo do Microsoft 365; engine personalizada é casa própria: hospeda, protege e responde por tudo.

💡 Implantações do Foundry

Padrão varia, Provisionada garante, Batch espera.

Tabela de memorização: não confunda

Quando é...	...e quando é

Agente Researcher Análise competitiva, relatório com várias fontes da web e do trabalho, com citações, em pesquisa de várias etapas.	**Agente Analyst** Arquivos CSV e planilhas: estatísticas, tendências, outliers, comparação por região e trimestre com raciocínio iterativo.
Copilot Chat sem licença paga Chat seguro sem custo extra, grounding principalmente na web, arquivos por upload, e no Outlook o conteúdo do próprio usuário.	**Microsoft 365 Copilot licenciado** Grounding amplo via Graph, Work IQ e indexação semântica, raciocínio avançado prioritário, agentes avançados e Copilot Analytics; sempre dentro das permissões.
Agente declarativo Menor esforço de compliance e segurança, uso individual, orquestrador e modelos do Copilot, criado no Agent Builder ou pro-code.	**Agente de engine personalizada** Modelos próprios, orquestração complexa, colaboração em grupo no Teams, mensagens proativas ou uso fora do Microsoft 365, com hospedagem e compliance próprios.
Copilot Studio (SaaS) Agentes low-code ou em linguagem natural quando a personalização é suficiente; custo principal em créditos pelo uso dos agentes.	**Microsoft Foundry (PaaS)** Controle sobre modelos, lógica da aplicação e integração de dados que o Copilot Studio não oferece; custo pelo uso de modelos e ferramentas.
Copilot connectors Trazer dados de terceiros (Jira, CRM, ERP) para o Graph, para o Copilot e o Microsoft Search encontrarem.	**Microsoft Graph Data Connect** Levar dados do Graph em escala para armazenamentos do Azure, como análise de padrões de colaboração.
Conector sincronizado Indexar o conteúdo externo no Microsoft Graph para busca e Copilot.	**Conector federado** Dados que mudam a cada minuto ou sensíveis demais para indexar: consulta em tempo real via MCP.
Provisioned throughput Produção crítica com tráfego alto e previsível, em que variação de latência prejudica o cliente.	**Implantação batch** Milhões de documentos sem exigência de tempo de resposta, com resultados assíncronos.

Números que caem

Item	Valor
Idiomas do Azure Translator	**mais de cem**
Certificações citadas das Foundry Tools	**CSA STAR, FedRAMP Moderate e HIPAA BAA**
Unidade de capacidade provisionada no Foundry	**PTU**

Alertas de prova (pegadinhas)

⚠ **Alerta de prova:** Achar que a licença paga ou uma palavra no prompt libera conteúdo que o usuário não pode acessar.
Como não cair: Copilot Chat e Copilot licenciado respeitam as permissões existentes; nenhuma licença ou prompt concede acesso adicional.

⚠ **Alerta de prova:** Supor que o Copilot pronto lê sistemas externos por padrão.
Como não cair: Para consultar estoque ou ERP sem sair do Copilot é preciso estender: agentes e Copilot connectors.

⚠ **Alerta de prova:** Contratar projeto ou licença separada para criar o Microsoft Graph antes do Copilot.
Como não cair: O Graph já faz parte do Microsoft 365 e fornece ao Copilot o contexto que cada usuário pode ver.

⚠ **Alerta de prova:** Usar o Researcher para resumir três e-mails curtos e achar que a demora é defeito.
Como não cair: O Researcher demora por design (raciocínio profundo). Resumo rápido é com o Copilot padrão ou Copilot Chat.

⚠ **Alerta de prova:** Indicar o Copilot Studio como interface de chat dos usuários finais ou exigir agente para usar no celular.
Como não cair: Copilot Studio é para criar agentes. O chat está no app Copilot (web, desktop, celular), Teams, Outlook e Edge.

⚠ **Alerta de prova:** Dizer que o Foundry substitui o Copilot nos apps ou é cobrado por licença de usuário.
Como não cair: Foundry é plataforma para soluções sob medida, complementar ao Copilot, com custo pelo uso de modelos e ferramentas.

⚠ **Alerta de prova:** Tratar a infraestrutura do Azure (VMs com GPU) como abordagem padrão, ou o Foundry como opção de menor responsabilidade operacional.
Como não cair: Menor responsabilidade é SaaS; PaaS exige mais habilidades; infraestrutura só quando plataformas gerenciadas não suportam o runtime ou o controle exigido.

⚠ **Alerta de prova:** Escolher agente declarativo para colaboração em grupo, mensagens proativas ou troca do modelo do Copilot.
Como não cair: Declarativo é individual e usa orquestrador e modelos do Copilot; esses requisitos apontam para engine personalizada.

⚠ **Alerta de prova:** Afirmar que agentes do Copilot Studio usados por licenciados no Teams consomem Copilot Credits, ou que são gratuitos em qualquer canal.
Como não cair: Para licenciados, agentes no Copilot, Teams ou SharePoint com grounding do Graph não consomem créditos; sites públicos e outros canais consomem.

⚠ **Alerta de prova:** Achar que a revisão de IA responsável dos modelos vendidos pelo Azure dispensa a avaliação do cliente.
Como não cair: As revisões não garantem mitigação de todos os riscos; o cliente revisa a documentação e faz as próprias avaliações.

⚠ **Alerta de prova:** Usar Graph Data Connect para trazer Jira ao Copilot, ou Copilot connectors para levar dados ao Azure.
Como não cair: Connectors trazem dados externos para dentro; Graph Data Connect leva dados do Graph para o Azure. São sentidos opostos.

⚠ **Alerta de prova:** Liberar o serviço Face no nível gratuito sem registro ou usar detecção de emoções na catraca.
Como não cair: Face tem Acesso Limitado por critérios de elegibilidade e uso; inferência de emoção e gênero foi aposentada.

⚠ **Alerta de prova:** Escolher Translator, Language ou AI Search para extrair campos de notas fiscais.
Como não cair: Campos de faturas e formulários são Document Intelligence; mídias variadas (vídeo, voz, foto, PDF) em campos estruturados são Content Understanding.

⚠ **Alerta de prova:** Acreditar que prompts entram no treinamento de forma anônima ou que a troca de nome mudou os termos contratuais.
Como não cair: Prompts, respostas e dados do Graph não treinam modelos de base; a renomeação foi só de nome, com os mesmos compromissos de segurança e privacidade.

Como cai na prova

📝 **Como cai:** Cenário de um usuário com ou sem licença paga pedindo algo ao Copilot Chat.
Como resolver: Verifique a licença: sem licença, grounding é web mais o que o usuário envia (e, no Outlook, o próprio conteúdo); com licença, grounding amplo pelo Graph. Em qualquer caso, nada além das permissões existentes.

📝 **Como cai:** Executivo descreve uma tarefa e quer a experiência do Copilot mais adequada.
Como resolver: Classifique o tipo de trabalho (documento, dados, apresentação, e-mail, reunião, pesquisa, análise de dados, coedição) e associe ao app ou agente. Descarte ferramentas de construção (Copilot Studio, Foundry, Azure ML) quando a necessidade é imediata e individual.

📝 **Como cai:** Empresa quer criar ou integrar agentes e pergunta qual plataforma ou tipo de agente.
Como resolver: Pergunte: precisa só de tarefas comuns (comprar), integração com sistemas (estender) ou controle total (construir)? Depois: low-code dentro do Microsoft 365 (declarativo, Agent Builder, Copilot Studio) ou modelos e orquestração próprios (engine personalizada, Foundry).

📝 **Como cai:** Processo de negócio e a pergunta de qual Foundry Tool atende.
Como resolver: Identifique entrada e saída: texto e sentimento (Language), tradução (Translator), áudio (Speech), imagem e texto em imagem (Vision), campos de formulário (Document Intelligence), mídia variada em campos (Content Understanding), recuperar conhecimento (AI Search), moderar (Content Safety).

📝 **Como cai:** Pergunta de segurança, privacidade ou contrato sobre o Copilot.
Como resolver: Responda com os fatos fixos: permissões respeitadas, dados no limite de serviço do Microsoft 365, criptografia em trânsito, sem treino de modelos de base, proteção pelo DPA e Product Terms, admin controla grounding na web pelos Copilot controls.

Gatilhos do enunciado

Se o enunciado trouxer...	Pense em
planilhas e CSV, tendências, outliers, por região e trimestre	**Agente Analyst**
análise competitiva com fontes da web e internas, citadas	**Agente Researcher**
transformar resposta do Copilot em conteúdo que a equipe edita junto	**Copilot Pages**
reunir arquivos, notas, links e conversas de um projeto	**Copilot Notebooks**
agentes prontos da Microsoft e de parceiros sem desenvolver	**Agent Store do Microsoft Copilot**
agente sem código no Microsoft 365 Copilot com catálogos do SharePoint	**Agent Builder (agente declarativo)**
status atualizado de sistema externo na hora da pergunta, num agente declarativo	**Ação (API plugin)**
dados que mudam a cada minuto e não podem ser indexados	**Conector federado via MCP**

aprovação automática com pessoa revisando valores acima de um limite	**Agent flow do Copilot Studio com revisão humana**
aviso de sinistro, consultar apólice, pedir documentos, atualizar sistema em etapas	**Orquestração agêntica com Foundry Agent Service**
dados processados no próprio ambiente da empresa	**Foundry Tools em contêineres locais**
conteúdo sexual, violento, de ódio ou automutilação com severidade	**Azure AI Content Safety**
termos exatos de artigos de lei e perguntas conceituais com outras palavras	**Busca híbrida com ranking semântico no Azure AI Search**
plano de controle com identidade, visibilidade e políticas para agentes	**Microsoft Agent 365**

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

- O Copilot respeita sempre as permissões existentes; licença não amplia acesso.
- Prompts, respostas e dados do Graph não treinam modelos de base.
- Copilot Chat sem licença é fundamentado principalmente na web; grounding amplo exige a licença paga.
- No CAF, escolha a opção com menor responsabilidade operacional que atende aos requisitos; infraestrutura nunca é o padrão.

Essenciais para a prova (checklist)

☐ Sei o que o Copilot Chat faz sem licença e o que a licença paga acrescenta.

☐ Sei mapear processos de documentos, dados, apresentações, e-mail e reuniões ao app certo do Copilot.

☐ Sei quando usar Researcher, Analyst, Pages, Notebooks e Copilot Search.

☐ Sei explicar o Microsoft Graph, o Work IQ e a diferença entre Copilot connectors sincronizados, federados e Graph Data Connect.

☐ Sei aplicar comprar, estender ou construir e escolher entre agente declarativo e engine personalizada.

☐ Sei o que o Copilot Studio oferece (criação, canais, agent flows, governança, Copilot Credits).

