Guia de Implementação de IA Generativa Responsável

Empoderando práticas de IA responsável com a Microsoft

Relatório encomendado pela Microsoft e produzido pela EY

Shape the future with confidence

Sumário Executivo

1

Seção 1.

Introdução

1 Introdução

A IA Generativa, também chamada de *GenAI*, teve uma adoção digital sem precedentes. O ChatGPT atingiu 100 milhões de usuários em janeiro de 2023, o crescimento de usuários mais rápido de qualquer aplicativo na história[1]. A proliferação dessa tecnologia está compelindo empresas de diversos setores a reavaliar suas estratégias de negócio e a buscar a adaptação ágil para garantir a competitividade em um mercado dinâmico.

Nesse contexto, no vasto portfólio da IA Generativa, podem ser citados algoritmos avançados capazes de gerar textos, imagens, vídeos e até mesmo códigos de programação, atuando como aceleradores na produtividade das empresas e no desenvolvimento de novas soluções. A *GenAI* não se limita somente a interpretar e analisar dados existentes, mas é capaz também de criar conteúdo, utilizando-se, por exemplo, de modelos mais avançados como redes generativas adversárias (*GANs*) e modelos de linguagem autorregressivos. O diferencial disruptivo da *GenAI* reside na sua habilidade de gerar conteúdos novos com uma eficiência e velocidade sem precedentes, o que potencializa a capacidade das empresas de não apenas acompanhar as tendências de mercado, mas de serem as pioneiras na criação destas, estabelecendo novos padrões e expectativas. Ao desenvolver produtos e serviços que antecipam e moldam as preferências dos consumidores, as empresas podem oferecer experiências personalizadas e originais, atendendo às necessidades dos clientes de forma mais criativa e impactante, e cada vez mais integradas às ferramentas usuais de trabalho, como é o caso do ***Microsoft Copilot,*** que já pode ser utilizado junto ao Word, Teams, Power Point e Excel, por exemplo. Essa capacidade de autonomia na construção de conteúdos abre um leque de possibilidades para a inovação em larga escala.

> ***"Para lidar com o potencial disruptivo dessa tecnologia, é fundamental adotar uma postura informada, buscando um equilíbrio entre inovação e responsabilidade."***

A *GenAI* compõe um subgrupo do campo maior da Inteligência Artificial, com diferenças significativas das IA tradicionais[2]. A IA tradicional é especializada em identificar padrões e extrair *insights* de conjuntos de dados, sendo altamente eficaz em tarefas como classificação, previsão, clusterização e detecção de anomalias. Na maioria dos casos, ela opera dentro de um escopo específico, baseado no consumo de dados de entrada estruturados, aprendizado de padrões e geração de resultados a partir da replicação desses padrões identificados. Em contraste, a IA Generativa revoluciona esse paradigma pela capacidade de criar conteúdo, sejam textos, imagens, vídeos ou áudios, indo além da reprodução de padrões existentes.

Essa versatilidade é consequência do seu tipo de treinamento. As *LLMs (Large Language Models), por exemplo,* são treinadas em enormes quantidades de dados textuais, o que lhes permite aprender padrões linguísticos, gramaticais e até mesmo nuances estilísticas. Essa capacidade de compreensão e geração de linguagem natural as torna ferramentas poderosas para uma variedade de aplicações,

como *chatbots* sem *scripts*, tradução automática, criação de conteúdo e análises de conteúdo.

Com a crescente popularidade da Inteligência Artificial Generativa no mundo corporativo, os executivos se deparam com novos desafios na hora de integrar essa tecnologia em suas operações, principalmente em relação à compatibilidade com os valores e objetivos da empresa[3].

De acordo com o CEO Outlook Pulse da EY[4], 65% dos CEOs reconhecem o potencial da *GenAI* para impulsionar a produtividade e resultados positivos para todas as partes interessadas, sem deixar de atentar para[5]: 1. a rastreabilidade limitada das fontes; 2. a tomada de decisões factualmente erradas; 3. comprometimento da privacidade dos dados pessoais; 4. aumento do risco de violação de dados; e 5. resultados não reproduzíveis.