☐ Sei aplicar a árvore de decisão do CAF entre Copilot, Copilot Studio, Foundry, Fabric, Azure ML e infraestrutura.

☐ Sei associar cada Foundry Tool ao seu cenário de negócio.

☐ Sei os controles de segurança do Foundry e os tipos de implantação (padrão, prioridade, provisionada, batch).

☐ Sei os compromissos de proteção de dados corporativos do Copilot.

Questões de revisão do capítulo

R1.1 · Identificar benefícios, recursos e oportunidades dos aplicativos e serviços de IA da Microsoft

A diretoria de atendimento da Telecom Sinal Verde quer usar IA para rotear automaticamente os chamados de clientes para a equipe certa, integrando o sistema de tickets. Segundo o Cloud Adoption Framework, como esse caso de uso deve ser classificado?

A. Trabalho individual, dentro das ferramentas atuais
B. Infraestrutura de rede, sem relação com a IA da empresa
C. Automação de negócio, com integração a sistemas
D. Pesquisa acadêmica, sem impacto na operação diária

R1.2 · Identificar benefícios, recursos e oportunidades dos aplicativos e serviços de IA da Microsoft

Na Companhia de Celulose Fibra Verde, os funcionários fazem perguntas complexas ao assistente de RH, como regras de férias para trabalhadores remotos contratados recentemente, enquanto os documentos usam termos diferentes. Qual capacidade do Azure AI Search atende melhor a esse cenário?

A. Exportação do índice para planilhas que os usuários filtram por conta própria
B. Recuperação agêntica, em que um LLM divide a pergunta em subconsultas executadas em paralelo
C. Tradução da pergunta para o inglês antes de consultar o índice da empresa
D. Busca por palavra-chave exata em um único índice, sem planejamento de consulta

R1.3 · Identificar benefícios, recursos e oportunidades dos aplicativos e serviços de IA da Microsoft

A rede de farmácias Bem Viver vai classificar perguntas curtas de clientes em cinco categorias. A equipe técnica propôs usar o maior modelo disponível no Microsoft Foundry. Qual recomendação está alinhada à orientação da Microsoft?

A. Usar um modelo menor, adequado à complexidade da tarefa, com custo e latência menores
B. Treinar um modelo grande do zero para garantir a melhor classificação possível
C. Usar dois modelos grandes em paralelo e escolher a resposta de maior confiança em cada caso
D. Usar o maior modelo disponível, pois ele sempre entrega o melhor resultado

R1.4 · Identificar benefícios, recursos e oportunidades dos aplicativos e serviços de IA da Microsoft

O diretor jurídico do Grupo Hospitalar Bom Pastor percebeu que a documentação passou a citar Microsoft Copilot em vez de Microsoft 365 Copilot e teme que os compromissos contratuais tenham mudado. Qual esclarecimento está correto?

A. O nome novo indica que o Copilot passou a usar só dados da internet
B. Só o nome mudou: agora é Microsoft Copilot, sem mudanças em segurança, compliance e privacidade
C. A mudança encerra a proteção de dados corporativos de quem não migrar a licença
D. A mudança indica um novo produto, com termos de privacidade a negociar

Respostas comentadas

R1.1. C) Automação de negócio, com integração a sistemas

Por quê: Classificar o caso de uso (trabalho individual ou automação de negócio) orienta a escolha da tecnologia: Microsoft 365 Copilot para o primeiro e soluções integradas para o segundo.

✗ A) Incorreta. Trabalho individual é melhorar o trabalho de pessoas nas ferramentas atuais, como redação ou preparação de reuniões.
✗ B) Incorreta. Trata-se de um caso de uso de IA, não de infraestrutura de rede.
✗ D) Incorreta. O caso tem impacto direto na operação.

> **Dica de prova:** Mudar como a empresa opera, com integração: automação de negócio.

R1.2. B) Recuperação agêntica, em que um LLM divide a pergunta em subconsultas executadas em paralelo

Por quê: O Azure AI Search oferece a busca clássica, previsível e de baixa latência, e a recuperação agêntica, com planejamento de consulta, várias fontes e respostas otimizadas para agentes.

✗ A) Incorreta. Planilhas não resolvem perguntas complexas em linguagem natural.
✗ C) Incorreta. O problema não é idioma, e sim intenção e vocabulário diferente.
✗ D) Incorreta. Busca por palavra-chave falha quando a pergunta não usa os termos dos documentos.

> **Dica de prova:** Pergunta complexa com vocabulário diferente: recuperação agêntica.

R1.3. A) Usar um modelo menor, adequado à complexidade da tarefa, com custo e latência menores

Por quê: Escolher o modelo do tamanho certo entrega a capacidade necessária sem gastar demais, mantém governança e facilita escalar.

✗ B) Incorreta. Treinar do zero é caro e desnecessário para uma classificação simples.
✗ C) Incorreta. Dois modelos grandes dobram custo e complexidade sem necessidade.
✗ D) Incorreta. O maior modelo não é necessariamente o mais adequado e eleva custo e latência.

> **Dica de prova:** Tarefa simples: modelo menor do tamanho certo.

R1.4. B) Só o nome mudou: agora é Microsoft Copilot, sem mudanças em segurança, compliance e privacidade

Por quê: Durante a transição, algumas experiências, licenças e documentos ainda citam Microsoft 365 Copilot e Microsoft 365 Copilot Chat; os nomes se referem aos mesmos serviços.

✗ A) Incorreta. A renomeação não altera o grounding nos dados da organização.

✗ C) Incorreta. A proteção de dados corporativos continua valendo.

✗ D) Incorreta. Trata-se de renomeação, não de um novo produto com novos termos.

> **Dica de prova:** Microsoft 365 Copilot agora se chama Microsoft Copilot, com as mesmas proteções.

02

CAPÍTULO 2 · 35-40% DA PROVA

Identificar o valor de negócio das soluções de IA generativa

Aqui eu preciso separar com segurança IA generativa de IA tradicional (machine learning), saber quando usar modelo pré-construído, personalizado, RAG ou fine-tuning e entender de onde vem o custo (tokens) e o ROI. O enunciado quase sempre traz um executivo com um problema de negócio e pede a técnica ou o conceito certo: grounding, few-shot, mensagem de sistema, temperatura, destilação, clustering, inferência. Também cai muito desafio da IA generativa (fabricação, confiabilidade, viés) com a mitigação correta, e segurança de IA pelo Cloud Adoption Framework (RBAC, DLP, Defender for Cloud, red teaming, endpoints privados). Onde mais erro: confundir RAG (traz dado atual) com fine-tuning (ajusta tom, formato e terminologia) e cair na alternativa de 'aumentar a temperatura', que quase nunca é a resposta. O papel do líder é sempre de negócio: definir o problema, garantir dados de qualidade e medir valor, nunca escolher algoritmo ou hiperparâmetro.

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

Tópicos-chave

IA generativa x IA tradicional (não generativa)

A IA tradicional analisa dados, classifica e faz previsões com base em padrões; a generativa cria conteúdo novo (texto, imagem, áudio, código) a partir de prompts. A generativa é não determinística: pode variar para a mesma entrada. Pelo CAF, quando a mesma entrada precisa levar ao mesmo resultado (score de crédito, comissão por regra fixa), a escolha é IA não generativa (determinística); quando o fluxo varia e alguma variação é aceitável (e-mails em texto livre, pareceres), é generativa.

IA agêntica

IA que usa modelos com instruções e ferramentas para buscar um objetivo e executar tarefas dentro de permissões definidas. Os tipos se combinam: um agente pode usar IA generativa para interpretar o pedido e IA tradicional para prever algo antes de executar uma ação autorizada. Agente sem permissões definidas contraria a recomendação: agentes agem dentro de permissões e controles definidos, e decisões como concessão de crédito pedem supervisão.

Data science, machine learning e deep learning

Data science gera insight com estatística e conhecimento de domínio; machine learning aprende padrões nos dados para prever ou classificar. Deep learning é um subconjunto do ML baseado em redes neurais em camadas, costuma exigir grandes volumes de dados e muito poder de computação e se destaca com dados não estruturados (imagens, texto, áudio). Não são sinônimos.

Tipos de modelos generativos

LLM gera e entende texto; SLM (modelo de linguagem pequeno) atende tarefas restritas com menor custo e mais velocidade; modelo de código gera e depura programas; modelo de difusão gera imagens a partir de texto; modelo multimodal trata várias entradas (texto, imagem, fala) na mesma interação; há ainda modelos específicos de domínio. A arquitetura transformer dá compreensão contextual de documentos inteiros, processamento paralelo e versatilidade entre texto, imagem e código.

Pré-construído x personalizado (comprar x construir)

Modelo pré-construído serve a casos gerais, com implantação rápida e pouca configuração; personalizado atende necessidades especializadas e dá controle dos dados, com custo e complexidade maiores. Comprar é o caminho quando a capacidade é padrão e há pouca expertise interna em ML. Construir faz sentido para propriedade intelectual única, domínio muito especializado, diferenciação de longo prazo ou dados proprietários com compliance complexo.

Fine-tuning

Adapta um modelo pré-treinado com treinamento adicional em dados da tarefa, ajustando os pesos; não treina do zero. Ajuda a manter tom, formato ou esquema consistentes, aprender terminologia de nicho e reduzir exemplos no prompt (menos tokens e latência). O benefício não é garantido: compare o modelo ajustado com a linha de base em dados de avaliação separados. Não traz informação atual e exige dados curados, governança e gestão de drift.

Destilação

Usa exemplos gerados por um modelo maior (professor) para ajustar um modelo menor, buscando qualidade parecida na tarefa com menor custo e latência. É a resposta quando o cenário fala em modelo grande caro e lento numa tarefa específica e a meta é um modelo menor com qualidade semelhante.

Fatores de custo e ROI

No Copilot os tokens existem nos bastidores, mas o custo é de assinatura previsível, sem cobrança por token e com infraestrutura gerida pela Microsoft. Em soluções sob medida no Foundry, a cobrança é por uso (tokens processados) e os recursos de nuvem são geridos pelo cliente. Fatores de custo de solução sob medida: treinamento, inferência, preparo de dados, infraestrutura e talentos; o RAG aproveita documentos como estão e reduz o preparo de dados. Prompts e respostas maiores (histórico completo, respostas longas) consomem mais tokens.

Desafios da IA generativa

Fabricação: o modelo prevê padrões de texto em vez de verificar fatos, por isso cita com segurança algo que não existe; mitigação é checagem cruzada com dados confiáveis e RAG. Confiabilidade: a natureza probabilística gera respostas diferentes para perguntas parecidas; mitigação com fine-tuning, engenharia de prompt e guardrails. Viés: dados históricos enviesados geram recomendações injustas. Outros desafios citados: privacidade de dados, consumo de energia e falta de explicabilidade. Resolver desafios com transparência e precisão pode virar inovação e novo mercado.

Engenharia de prompt para líderes

Os quatro elementos de um bom prompt são objetivo, contexto, fontes e expectativas sobre a resposta; diga o que quer que o Copilot faça, em vez de listar proibições. É responsabilidade compartilhada: usuários precisam de alfabetização em prompts e desenvolvedores aplicam templates, few-shot e geração automatizada de prompts. Prompts ruins geram resultados irrelevantes, informações imprecisas e tempo perdido; investir em alfabetização traz fluxos mais rápidos, funcionários confiantes e menor risco de saídas enganosas, mas validar as respostas continua necessário.

Técnicas de prompt (Microsoft)

Papel e limites do assistente ficam na mensagem de sistema. Few-shot inclui exemplos de entrada e saída para guiar formato sem retreinar. Sintaxe clara e divisão da tarefa em etapas melhoram tarefas compostas; a ordem das informações altera a saída (viés de recência). Dar uma saída alternativa ('não encontrado'), pedir citações e usar dados de grounding reduzem respostas inventadas. Templates padronizados reduzem variação entre canais.

Temperatura

Parâmetro de 0 a 2 que controla a aleatoriedade: valores baixos (como 0,2) dão respostas focadas, adequadas a documentos legais e classificação; valores altos (como 0,7) dão respostas divergentes, para ficção e ideias. A recomendação é alterar temperatura ou top_p, um de cada vez. Temperatura não controla escopo, segurança, auditoria, atualização de dados nem proteção contra injeção de prompt.

Grounding

Conecta as respostas da IA a fontes de dados autorizadas, indo além do treinamento do modelo, e reduz informações convincentes porém incorretas. É prioritário em setores regulados, cenários de compliance e aplicações voltadas a clientes. Para funcionar, o líder garante acesso a fontes autorizadas, integração com sistemas corporativos e políticas de governança sobre a precisão dos dados. Sem grounding: respostas incorretas ou desatualizadas, falhas de compliance e perda de confiança que reduz a adoção.

RAG (geração aumentada por recuperação)

Técnica de grounding em três etapas: recuperar dados, aumentar o prompt e gerar a resposta; funciona como um especialista que consulta o manual atualizado antes de responder. Não altera pesos nem retreina o modelo. Envia só os trechos relevantes porque o modelo aceita tokens limitados; a recuperação precisa respeitar as permissões de cada usuário. Exibir as fontes na resposta aumenta a confiança. No Azure AI Search, vetores não são obrigatórios: servem para busca por similaridade, e o enriquecimento faz chunking e vetorização quando necessário.

Impacto e prontidão dos dados

Tipo (estruturado ou não estruturado), qualidade e representatividade determinam o desempenho da IA; dados que não representam um grupo geram resultados enviesados ou irrelevantes para ele, e dados representativos sustentam a equidade. Prontidão de dados: inventário e classificação, acessibilidade, metadados e rotulagem, governança e ciclos de feedback. O Microsoft 365 Copilot usa sobretudo conteúdo não estruturado (e-mails, documentos), que precisa estar organizado e acessível. IA confiável se apoia em dois pilares: dados de alta qualidade e segurança robusta.

Ciclo de vida de machine learning

Etapas: definir a tarefa, coletar e preparar dados, treinar e validar, implantar e monitorar e gerenciar com MLOps; o ciclo é iterativo, e drift ou mudança de comportamento exige retreinar e reavaliar. Inferência é quando o modelo integrado aos sistemas gera previsões. O líder define a pergunta de negócio e garante dados precisos, completos e relevantes; não escreve código nem escolhe hiperparâmetros. Com a IA generativa, o foco sai de construir modelos (MLOps) para usar modelos pré-treinados (GenAIOps ou LLMOps), que continuam exigindo operação e monitoramento.

Quando o machine learning agrega valor

ML é mais adequado que LLM para previsão numérica, classificação e clustering: prever churn e demanda, detectar fraude e transações fora do padrão, segmentar clientes, personalizar sugestões. Antes de construir, pergunte se há dados e habilidades e se o benefício justifica o custo frente a serviços prontos; os requisitos são dados suficientes e representativos, especialistas, infraestrutura e análise de ROI. Categorias de valor: descritiva (busca, classificação, resumo), preditiva (previsão, anomalias, score) e prescritiva (recomendação, otimização).

Segurança de sistemas de IA

A base é Zero Trust com governança de dados robusta, criptografia em repouso e em trânsito e autenticação. Pelo CAF: RBAC limitando acesso por carga e grupo, Purview DLP contra exposição de dados sensíveis, Defender for Cloud para descobrir recursos de IA e manter inventário, red teaming e testes adversariais antes do lançamento, planos de resposta a incidentes de IA testados e redes virtuais com endpoints privados. Injeção indireta (instruções escondidas em documento ou e-mail) é detectada pelo Prompt Shields. Cada recurso das Foundry Tools tem duas chaves para rodar uma sem interromper o serviço, e o conteúdo gerado deve ser rotulado conforme a sensibilidade da fonte (resumo de relatório confidencial continua confidencial).

Macetes para memorizar

RAG x fine-tuning

RAG = Recente (dado que muda). Fine-tuning = Forma (tom, formato, termos).

Etapas do RAG

R-A-G na ordem do nome: Recuperar, Aumentar o prompt, Gerar.

Elementos do prompt

OCFE: Objetivo, Contexto, Fontes, Expectativas. E diga o que fazer, não o que não fazer.

Temperatura

Contrato frio (0,2), ficção quente (0,7). Ajuste temperatura ou top_p, um de cada vez.

Difusão x multimodal

Difusão desenha (imagem sai). Multimodal enxerga e ouve (imagem e fala entram).

Categorias de valor

Descreve (busca, resumo), Prevê (previsão, anomalia, score), Prescreve (recomenda, otimiza).

Desafios da IA generativa

FCV + PEE: Fabricação, Confiabilidade, Viés; depois Privacidade, Energia, Explicabilidade (falta).

Fatores de custo sob medida

TIPIT: Treinamento, Inferência, Preparo de dados, Infraestrutura, Talentos.

Tabela de memorização: não confunda

Quando é...	...e quando é
RAG Informação que muda (preços, estoque, políticas, normas vigentes), precisão factual com citação de fontes e menor custo de preparo de dados; exige conteúdo curado.	**Fine-tuning** Padrões estáveis de tarefa: tom, formato ou esquema consistentes, terminologia de nicho e formato rígido quando few-shot não basta; exige dados curados, avaliação contra a linha de base e gestão de drift.
IA generativa (não determinística) Fluxo não fixo, entrada em texto livre e variação aceitável na saída: responder e-mails, redigir pareceres, resumir documentos, criar conteúdo de campanha.	**IA não generativa (determinística)** A mesma entrada deve gerar o mesmo resultado a partir de dados estruturados: score de crédito, cálculo de comissão por regra fixa, previsão de ruptura, detecção de anomalias.
Few-shot no prompt Mostrar alguns exemplos de entrada e saída para o modelo seguir um formato próprio sem retreinar; mas acrescenta tokens em cada chamada.	**Fine-tuning** Quando os exemplos no prompt já não dão a precisão exigida ou se quer reduzir exemplos (tokens e latência), aceitando o custo de treinar e avaliar.
Modelo de difusão Gerar imagens novas a partir de descrições em texto (imagem é a saída).	**Modelo multimodal** Receber foto, fala e texto na mesma interação e produzir uma análise (imagem é entrada).

Custo do Microsoft 365 Copilot Assinatura previsível por usuário, tokens embutidos, infraestrutura gerida pela Microsoft e ROI rápido.	**Custo de solução sob medida no Foundry** Cobrança por uso (tokens processados na inferência), recursos geridos pelo cliente, mais flexibilidade e controle com gestão de custos mais cuidadosa.
Modelo menor (SLM ou destilado) Tarefa restrita como classificar tickets ou avaliações, com baixo custo e boa velocidade.	**Maior LLM disponível** Tarefas amplas e complexas que justificam custo e latência maiores; nunca escolhido só por ser o maior.
Comprar (solução pronta) Capacidade padrão, pouca expertise interna em ML, sem diferencial competitivo e necessidade de valor rápido.	**Construir (modelo próprio)** Propriedade intelectual única, domínio muito especializado, diferenciação de longo prazo ou dados proprietários com compliance complexo, havendo dados e habilidades.

Números que caem

Item	Valor
Faixa de temperatura	**0 a 2**
Temperatura baixa citada para documentos legais	**0,2**
Temperatura mais alta citada (respostas mais variadas)	**0,7**
Folga recomendada no cronograma de projetos de IA (CAF)	**20% a 30%**
Prazo de retorno dos Copilots da Microsoft	**dias ou semanas**
Prazo de cargas de IA sob medida no Azure até a produção	**várias semanas a meses**
KPIs por caso de uso para provar valor	**dois ou três, com linha de base e testes A/B/N**
Chaves de acesso por recurso das Foundry Tools	**duas (para rodar e regenerar sem interromper)**

Alertas de prova (pegadinhas)

⚠ **Alerta de prova:** Aumentar a temperatura para o modelo ser mais criativo, achar respostas, cobrir mais cláusulas ou dificultar ataques.
Como não cair: Temperatura alta só aumenta a variação. Em contratos, classificação, relatórios padronizados e segurança, a resposta nunca é subir a temperatura; em documentos legais, a indicação é temperatura baixa.

⚠ **Alerta de prova:** Fazer fine-tuning diário ou semanal para manter preços e disponibilidade atualizados.
Como não cair: Fine-tuning ajusta comportamento e formato, não traz informação atual. Dado que muda todo dia vem de recuperação (RAG).

⚠ **Alerta de prova:** Usar RAG para garantir tom e estrutura de campos sempre iguais.
Como não cair: RAG traz informação para a resposta, mas não garante formato. Tom e formato consistentes com bons dados históricos apontam para fine-tuning.

⚠ **Alerta de prova:** Acreditar que o RAG altera os pesos do modelo ou retreina a cada documento indexado.
Como não cair: RAG traz contexto no momento da pergunta (recuperar, aumentar, gerar). Quem altera pesos é o fine-tuning; as definições trocadas são distrator clássico.

⚠ **Alerta de prova:** Considerar o ganho do fine-tuning garantido ou avaliá-lo com os mesmos exemplos do treino.
Como não cair: A documentação diz que o benefício não é garantido; compare com a linha de base em dados de avaliação separados. Avaliar com dados de treino superestima o desempenho.

⚠ **Alerta de prova:** Escolher modelo personalizado por 'menor custo e menor complexidade' ou treinar do zero para ter 'controle total com pouco esforço'.
Como não cair: Personalizado tem custo e complexidade maiores; treinar do zero exige muito esforço, dados e especialistas. Pressa e equipe pequena apontam para pré-construído.