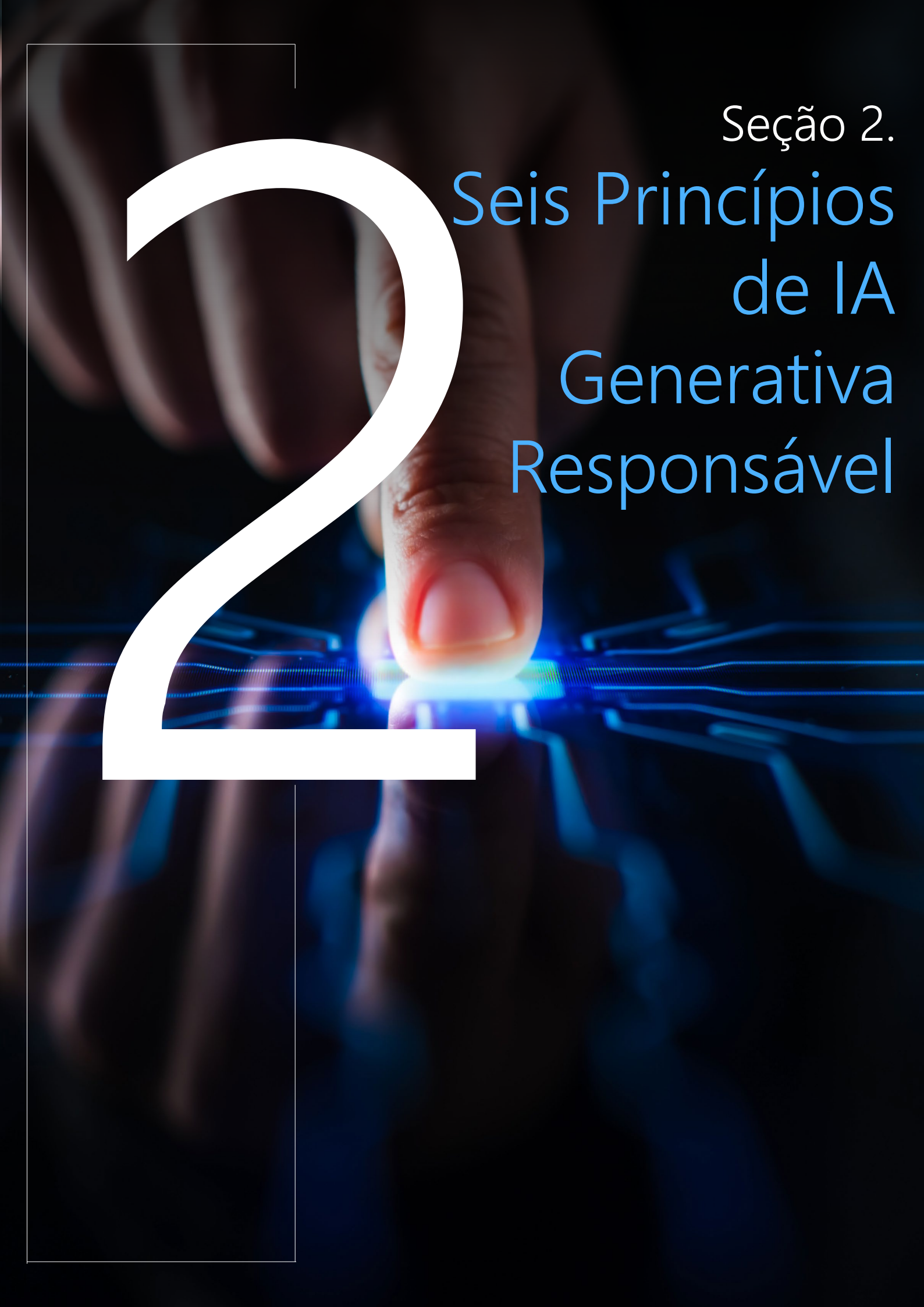
CEOs estão elaborando estratégias de investimento para maximizar os benefícios atuais e futuros que a Inteligência Artificial pode trazer para os seus negócios[4].

Não obstante as apreensões inerentes à disrupção tecnológica, a comunidade empresarial reconhece a inevitabilidade da adoção da Inteligência Artificial. Em resposta, os líderes executivos estão priorizando o desenvolvimento de estruturas de governança robustas, incluindo controles, processos e políticas, para assegurar que a implementação da IA seja conduzida com equidade, confiabilidade, segurança e privacidade de dados.

Embora a implementação segura e ética da *GenAI* seja um tema complexo e com diversas nuances, é possível integrar essa tecnologia de forma alinhada aos valores e objetivos da empresa, desde que haja cuidado e diligência.

Na próxima seção, mostramos como os princípios da IA Responsável da Microsoft[6] abordam as preocupações do *C-Level* e como eles podem ser usados para criar uma diretriz prática e adaptável para produtos de IA confiáveis *by-design* e uma governança robusta e eficaz no gerenciamento de riscos.