⚠ **Alerta de prova:** Trocar pelo maior modelo disponível para reduzir chamadas ou responder melhor.
Como não cair: Modelos maiores elevam custo e latência; a recomendação é dimensionar o modelo à tarefa e otimizar tarefas repetitivas (reutilizar respostas frequentes, agrupar solicitações).

⚠ **Alerta de prova:** Achar que a fatura do Microsoft 365 Copilot varia por token ou tem franquia de tokens por usuário.
Como não cair: O Copilot usa tokens nos bastidores, mas o custo é assinatura previsível, sem cobrança por token. Também é falso que o Copilot não use LLMs.

⚠ **Alerta de prova:** Repetir 'não invente informações' várias vezes na mensagem de sistema ou pedir que o modelo estime a resposta provável.
Como não cair: O ajuste indicado é pedir citações das fontes e dar uma saída alternativa ('não encontrado'). Estimar sem base incentiva justamente a invenção.

⚠ **Alerta de prova:** Explicar a fabricação como intenção de enganar, falta de licença premium ou cópia de erro de fonte oficial.
Como não cair: Fabricação acontece porque o modelo prevê padrões de texto em vez de verificar fatos. Mitiga-se com checagem cruzada e RAG, não abandonando o projeto.

⚠ **Alerta de prova:** Usar LLM ou modelo de difusão para prever demanda, churn ou fraude.
Como não cair: Previsão, classificação, detecção de padrões incomuns e segmentação são tarefas de machine learning (preditivo ou clustering), não de IA generativa.

⚠ **Alerta de prova:** Dizer que a IA garante neutralidade ou que dados representativos dispensam revisão humana.
Como não cair: A IA pode herdar vieses dos dados históricos; decisões que afetam pessoas continuam com revisão humana. Dados representativos ajudam a evitar viés, mas não substituem supervisão.

⚠ **Alerta de prova:** Juntar RH e finanças no mesmo índice sem filtro, para respostas mais completas.
Como não cair: Requisito essencial do RAG corporativo: a recuperação respeita as permissões e devolve a cada usuário só o conteúdo autorizado.

⚠ **Alerta de prova:** Dispensar a validação humana porque o prompt foi ótimo em poucos casos.
Como não cair: Um bom prompt pode não generalizar; técnicas de prompt aumentam a precisão, mas validar saídas e monitorar em produção continua essencial.

Como cai na prova

📝 **Como cai:** Executivo descreve um assistente que erra ou inventa e pergunta qual técnica resolve.
Como resolver: Identifique a causa: falta de dado atual ou de fonte (grounding, RAG, citações, saída alternativa), formato inconsistente (template, few-shot, fine-tuning) ou tarefa composta (dividir em etapas). Elimine primeiro qualquer alternativa que aumente a temperatura ou remova dados de grounding.

📝 **Como cai:** Lista de iniciativas e a pergunta de qual é generativa, qual é ML ou que tipo de modelo usar.
Como resolver: Pergunte se a saída é conteúdo novo (generativa) ou previsão, classificação, agrupamento (ML). Depois case o tipo de modelo: texto (LLM/SLM), imagem como saída (difusão), várias entradas (multimodal), código (modelo de código).

📝 **Como cai:** CFO pergunta como funciona o custo ou por que a fatura subiu.
Como resolver: Copilot = assinatura previsível sem cobrança por token. Foundry = tokens por uso; fatura sobe com histórico completo e respostas longas. Para reduzir: RAG em vez de treinar, modelo dimensionado, reutilizar respostas frequentes.

📝 **Como cai:** Pergunta sobre o papel do líder numa etapa do ciclo de ML ou antes de aprovar um projeto.
Como resolver: Escolha a ação de negócio: definir a pergunta e a decisão apoiada, garantir dados precisos, completos e relevantes, avaliar ROI frente a serviços prontos. Descarte escrever código, escolher algoritmo ou aprovar orçamento sem participar.

📝 **Como cai:** CISO quer proteger dados, comunicações ou detectar ameaças em soluções de IA.
Como resolver: Mapeie o risco ao controle: acesso amplo (RBAC), exposição de dados sensíveis (Purview DLP), recursos desconhecidos (Defender for Cloud), interceptação (redes virtuais e endpoints privados), vulnerabilidades antes do lançamento (red teaming), injeção (Prompt Shields), incidentes (plano de resposta testado).

📝 **Como cai:** Questões de múltipla resposta (Escolha DUAS/TRÊS) sobre RAG, grounding, fine-tuning ou deep learning.
Como resolver: Marque as afirmações que descrevem a função correta e descarte as absolutas ('dispensa avaliação', 'altera pesos', 'treina do zero', 'usa só o treinamento original', 'dispensa governança').

Gatilhos do enunciado

Se o enunciado trouxer...	Pense em
dado que muda todo dia, versão vigente, tabela de preço atual	**RAG / grounding em fontes autorizadas**
mesmo tom e mesma estrutura, terminologia própria, formato rígido de submissão	**Fine-tuning com exemplos curados**
formato próprio de categorias sem retreinar o modelo	**Few-shot no prompt**
modelo grande caro e lento, quer menor com qualidade parecida	**Destilação (professor gera exemplos para o menor)**
papel bem definido e limites do assistente	**Mensagem de sistema**
respostas diferentes conforme a posição das instruções	**Ordem das informações no prompt (viés de recência)**
inventa resposta quando a informação não está no texto	**Saída alternativa ('não encontrado') e pedir citações**
mesma entrada deve dar sempre o mesmo resultado	**IA não generativa (determinística)**
agrupar clientes com hábitos parecidos	**Clustering (machine learning)**
roteiro de entregas, otimizar frota, recomendar a melhor ação	**IA prescritiva**

modelo já treinado integrado ao app gera previsão a cada transação	**Inferência**
modelo perdeu precisão porque hábitos mudaram	**Retreinar e reavaliar; MLOps detecta drift**
instruções escondidas num e-mail ou documento lido pelo assistente	**Ataque por documento (injeção indireta) detectado pelo Prompt Shields**
não sabe quantos recursos de IA existem nas assinaturas do Azure	**Microsoft Defender for Cloud para descobrir e inventariar**

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

- RAG traz dado atual sem mexer nos pesos; fine-tuning ajusta pesos para tom, formato e terminologia e não atualiza informação.
- Copilot tem custo de assinatura previsível; Foundry cobra por tokens processados.
- Fabricação vem de o modelo prever padrões em vez de verificar fatos; mitigação é checagem cruzada, RAG e citações.
- O líder define o problema e garante dados de qualidade; escolhas técnicas ficam com especialistas.

Essenciais para a prova (checklist)

☐ Sei diferenciar IA generativa, IA tradicional e IA agêntica e quando o CAF pede IA determinística.
☐ Sei escolher entre LLM, SLM, modelo de código, difusão e multimodal pelo tipo de entrada e saída.
☐ Sei quando usar pré-construído, personalizado, few-shot, RAG, fine-tuning ou destilação.
☐ Sei listar os fatores de custo de solução sob medida e explicar por que a fatura do Foundry sobe.
☐ Sei relacionar fabricação, confiabilidade e viés às suas causas e mitigações.
☐ Sei os quatro elementos de um bom prompt e as técnicas de mensagem de sistema, saída alternativa, citações e ordem das informações.
☐ Sei as três etapas do RAG e por que a recuperação envia só trechos relevantes e respeita permissões.
☐ Sei os fatores de dados (tipo, qualidade, representatividade) e os elementos de prontidão de dados.
☐ Sei as etapas do ciclo de vida de ML e o papel do líder em cada uma.
☐ Sei associar cada risco de segurança de IA ao controle indicado pelo CAF.

Questões de revisão do capítulo

R2.1 · Identificar o valor de negócio das soluções de IA generativa

Na etapa de preparação de dados de um projeto de machine learning da Rede de Óticas Bom Foco, a equipe de dados está limpando e organizando as bases. Qual é o papel esperado do líder de negócio nessa etapa?

A. Garantir que os dados sejam precisos, completos e relevantes para o problema
B. Limpar pessoalmente cada tabela de dados antes do treinamento
C. Dispensar a etapa de preparo para acelerar o projeto
D. Escolher a arquitetura da rede neural e os hiperparâmetros do modelo em produção

R2.2 · Identificar o valor de negócio das soluções de IA generativa

A varejista Casa Bela usa um modelo grande para classificar avaliações de clientes, mas o custo e a latência subiram com o volume. A equipe quer um modelo menor com qualidade parecida nessa tarefa. Qual técnica atende a esse objetivo?

A. RAG, recuperando trechos do catálogo de produtos antes de cada classificação
B. Aumento da temperatura, para que o modelo menor produza respostas mais variadas
C. Destilação, com exemplos gerados pelo modelo maior para treinar um menor
D. Few-shot no prompt, incluindo três exemplos de resposta em cada chamada ao modelo

R2.3 · Identificar o valor de negócio das soluções de IA generativa

O gerente da central de reclamações da Companhia de Saneamento Águas Claras quer que o modelo classifique as mensagens num formato próprio de categorias, sem retreinar o modelo. Qual técnica atende a esse objetivo?

A. Pré-treinamento de um modelo novo do zero com as reclamações recebidas
B. Temperatura alta, para que o modelo crie novas categorias conforme cada reclamação que receber
C. Fine-tuning completo do modelo com o histórico inteiro de reclamações da companhia
D. Few-shot, incluindo no prompt alguns exemplos de entrada e da classificação esperada

R2.4 · Identificar o valor de negócio das soluções de IA generativa

No Hospital Infantil Sorriso, todos os membros da equipe de dados receberam acesso de administrador aos recursos e dados de IA para não atrasar os projetos. Segundo o Cloud Adoption Framework, qual prática deveria ser adotada?

A. Manter o acesso de administrador para a equipe inteira, para não atrasar os projetos
B. Publicar as chaves de acesso no wiki interno para que qualquer pessoa da equipe consiga testar os modelos
C. Compartilhar uma conta única de serviço entre as equipes para simplificar a gestão
D. Aplicar controle de acesso baseado em função, limitando o acesso aos dados por carga e grupo

Respostas comentadas

R2.1. A) Garantir que os dados sejam precisos, completos e relevantes para o problema

Por quê: Em cada etapa do ciclo de ML, o líder tem um papel: definir o problema, garantir fontes e qualidade dos dados, apoiar a experimentação, integrar ao processo e acompanhar o desempenho.

✗ B) Incorreta. A limpeza é trabalho das equipes de dados.

✗ C) Incorreta. Dados mal preparados comprometem o modelo.

✗ D) Incorreta. Escolhas técnicas de modelo cabem aos especialistas.

> **Dica de prova:** Preparo de dados: o líder garante dados precisos, completos e relevantes.

R2.2. C) Destilação, com exemplos gerados pelo modelo maior para treinar um menor

Por quê: A destilação usa dados gerados por um modelo maior (professor) para ajustar um modelo menor, buscando qualidade semelhante com menor custo e latência.

✗ A) Incorreta. RAG traz dados atuais para a resposta, mas não transfere a qualidade do modelo maior para um menor.

✗ B) Incorreta. Temperatura alta aumenta a variação das respostas, o oposto do necessário para classificação consistente.

✗ D) Incorreta. Few-shot melhora o formato da resposta, mas não reduz o tamanho do modelo usado; ainda acrescenta tokens a cada chamada.

> **Dica de prova:** Modelo menor imitando um maior: destilação.

R2.3. D) Few-shot, incluindo no prompt alguns exemplos de entrada e da classificação esperada

Por quê: Few-shot é uma técnica avançada de prompt em que exemplos guiam o raciocínio e o formato da resposta; ajuda o modelo a entender o comportamento esperado sem retreinar.

✗ A) Incorreta. Treinar do zero é caro e desnecessário para orientar o formato.

✗ B) Incorreta. Temperatura alta aumenta a variação, o oposto de um formato fixo de categorias.

✗ C) Incorreta. Fine-tuning retreina o modelo, o que o gestor quer evitar.

> **Dica de prova:** Exemplos no prompt para guiar o formato: few-shot.

R2.4. D) Aplicar controle de acesso baseado em função, limitando o acesso aos dados por carga e grupo

Por quê: Definir limites claros de dados garante que cargas de IA acessem só o que é adequado ao público e ao caso de uso.

✗ A) Incorreta. Acesso amplo aumenta o risco de vazamento e de ameaças internas.

✗ B) Incorreta. Expor chaves amplia o acesso indevido.

✗ C) Incorreta. Contas compartilhadas prejudicam auditoria e responsabilização.

> **Dica de prova:** Dados de IA: RBAC e menor privilégio.

03

CAPÍTULO 3 · 20-25% DA PROVA

Identificar uma estratégia de implementação e adoção dos aplicativos e serviços de IA da Microsoft

Domínio de menor peso, mas cheio de detalhes decorados: os seis princípios de IA responsável com uma prática típica de cada, os três gatilhos de usos sensíveis, a estrutura de governança da Microsoft (Responsible AI Council, Office of Responsible AI, Aether, Champs), o conselho de IA, o AI Center of Excellence e as políticas do Cloud Adoption Framework. A outra metade é dinheiro e adoção: licenças do Copilot (Copilot Chat incluído, add-on, Copilot Business, E7, pay-as-you-go), Copilot Credits e modelos de assinatura das Foundry Tools (camada gratuita, pay-as-you-go, commitment tier), além de barreiras de adoção, early adopters e programa de campeões. As alternativas erradas costumam ser extremas: extinguir o conselho, punir quem não usa, esperar um incidente, centralizar tudo na TI ou comprar licenças para todos sem medir. Onde mais erro: confundir qual órgão decide (Council) e qual implementa (Office), e quando usar pay-as-you-go x commitment tier.

Neste capítulo

19 tópicos-chave · 14 pegadinhas · 7 macetes · 14 gatilhos · 4 questões comentadas

Tópicos-chave

Equidade (fairness)

Pessoas em situações semelhantes devem receber tratamento ou recomendações semelhantes; se dois clientes parecidos diferem só pelo bairro, a equidade foi afetada. Práticas: equipe de design com origens e perspectivas diversas, avaliar origem e representatividade dos dados, analisar viés em dados e algoritmos e manter revisão humana com especialistas. Dados históricos de aprovações podem reproduzir vieses do passado.

Confiabilidade e segurança

O sistema deve funcionar como projetado, responder com segurança a condições imprevistas e resistir a manipulações. Práticas: testes rigorosos em laboratório e em cenários reais, inclusive imprevistos, definir quando passar o controle a uma pessoa e verificar em produção com auditorias e monitoramento. Riscos à segurança são tratados antes da implantação, não depois de acidentes.

Privacidade e segurança

Cumprir leis de proteção de dados, minimizar e limitar o uso à finalidade, proteger contra agentes mal-intencionados e dar às pessoas controle sobre seus dados. Dados pessoais são usados só pelo tempo e finalidade definidos, e o restante é excluído; criptografia não justifica retenção sem finalidade, e dados transferidos a terceiros devem ser protegidos, considerando anonimização.

Inclusão

Atender necessidades diversas, como acessibilidade, desde o design e não como correção posterior. Prática: incluir pessoas com deficiência nos testes e adotar padrões reconhecidos de acessibilidade. Princípios de design inclusivo da Microsoft: reconhecer a exclusão, resolver para um e estender para muitos, aprender com a diversidade.

Transparência

Tornar decisões compreensíveis: melhorar a inteligibilidade do modelo, explicar como funciona e onde pode falhar e treinar funcionários para interpretar as saídas. Ser honesto sobre o uso de IA: informar ao cliente que ele conversa com um agente. Publicar pesos do modelo ou uma taxa agregada não é transparência; simular um corretor real ou ocultar a IA contraria o princípio.

Responsabilidade (accountability)

Quem projeta e implanta IA responde por ela, e humanos mantêm controle significativo mesmo sobre sistemas autônomos. Práticas: responsáveis claros, conselhos de revisão e processo para corrigir o sistema quando age de forma injusta. A responsabilidade não é transferida ao fornecedor, a IA não é autoridade final em decisões que afetam pessoas e logs apoiam a auditoria.

Usos sensíveis (framework da Microsoft)

Três gatilhos: negação de serviços consequenciais (crédito, moradia, seguros, educação, emprego, saúde), risco de dano físico ou psicológico (ambientes críticos, como desligar reatores sem humano) e violação de direitos humanos (liberdade, expressão, privacidade, como reconhecimento facial para identificar pessoas em público). O processo tem três estágios: identificação (reportar por canal central), avaliação (fatos e efeito sobre pessoas e sociedade) e mitigação (podendo limitar, exigir supervisão ou desaconselhar o caso), com escalonamento ao Aether e ao Responsible AI Council.

Governança de IA responsável na Microsoft (hub-and-spoke)

Governança centralizada com o Responsible AI Council (fórum de líderes de negócio, pesquisa, políticas e engenharia, decisor final nos casos mais sensíveis e definidor dos princípios), o Office of Responsible AI (implementa e mantém o framework, políticas, capacitação e casos) e o Aether Committee (consultivo, com grupos de trabalho). Execução descentralizada pelos Responsible AI Champs, que mantêm suas funções, orientam equipes e líderes e escalam usos sensíveis sem papel de fiscalização. Lição do Tay (2016): antecipar uso indevido e ter salvaguardas desde o primeiro dia.

Sistema de governança eficaz

Tem mais chance de gerar mudança real quando reporta ao conselho de administração e tem orçamento, pessoal dedicado e autoridade clara para decidir e fazer cumprir. Para ética, a recomendação é um modelo híbrido (Chief Ethics Officer, escritório e comitê de ética), com autoridade central e execução descentralizada. O comitê de governança é multidisciplinar: TI, jurídico, compliance, negócio, gestão de riscos e RH. Para começar sem burocracia: pacote piloto com uma política, uma avaliação de impacto e um checklist de monitoramento.

Conselho de IA (modelo de maturidade)

Composição típica: patrocinador executivo, TI e plataforma, gestão da mudança e risco, jurídico e compliance. Alinha o uso de IA aos valores, revisa casos de alto impacto e define direitos de decisão. Conselho sem mandato, direitos de decisão e patrocínio é antipadrão: a correção é formalizar papel, direitos de decisão e caminhos de escalonamento, não extinguir nem aumentar reuniões.

Políticas de governança do CAF

Cobrem seleção e onboarding de modelos (sandbox para experimentar e validação antes do catálogo de produção), ferramentas e dados de terceiros (avaliação de privacidade, segurança e ética), padrões de sensibilidade (dados sensíveis separados dos públicos), qualidade de dados (golden dataset), manutenção (retreinamento conforme risco), compliance regulatório e conduta de usuários (termos de uso comunicados e controles contra uso indevido). A aplicação combina automação (Azure Policy, Purview) com treinamento e auditorias.

Avaliação e monitoramento de risco (CAF)

Avaliar cada carga de IA à luz dos princípios de IA responsável, num processo alinhado ao NIST AI RMF, incluindo riscos de dependências externas e de integração. Revisões trimestrais para alto risco, anuais para menor risco e avaliações independentes. Manter inventário central dos sistemas de IA com finalidade, local de execução e dono. Em compras de terceiros, avaliar práticas do fornecedor, exigir documentação e incluir requisitos em contratos e SLAs.

AI Center of Excellence (CoE)

Primeiro passo: garantir patrocínio executivo com comitê de direção de líderes de negócio e de TI; depois líder dedicado e time multidisciplinar (líderes de negócio e especialistas em dados, ML, governança, segurança e operações de IA). Responsabilidades: estratégia, capacitação, pilotos, padrões, intake e priorização, medir e reportar resultados e governar serviços implantados. Começa centralizado para acelerar e evolui para modelo consultivo com guardrails; se já existe Cloud CoE maduro, integra-se a ele.

Papéis da liderança na transformação com IA

Líderes de negócio impulsionam ideias, o Chief Digital Officer define a visão, o RH cuida de capacitação, requalificação e contratação (como cientistas de dados) e a TI operacionaliza infraestrutura e MLOps. A alta liderança garante patrocínio executivo, integra a IA ao planejamento e às métricas, comunica resultados esperados e compartilha as primeiras vitórias. Gestão de IA como portfólio: ganhos rápidos, apostas estratégicas e revisões trimestrais.

Barreiras comuns à adoção

Lacuna de habilidades: combinar capacitação interna, contratação de especialistas e parceiros da Microsoft. Resistência por medo: políticas claras de uso e treinamento sobre uso e limites. Shadow AI: catálogo de ferramentas aprovadas com política de uso e treinamento em alfabetização em IA, IA responsável, manuseio de dados e riscos do shadow AI. Pilotos dispersos: focar em 3 a 5 casos priorizados por valor, dados prontos e tempo até o valor. Dependência de poucos entusiastas: modelo operacional documentado com papéis, donos, intake e escalonamento.

Equipe de adoção, early adopters e campeões

A primeira turma do Copilot reúne usuários de diferentes áreas, de preferência com alto uso do Microsoft 365; eles viram campeões que mostram onde o Copilot traz valor e ajudam a escalar o suporte. Na comunidade do Copilot Studio, eventos são a faísca e a comunidade contínua é o motor; o ciclo autossustentável é eventos formam campeões, campeões ampliam a adoção e a adoção gera eventos. Papéis: líder da comunidade, líder de conteúdo, líder de eventos, coordenador de campeões (elo com o time central) e patrocinador executivo. Sustentar exige comunidade, reconhecimento, histórias de sucesso e medição de sentimento.