"É possível integrar essa tecnologia de forma alinhada aos valores e objetivos da empresa, desde que haja cuidado e diligência."



Seção 2.

2 Seis Princípios de IA Generativa Responsável

2 Seis Princípios da IA Responsável no Contexto da *GenAI*

A aplicação de IA Generativa tem o potencial de transformar diversos setores, automatizando e criando conteúdo de novas formas. No entanto, a utilização responsável dessas tecnologias é crucial para garantir resultados precisos, éticos e seguros.

A IA responsável enfatiza o uso ético da IA, considerando as implicações sociais de seu desenvolvimento e aplicação. Pioneira na compreensão da necessidade de uma IA responsável, a Microsoft publicou seis princípios de IA responsável (Figura 1), os quais entendemos fundamentais para nortear o uso eficaz da Inteligência Artificial[6,7]:

1. Equidade: estabelece que a IA deve tratar todas as pessoas de maneira equânime. Por exemplo, um modelo de IA de recomendação de empregos deve oferecer as mesmas oportunidades para candidatos com as mesmas qualificações, independentemente de raça, gênero, etnia e outras características protegidas por lei[6]. Para assegurar a equidade na modelagem de IA tradicional, é importante realizar auditorias regulares nos dados de treinamento para identificar e eliminar vieses, implementar técnicas de pré-processamento para equilibrar a distribuição de classes nos dados e utilizar métricas de equidade para avaliar o desempenho do modelo.

2. Confiabilidade: determina que a IA deve ser confiável e segura. Um exemplo é um modelo de diagnóstico médico baseado em IA que deve fornecer resultados precisos e ser resiliente a manipulações mal intencionadas. Na modelagem de IA tradicional, isso pode ser alcançado por meio de testes de validação dos dados de entrada e dos resultados do modelo, detecção de anomalias, autenticação multifator (*MFA*), *patch management* e criptografia de dados em trânsito.

3. Privacidade e segurança: assegura que a IA deve garantir a privacidade e a segurança dos dados[6]. Para estar de acordo com a lei geral de proteção dos dados pessoais (LGPD), é necessário dar transparência de como os dados são coletados, processados e armazenados, além do controle apropriado ao usuário para escolher como seu dado é usado. Um exemplo seria um *chatbot* de atendimento ao cliente que protege a privacidade das informações pessoais dos usuários e pede autorização para armazenar os dados. Na modelagem de IA tradicional, isso pode ser feito utilizando técnicas de anonimização e protocolos de criptografia de dados, implementando políticas de acesso e controle e realizando auditorias de conformidade regulares.

4. Inclusão: defende que a IA deve capacitar e envolver a todos[6]. Por exemplo, um sistema de tradução automática deve suportar múltiplos idiomas e dialetos, incluindo aqueles menos representados. Para garantir a inclusão na modelagem de IA tradicional, é importante incluir dados de populações diversas, desenvolver interfaces e *outputs* acessíveis e realizar testes de usabilidade com diversos grupos de usuários.

5. Transparência: afirma que a IA deve ser compreensível e explicável[6]. Um exemplo é um sistema de pontuação de crédito que fornece explicações claras sobre como as decisões são tomadas. Na modelagem de IA tradicional, técnicas de explicabilidade, que a Gartner define como "as capacidades de esclarecer o funcionamento de um modelo[8]", como *LIME* ou *SHAP*, podem ser

utilizadas para interpretar decisões do modelo, além de documentar o processo de desenvolvimento e fornecer relatórios de transparência detalhados[7].

6. **Responsabilidade:** determina que as organizações devem ser responsáveis pelo funcionamento e uso da IA[6]. Por exemplo, uma empresa que utiliza IA para avaliações de desempenho deve ser responsável pelos impactos dessas avaliações nos funcionários. Na modelagem de IA tradicional, isso pode ser feito implementando esteiras de *MLOps* (*Machine Learning Operations*) e uma governança de IA clara em todo o ciclo de vida da modelagem, treinando equipes sobre as implicações éticas e legais do uso de IA e mantendo canais abertos para *feedback*.

A Inteligência Artificial Generativa exige a adaptação dos seis princípios de IA responsável para enfrentar seus desafios únicos. A principal diferença entre a *GenAI* e a IA tradicional reside na natureza dos resultados: a *GenAI* cria conteúdo novo, enquanto a IA tradicional executa tarefas específicas baseadas em padrões nos dados.