Impactos em dados, segurança e privacidade na implantação do Copilot

O Copilot só acessa o que o usuário já pode ver, por isso permissões amplas demais precisam ser corrigidas antes: SharePoint Advanced Management para revisar acessos, Restricted SharePoint Search como medida intermediária, rótulos de sensibilidade (com criptografia, o usuário precisa dos direitos de uso), Purview Audit para registrar interações, DSPM para IA para visibilidade do uso de apps de IA, MFA e acesso condicional. Agentes devem ser tratados como identidades gerenciadas para evitar agent sprawl. Adoção e impacto se medem com o Copilot Dashboard do Viva Insights e os relatórios de uso do admin center.

Licenças do Copilot e Copilot Credits

Copilot Chat é incluído sem custo extra para assinaturas elegíveis; o Copilot é add-on por usuário para planos elegíveis (custo previsível); Copilot Business atende organizações com até 300 usuários; o Microsoft 365 E7 já inclui o Copilot. O pay-as-you-go do Copilot liga uma política de cobrança a uma assinatura do Azure, vem desabilitado por padrão e aceita orçamento e alertas; é ideal quando a adoção é incerta. Não se atribui licença a convidados de outros locatários. Copilot Credits vêm de pay-as-you-go, pré-compra ou pacotes pré-pagos e o saldo não passa ao mês seguinte.

Modelos de assinatura das Foundry Tools

Camada gratuita para testar, preços por volume de transações, pay-as-you-go para pilotos e cargas variáveis e commitment tier para uso estável, com taxa fixa, desconto e previsibilidade. Commitment tier exige recurso de serviço único no nível Standard (não multisserviço), é mensal para web e contêineres conectados e anual, pago integralmente na compra, para contêineres desconectados. O excedente é cobrado pela tarifa de overage, o compromisso não é reembolsável e é possível trocar de plano para o mês seguinte ou desligar a renovação automática. Custos se acompanham no Microsoft Cost Management com alertas.

Macetes para memorizar

💡 Seis princípios de IA responsável

ECPITR: Equidade, Confiabilidade e segurança, Privacidade e segurança, Inclusão, Transparência, Responsabilidade.

💡 Usos sensíveis

SDD: Serviço consequencial negado, Dano físico ou psicológico, Direitos humanos violados. Processo IAM: Identificar, Avaliar, Mitigar.

💡 Governança da Microsoft

Council decide, Office opera, Aether aconselha, Champs orientam.

💡 Pay-as-you-go x commitment

Não sei o volume: pago pelo uso. Sei e é estável: me comprometo e ganho desconto.

💡 Commitment tier por local

Conectado = mensal; desconectado = anual e pago na hora.

💡 Licenças do Copilot

Chat grátis, add-on por usuário, Business até 300, E7 já vem com Copilot.

💡 Comunidade do Copilot Studio

Evento é faísca, comunidade é motor; o coordenador de campeões é a ponte.

Tabela de memorização: não confunda

Quando é...	...e quando é

Pay-as-you-go (Foundry Tools) Piloto, teste curto ou volume ainda desconhecido: sem compromisso, cobrado pelo consumo real.	**Commitment tier** Volume alto e estável já conhecido: taxa fixa com desconto e custo previsível; recurso de serviço único Standard.
Pay-as-you-go do Copilot Adoção incerta ou piloto departamental: política de cobrança na assinatura do Azure, com orçamento e alertas, decidindo depois sobre pré-pago.	**Licença add-on por usuário** Milhares de usuários com uso amplo e longo e necessidade de orçamento previsível.
Responsible AI Council Decisor final nos casos mais sensíveis e definidor dos princípios da empresa.	**Office of Responsible AI** Implementa e mantém a governança: políticas internas, estruturas, capacitação, casos e políticas públicas.
Cultura resistente a mudanças Começar com vitórias rápidas e visíveis, sustentadas por governança forte.	**Cultura orientada à inovação** Dar autonomia às equipes, sempre com guardrails definidos.
AI CoE centralizado Início da jornada, para acelerar a adoção com padrões comuns.	**AI CoE consultivo** Maturidade, quando o CoE virou gargalo: guardrails e entrega levada às equipes de plataforma e produto.
Negação de serviços consequenciais IA decide acesso a crédito, moradia, seguros, educação (bolsas), emprego ou saúde.	**Violação de direitos humanos** IA restringe liberdade, expressão ou privacidade, como reconhecimento facial para identificar pessoas de interesse em público.
Equidade Resultados diferentes para pessoas em situação semelhante; tratar com dados representativos, equipes diversas e revisão humana.	**Transparência** Clientes e auditores não entendem como a decisão é tomada; tratar com inteligibilidade, explicação e aviso de uso de IA.

Números que caem

Item	Valor
Copilot Business atende organizações com até	**300 usuários**
Casos de uso priorizados para sair de pilotos dispersos	**3 a 5**
Revisão de risco de cargas de IA de alto risco (CAF)	**trimestral**
Revisão de risco de cargas de menor risco (CAF)	**anual**
Período do commitment tier para web e contêineres conectados	**mensal**
Período do commitment tier para contêineres desconectados	**anual, pago integralmente na compra**

Chatbot Tay da Microsoft	**lançado em 2016; passou a repetir conteúdo de ódio em menos de 24 horas**

Alertas de prova (pegadinhas)

⚠ **Alerta de prova:** Escolher o Office of Responsible AI ou o Aether como decisor final dos casos sensíveis.
Como não cair: Decisor final e definidor de princípios é o Responsible AI Council; o Office implementa e mantém, o Aether é consultivo.

⚠ **Alerta de prova:** Dar aos Responsible AI Champs papel de fiscalizar e punir ou de decidir sozinhos casos sensíveis.
Como não cair: Champs orientam equipes e líderes, mantêm suas funções e escalam usos sensíveis pelos canais; não policiam.

⚠ **Alerta de prova:** Extinguir o conselho de IA sem autoridade ou resolver com reuniões diárias.
Como não cair: Conselho sem mandato é antipadrão; a correção é formalizar papel, direitos de decisão e escalonamento.

⚠ **Alerta de prova:** Formar o comitê de governança só com TI, só com cientistas de dados ou com fornecedores.
Como não cair: Governança é multidisciplinar (TI, jurídico, compliance, negócio, riscos, RH); fornecedores são avaliados, não governam.

⚠ **Alerta de prova:** Comprar licenças anuais para a empresa inteira no primeiro dia ou proibir o uso até estimar o número exato de usuários.
Como não cair: Com adoção incerta, comece com pay-as-you-go para medir e depois avalie pré-pago.

⚠ **Alerta de prova:** Achar que o pay-as-you-go do Copilot vem habilitado para todos, não permite limitar gastos ou é cobrado no cartão de cada usuário.
Como não cair: Vem desabilitado; o admin cria política de cobrança ligada a uma assinatura do Azure, com orçamento e alertas.

⚠ **Alerta de prova:** Usar Copilot Studio com Copilot Credits ou Foundry para liberar um chat básico sem custo.
Como não cair: Chat seguro sem custo extra é o Copilot Chat nas assinaturas elegíveis; Credits pagam agentes do Copilot Studio.

⚠ **Alerta de prova:** Tratar o Microsoft 365 E7 como obrigatório para PMEs usarem o Copilot.
Como não cair: E7 inclui o Copilot, mas não é obrigatório; para até 300 usuários existe o Copilot Business.

⚠ **Alerta de prova:** Contratar commitment tier com volume desconhecido, em recurso multisserviço, ou esperar reembolso e bloqueio no excedente.
Como não cair: Commitment tier é para demanda estável, exige recurso de serviço único Standard, cobra overage no excedente e não é reembolsável.

⚠ **Alerta de prova:** Liberar o Copilot e corrigir permissões quando alguém reclamar.
Como não cair: O Copilot respeita permissões existentes, então o excesso de compartilhamento se corrige antes (SharePoint Advanced Management, Restricted SharePoint Search, rótulos).

⚠ **Alerta de prova:** Medir adoção pelo número de licenças compradas ou de logins.
Como não cair: Use o Copilot Dashboard do Viva Insights e os relatórios de uso do admin center, ouvindo também o sentimento dos usuários.

⚠ **Alerta de prova:** Escolher a primeira turma do Copilot só de uma área, do conselho de administração ou de quem nunca usou o Microsoft 365.
Como não cair: Early adopters vêm de várias áreas, de preferência usuários intensivos do Microsoft 365.

⚠ **Alerta de prova:** Avaliar riscos só antes da implantação, só quando a imprensa noticiar ou pela própria equipe que desenvolveu.
Como não cair: Revisões trimestrais (alto risco), anuais (menor risco) e avaliações independentes, de forma proativa.

⚠ **Alerta de prova:** Contornar a lacuna de habilidades só com consultores temporários ou suspendendo projetos.
Como não cair: Combinar capacitação interna, contratação de especialistas e parceiros da Microsoft constrói capacidade sustentável.

Como cai na prova

📝 **Como cai:** Cenário de IA com um problema ético e a pergunta de qual princípio foi afetado ou qual prática está alinhada.
Como resolver: Identifique o sintoma: tratamento desigual (equidade), decisão incompreensível (transparência), ninguém responde (responsabilidade), comportamento inseguro em imprevistos (confiabilidade), exclusão de grupos (inclusão), uso indevido de dados pessoais (privacidade). Descarte as práticas que escondem, adiam ou removem a revisão humana.

📝 **Como cai:** Uso de IA descrito e a pergunta de em que categoria de uso sensível se enquadra.
Como resolver: Pergunte o que pode dar errado para a pessoa: perder acesso a crédito, educação, saúde (serviço consequencial), sofrer dano físico ou emocional (risco de dano) ou ter liberdade e privacidade violadas (direitos humanos). Nunca é 'baixo risco' só por ser interno ou em espaço público.

📝 **Como cai:** CFO compara opções de licenciamento ou de preço.
Como resolver: Pergunte se o volume é conhecido e estável (licença por usuário ou commitment tier) ou incerto (pay-as-you-go). Confira detalhes: Copilot Chat sem custo, Business até 300, E7 inclui, Credits sem acúmulo, commitment não reembolsável e com overage.

📝 **Como cai:** Organização com governança que não funciona ou adoção que travou.
Como resolver: A resposta formaliza e estrutura: dar autoridade, orçamento e direitos de decisão, modelo operacional com donos, foco em poucos casos priorizados, campeões e comunidade. Elimine extinguir, punir, esperar, centralizar tudo na TI ou abrir mais pilotos.

📝 **Como cai:** Preparação técnica para liberar o Copilot com segurança.
Como resolver: Lembre que o Copilot herda as permissões: revisar compartilhamento excessivo (SharePoint Advanced Management), medida intermediária (Restricted SharePoint Search), rótulos de sensibilidade, Purview Audit, MFA e acesso condicional; depois piloto com early adopters e medição com Copilot Dashboard.

Gatilhos do enunciado

Se o enunciado trouxer...	Pense em
dois clientes parecidos com resultados diferentes por um atributo como bairro	**Equidade**
clientes e auditores querem entender como o sistema decide	**Transparência (inteligibilidade e treinar quem interpreta)**
ninguém sabe quem responde pela solução que errou	**Responsabilidade (accountability)**
sistema seguro em situações inesperadas, passar o controle a pessoa	**Confiabilidade e segurança**
assistente de voz para pessoas com deficiência	**Inclusão**
guardar dados de candidatos para uso futuro	**Privacidade e segurança: limitar à finalidade e excluir**
decidir bolsa de estudo, crédito ou moradia por IA	**Uso sensível: negação de serviços consequenciais**
registrar e revisar interações com o Copilot para investigações	**Microsoft Purview Audit**

revisão de permissões do SharePoint ainda não terminou	**Restricted SharePoint Search como medida intermediária**
descobrir como funcionários usam apps de IA, inclusive de terceiros	**DSPM para IA no Microsoft Purview**
centenas de agentes sem responsável claro	**Agente como identidade gerenciada (evitar agent sprawl)**
volume alto e estável há um ano, quer desconto	**Commitment tier**
local sem internet, contêineres desconectados	**Commitment tier anual, pago integralmente na compra**
CoE virou gargalo nas aprovações	**Evoluir para modelo consultivo com guardrails**

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

- Responsible AI Council decide; Office of Responsible AI implementa; Champs orientam e escalam, sem fiscalizar.
- Os três gatilhos de uso sensível: serviços consequenciais, risco de dano e direitos humanos.
- Adoção incerta começa em pay-as-you-go; demanda estável vai para licença previsível ou commitment tier.
- O Copilot respeita permissões existentes, então corrigir compartilhamento excessivo vem antes da implantação.
- Respostas certas formalizam, capacitam e medem; erradas punem, escondem, esperam ou centralizam tudo.

Essenciais para a prova (checklist)

☐ Sei os seis princípios de IA responsável e uma prática típica de cada um.

☐ Sei classificar cenários nos três gatilhos de uso sensível e as etapas identificação, avaliação e mitigação.

☐ Sei o papel do Responsible AI Council, do Office of Responsible AI, do Aether e dos Responsible AI Champs.

☐ Sei o que torna um sistema de governança eficaz e a composição de um conselho de IA.

☐ Sei as políticas de governança do CAF, a frequência de revisão de risco e o papel do inventário central.

☐ Sei montar e evoluir um AI Center of Excellence.

☐ Sei tratar barreiras de adoção (habilidades, resistência, shadow AI, pilotos dispersos) e montar programa de campeões.

☐ Sei os controles de preparação do Copilot (SharePoint Advanced Management, Restricted SharePoint Search, rótulos, Purview Audit, DSPM para IA).

☐ Sei as opções de licença do Copilot, o pay-as-you-go e as regras dos Copilot Credits.

☐ Sei quando usar pay-as-you-go ou commitment tier nas Foundry Tools e as regras de período, overage e reembolso.

Questões de revisão do capítulo

R3.1 · Identificar uma estratégia de implementação e adoção dos aplicativos e serviços de IA da Microsoft

A Indústria de Cosméticos Bela Pele vai escolher o primeiro grupo de usuários do Microsoft 365 Copilot para gerar aprendizado e futuros campeões. Segundo a Microsoft, quem deve compor esse grupo?

A. Usuários de diferentes áreas de negócio, de preferência com alto uso dos recursos do Microsoft 365
B. Pessoas de uma única área, escolhidas por sorteio entre os voluntários
C. Os membros do conselho de administração, por terem a maior visibilidade dentro da empresa
D. Funcionários que nunca usaram o Microsoft 365, para medir o impacto máximo

R3.2 · Identificar uma estratégia de implementação e adoção dos aplicativos e serviços de IA da Microsoft

A Rede de Lojas Casa Bonita já liberou agentes para todas as áreas, mas o uso cai depois das primeiras semanas. Segundo o modelo de maturidade de agentes da Microsoft, como sustentar a adoção?

A. Encerrar o programa de adoção logo após a entrega das licenças
B. Reforçar a adoção com comunidades, campeões, reconhecimento e histórias de sucesso, medindo o sentimento
C. Impor o uso por metas punitivas, sem comunicação nem reconhecimento das equipes que avançam mais rápido no uso
D. Medir o número de logins, sem ouvir os usuários

R3.3 · Identificar uma estratégia de implementação e adoção dos aplicativos e serviços de IA da Microsoft

A empresa de segurança privada Vigia Total quer usar reconhecimento facial em áreas abertas ao público para identificar pessoas de interesse. Pelo framework de usos sensíveis da Microsoft, em que categoria esse cenário se enquadra?

A. Negação de serviços consequenciais, como crédito, moradia e saúde
B. Risco de dano físico em ambientes industriais críticos de manufatura
C. Violação de direitos humanos, como liberdade e privacidade
D. Uso de baixo risco, pois o espaço monitorado é aberto ao público

R3.4 · Identificar uma estratégia de implementação e adoção dos aplicativos e serviços de IA da Microsoft

O Banco Popular Integra usa IA para apoiar decisões de crédito, e clientes e auditores querem entender como o sistema chega às recomendações. Qual prática está alinhada ao princípio de transparência?

A. Melhorar a inteligibilidade do modelo e treinar os funcionários para interpretar as saídas
B. Substituir o modelo por regras ocultas que ninguém precisa explicar
C. Manter o modelo em sigilo completo, para proteger a vantagem competitiva
D. Divulgar a taxa de aprovação mensal ao público, sem detalhar critérios, dados usados nem limitações do modelo

Respostas comentadas

R3.1. A) Usuários de diferentes áreas de negócio, de preferência com alto uso dos recursos do Microsoft 365

Por quê: Early adopters bem escolhidos mostram onde o Copilot gera valor e se tornam campeões que ajudam os colegas a adotar a ferramenta.

✗ B) Incorreta. Uma única área limita o aprendizado sobre os diferentes usos.
✗ C) Incorreta. Visibilidade não substitui diversidade de áreas e intensidade de uso.
✗ D) Incorreta. Quem já usa bem o Microsoft 365 tende a extrair valor mais rápido e servir de referência.

> **Dica de prova:** Primeira turma do Copilot: várias áreas e usuários intensivos do Microsoft 365.

R3.2. B) Reforçar a adoção com comunidades, campeões, reconhecimento e histórias de sucesso, medindo o sentimento

Por quê: Adoção liderada por comunidade, apoio visível da liderança e práticas estruturadas de mudança aceleram a adoção e reduzem custos de capacitação no longo prazo.

✗ A) Incorreta. A adoção precisa de reforço contínuo.
✗ C) Incorreta. Punição sem comunicação aumenta a resistência.
✗ D) Incorreta. Logins não mostram confiança nem valor percebido.

> **Dica de prova:** Sustentar a adoção: comunidade, campeões, reconhecimento e histórias.

R3.3. C) Violação de direitos humanos, como liberdade e privacidade

Por quê: No caso estudado pela Microsoft, identificar pessoas de interesse em abordagens policiais foi considerado prematuro, porque erros de identificação podem levar a detenções indevidas.

✗ A) Incorreta. Essa categoria trata de negar acesso a serviços como crédito, moradia, educação e saúde.
✗ B) Incorreta. Essa categoria trata de riscos à integridade física ou emocional.
✗ D) Incorreta. Identificar pessoas em ambientes não controlados é cenário de alto risco.

> **Dica de prova:** Vigilância e identificação de pessoas: uso sensível por direitos humanos.

R3.4. A) Melhorar a inteligibilidade do modelo e treinar os funcionários para interpretar as saídas

Por quê: Quando sistemas de IA informam decisões que afetam a vida das pessoas, como crédito e contratação, é essencial que elas entendam como as conclusões são alcançadas.

✗ B) Incorreta. Regras ocultas contrariam a transparência.

✗ C) Incorreta. Sigilo total impede que as partes interessadas entendam as decisões.

✗ D) Incorreta. Uma taxa agregada não explica como as decisões são tomadas.

> **Dica de prova:** Transparência: decisões compreensíveis e pessoas treinadas para interpretá-las.

DIA DA PROVA

Estratégia de prova

O AB-731 é uma prova para tomadores de decisão de negócio, não para quem escreve código: quase todo enunciado traz um executivo (CEO, CFO, CISO, diretora de RH) com um problema concreto e pede a decisão, o produto ou a prática que a Microsoft recomenda. As questões são de múltipla escolha e múltipla resposta (Escolha DUAS ou TRÊS), com cenários de liderança. As alternativas erradas costumam ser extremas ou fora de papel: punir, esconder, esperar o incidente, dispensar revisão humana, escolher o maior modelo, aumentar a temperatura, construir do zero sem necessidade. Muitas vezes duas alternativas citam produtos reais; ganha quem casa a necessidade exata com a função de cada um. A pontuação vai de 1 a 1000, com mínimo de 700, e a prova está disponível em Portuguese (Brazil).

Como responder

✓ Leia primeiro a pergunta final e sublinhe a restrição que decide: 'sem custo adicional', 'sem retreinar', 'dados que mudam todo dia', 'menor responsabilidade operacional', 'sem código'.

✓ Elimine logo as alternativas extremas ou absolutas (nunca, sempre, dispensa, garante, sem revisão), que raramente são a recomendação da Microsoft.

✓ Em questões de produto, identifique a entrada e a saída do cenário (texto, imagem, áudio, planilha, documento) e só então escolha a ferramenta.

✓ Nas questões Escolha DUAS ou TRÊS, avalie cada afirmação isoladamente como verdadeiro ou falso; a quantidade pedida vem no enunciado.

✓ Na dúvida entre opções do Cloud Adoption Framework, prefira a que atende aos requisitos com menor responsabilidade operacional e a que começa pequeno, mede e escala.

✓ Lembre que o Copilot sempre respeita as permissões existentes e que dados do cliente não treinam modelos de base; isso elimina várias alternativas de segurança.

✓ Com cerca de 45 minutos de prova, não trave: marque para revisão a questão em que estiver entre duas opções e volte no fim.

✓ Distribua o estudo pelos pesos: valor de negócio e aplicativos da Microsoft somam 35 a 40% cada; estratégia de adoção vale 20 a 25%, mas tem muitos detalhes decorados.

Erros que derrubam candidatos

✗ Confundir RAG (dado atual, sem alterar pesos) com fine-tuning (tom, formato e terminologia, alterando pesos).

✗ Escolher 'aumentar a temperatura' ou 'usar o maior modelo' como solução de qualidade ou custo.

✗ Achar que licença paga ou prompt amplia o acesso a dados que o usuário não pode ver.

✗ Trocar papéis da governança: Office of Responsible AI no lugar do Responsible AI Council, ou Champs fiscalizando.

✗ Confundir pay-as-you-go com commitment tier, ou Copilot Credits com licença do Copilot nos apps.

✗ Trocar ferramentas de função parecida: Translator x Document Intelligence, Researcher x Analyst, Pages x Notebooks, Copilot connectors x Graph Data Connect.

✗ Responder como técnico (escolher algoritmo, escrever código) quando a pergunta pede o papel do líder de negócio.

Na semana da prova

- ☐ Revise as comparações de cada domínio até conseguir explicar cada par em uma frase.
- ☐ Treine com simulados cronometrados em 45 minutos e anote cada erro com o conceito que faltou.
- ☐ Releia as listas que mais derrubam: seis princípios de IA responsável, três gatilhos de uso sensível, licenças do Copilot e modelos de preço das Foundry Tools.
- ☐ Monte uma tabela de uma página ligando cenário a produto (apps do Copilot, agentes, Foundry Tools, Copilot Studio x Foundry x infraestrutura).
- ☐ Na véspera, revise só macetes e pegadinhas, sem conteúdo novo, e use a área restrita do exame (exam sandbox) para conhecer a interface e como marcar questões para revisão.
- ☐ Planeje as pausas: o relógio da prova continua correndo durante a pausa, e depois de iniciá-la não dá para voltar às questões já vistas, mesmo as marcadas para revisão.

REVISÃO FINAL

Folha de cola

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

1. Identificar benefícios, recursos e oportunidades dos aplicativos e serviços de IA da Microsoft

‣ planilhas e CSV, tendências, outliers, por região e trimestre → **Agente Analyst**

‣ análise competitiva com fontes da web e internas, citadas → **Agente Researcher**

‣ transformar resposta do Copilot em conteúdo que a equipe edita junto → **Copilot Pages**

‣ reunir arquivos, notas, links e conversas de um projeto → **Copilot Notebooks**

‣ agentes prontos da Microsoft e de parceiros sem desenvolver → **Agent Store do Microsoft Copilot**

⚠ Copilot Chat e Copilot licenciado respeitam as permissões existentes; nenhuma licença ou prompt concede acesso adicional.

⚠ Para consultar estoque ou ERP sem sair do Copilot é preciso estender: agentes e Copilot connectors.

⚠ O Graph já faz parte do Microsoft 365 e fornece ao Copilot o contexto que cada usuário pode ver.

2. Identificar o valor de negócio das soluções de IA generativa

‣ dado que muda todo dia, versão vigente, tabela de preço atual → **RAG / grounding em fontes autorizadas**

‣ mesmo tom e mesma estrutura, terminologia própria, formato rígido de submissão → **Fine-tuning com exemplos curados**

‣ formato próprio de categorias sem retreinar o modelo → **Few-shot no prompt**

‣ modelo grande caro e lento, quer menor com qualidade parecida → **Destilação (professor gera exemplos para o menor)**

‣ papel bem definido e limites do assistente → **Mensagem de sistema**

⚠ Temperatura alta só aumenta a variação. Em contratos, classificação, relatórios padronizados e segurança, a resposta nunca é subir a temperatura; em documentos legais, a indicação é temperatura baixa.

⚠ Fine-tuning ajusta comportamento e formato, não traz informação atual. Dado que muda todo dia vem de recuperação (RAG).

⚠ RAG traz informação para a resposta, mas não garante formato. Tom e formato consistentes com bons dados históricos apontam para fine-tuning.

3. Identificar uma estratégia de implementação e adoção dos aplicativos e serviços de IA da Microsoft

‣ dois clientes parecidos com resultados diferentes por um atributo como bairro → **Equidade**

‣ clientes e auditores querem entender como o sistema decide → **Transparência (inteligibilidade e treinar quem interpreta)**

‣ ninguém sabe quem responde pela solução que errou → **Responsabilidade (accountability)**

‣ sistema seguro em situações inesperadas, passar o controle a pessoa → **Confiabilidade e segurança**

‣ assistente de voz para pessoas com deficiência → **Inclusão**

⚠ Decisor final e definidor de princípios é o Responsible AI Council; o Office implementa e mantém, o Aether é consultivo.

⚠ Champs orientam equipes e líderes, mantêm suas funções e escalam usos sensíveis pelos canais; não policiam.

⚠ Conselho sem mandato é antipadrão; a correção é formalizar papel, direitos de decisão e escalonamento.

GABARITO

Respostas do diagnóstico

D1. D) Microsoft Fabric, pois a empresa já usa a plataforma para seus dados e suas análises → revise o capítulo 1

Por quê: Para treinar ML com dados próprios: Fabric se a organização já trabalha nele, Azure Machine Learning nos demais casos.

✗ A) Incorreta. O Copilot Studio cria agentes; não é a plataforma de treinamento de modelos de ML.
✗ B) Incorreta. O Copilot no Excel ajuda na análise, mas não é a plataforma para treinar e implantar modelos de ML.
✗ C) Incorreta. O Content Safety detecta conteúdo nocivo, não treina modelos de previsão.

> **Dica de prova:** ML com dados próprios: Fabric se já está nele; senão, Azure Machine Learning.

D2. B) Acompanhar as datas de aposentadoria e planejar a migração para versões novas → revise o capítulo 1

Por quê: O catálogo do Foundry reúne milhares de modelos de vários provedores; acompanhar o ciclo de vida evita interrupções quando versões antigas deixam de existir.

✗ A) Incorreta. Trocar sem planejamento e avaliação aumenta riscos.
✗ C) Incorreta. Congelar o app não evita a aposentadoria da versão.
✗ D) Incorreta. Versões de modelo são aposentadas ao longo do tempo.

> **Dica de prova:** Modelo tem data de aposentadoria: planeje a migração.

D3. B) Não; para licenciados, agentes no Copilot, Teams ou SharePoint com grounding do Graph não consomem créditos → revise o capítulo 1

Por quê: A licença do Microsoft 365 Copilot inclui Copilot Chat, Copilot nos apps, capacidades de agentes do Copilot Studio quando usadas no Microsoft Copilot, Teams ou SharePoint, e Copilot Analytics.

✗ A) Incorreta. Outros canais, como sites públicos, consomem Copilot Credits.
✗ C) Incorreta. Esse cenário específico, para usuários licenciados, não consome créditos.
✗ D) Incorreta. Há várias formas de licenciamento, e a licença do Copilot cobre esse cenário.

> **Dica de prova:** Usuário licenciado + agente no Copilot, Teams ou SharePoint com Graph: sem consumo de créditos.

D4. A) Viés nas saídas, vazamento de dados e falhas de compliance → revise o capítulo 2

Por quê: Confiança impulsiona a adoção: dados de qualidade e segurança robusta não são opcionais, são a base da IA responsável.

✗ B) Incorreta. Cortar governança não traz esses benefícios e aumenta riscos.
✗ C) Incorreta. Licenças e treinamentos não ficam mais baratos e o risco aumenta.
✗ D) Incorreta. A confiança impulsiona a adoção; menos controles tendem a reduzi-la.

> **Dica de prova:** Dados e segurança fracos: viés, vazamento e falha de compliance.

D5. D) Criar e testar planos de resposta para incidentes de IA, com escalonamento claro → revise o capítulo 2

Por quê: Detectar ameaças de IA envolve detecção automatizada (como a gestão de postura de segurança de IA no Defender for Cloud), resposta a incidentes específica e monitoramento por plataforma.

✗ A) Incorreta. Logs são essenciais para detectar e investigar incidentes.
✗ B) Incorreta. Incidentes de IA têm características próprias que planos genéricos não cobrem.
✗ C) Incorreta. A resposta precisa estar definida antes do incidente.

> **Dica de prova:** Incidente de IA: plano específico, testado e com escalonamento claro.

D6. D) Setores regulados, cenários de compliance e aplicações voltadas a clientes → revise o capítulo 2

Por quê: Para grounding, líderes devem garantir acesso a fontes autorizadas, integração entre ferramentas de IA e sistemas corporativos e políticas de governança para a precisão dos dados.

✗ A) Incorreta. Usos criativos dependem menos de fatos verificados.
✗ B) Incorreta. Em atividades lúdicas, o impacto de imprecisões é baixo.
✗ C) Incorreta. Rascunhos descartáveis não são o foco prioritário de grounding.

> **Dica de prova:** Grounding é prioridade em setor regulado, compliance e atendimento a clientes.

D7. B) Testar com rigor, inclusive em cenários imprevistos, e definir quando passar o controle a uma pessoa → revise o capítulo 3

Por quê: Sistemas de IA devem funcionar como projetado, responder com segurança a condições imprevistas e resistir a manipulações; auditorias e monitoramento verificam isso em produção.

✗ A) Incorreta. Riscos à segurança devem ser tratados antes da implantação.
✗ C) Incorreta. Em decisões de alto impacto, as pessoas precisam ter o nível adequado de participação.
✗ D) Incorreta. Sistemas de alto risco devem ser testados em laboratório e em cenários reais.

> **Dica de prova:** Confiabilidade e segurança: testes rigorosos e passagem de controle a humanos.

D8. B) Copilot Chat, incluído sem custo extra para usuários com assinatura elegível → revise o capítulo 3

Por quê: O Copilot Chat é a experiência básica e segura de chat de IA incluída nos planos elegíveis do Microsoft 365; recursos nos apps, dados organizacionais amplos e agentes avançados dependem de licença.

✗ A) Incorreta. Copilot Credits são pagos e servem para agentes do Copilot Studio, não para liberar um chat básico.
✗ C) Incorreta. A licença add-on do Copilot tem custo por usuário.
✗ D) Incorreta. O Foundry é uma plataforma para criar soluções sob medida, com cobrança por uso.

> **Dica de prova:** Chat de IA seguro sem custo extra: Copilot Chat nas assinaturas elegíveis.

D9. A) Antecipar o uso indevido e projetar salvaguardas desde o primeiro dia → revise o capítulo 3

Por quê: Responsabilidade precisa acompanhar a capacidade: o comportamento humano pode explorar sistemas de aprendizado, por isso salvaguardas e monitoramento devem existir desde o primeiro dia.

✗ B) Incorreta. A lição não é abandonar o aprendizado com dados, e sim prever abusos e criar defesas.

✗ C) Incorreta. O caso mostra justamente o custo de não ter salvaguardas no lançamento.

✗ D) Incorreta. Riscos de uso indevido também existem em ambientes internos.

> **Dica de prova:** Lição do Tay: prever o abuso e ter salvaguardas desde o lançamento.

Agora, treine no formato real

Faça as 6 provas completas do curso AB-731: 300 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 Microsoft. As questões são inéditas e não reproduzem a prova real.