Para se treinar esses modelos capazes de gerar novo conteúdo, é necessário um grande volume de dados, comparado ao tamanho típico dos *datasets* utilizados no treino da IA tradicional. Essa distinção fundamental implica em diferentes necessidades de controles e mitigação. Ambas inteligências necessitam de controle. A diferença básica está em onde esses controles serão posicionados. Enquanto na IA tradicional esses controles estão posicionados na entrada e no treino do modelo, na *GenAI* esses controles estão posicionados na saída do modelo e nos *prompts*.

A natureza dos modelos de linguagem de grande escala (LLMs) modifica a abordagem de se interpretar suas previsões. Ao contrário dos sistemas tradicionais de IA, as LLMs não codificam regras explícitas, elas utilizam representações ocultas, chamadas de *embeddings*, que são abstratas e de difícil interpretação, por isso é mais custoso computacionalmente identificar as características que influenciam seus resultados.

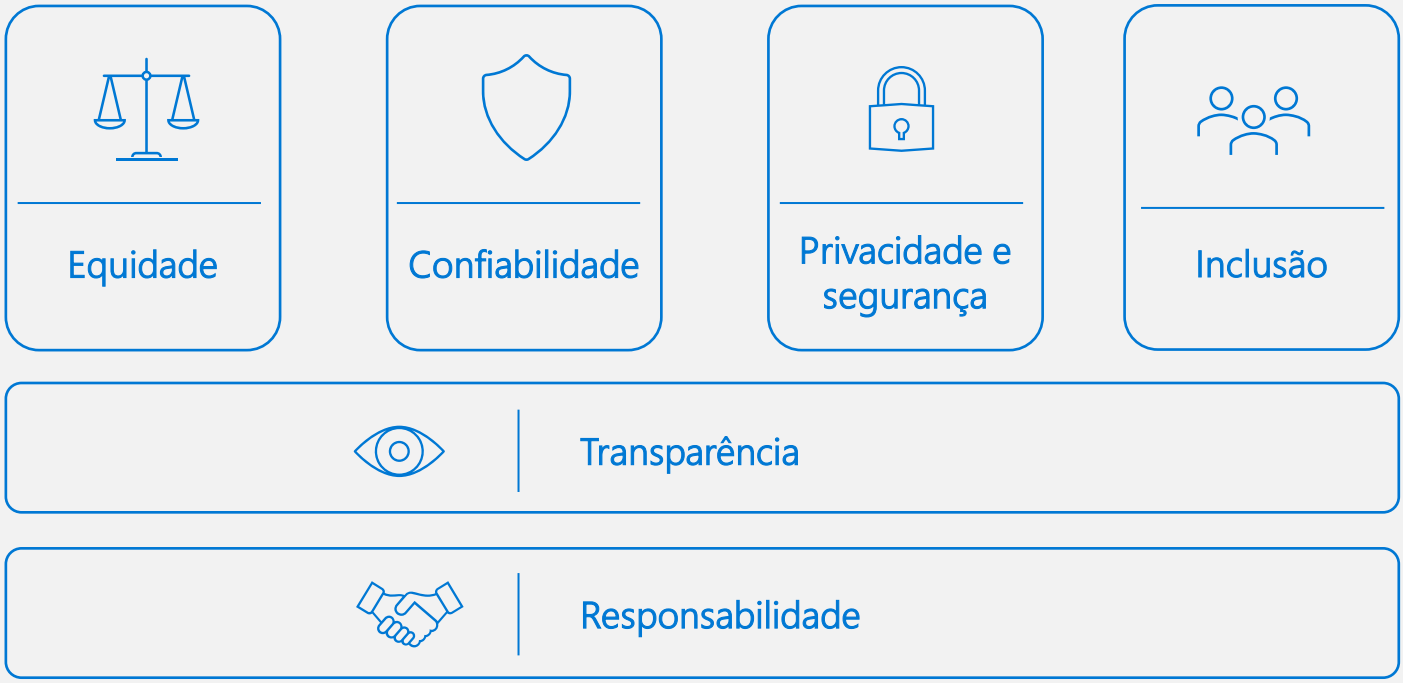


Figura 1. Princípios da IA Responsável[1]

A chave para aproveitar ao máximo o potencial dos LLMs reside na escolha do modelo e na formulação precisa dos *prompts*. A engenharia de *prompt*, que envolve a elaboração das instruções fornecidas ao modelo, permite direcionar o LLM para gerar resultados relevantes e alinhados com as expectativas do usuário. Por exemplo, ao solicitar a um LLM que gere um texto criativo, um *prompt* bem elaborado pode especificar o gênero, o tom, o público-alvo e outros detalhes.

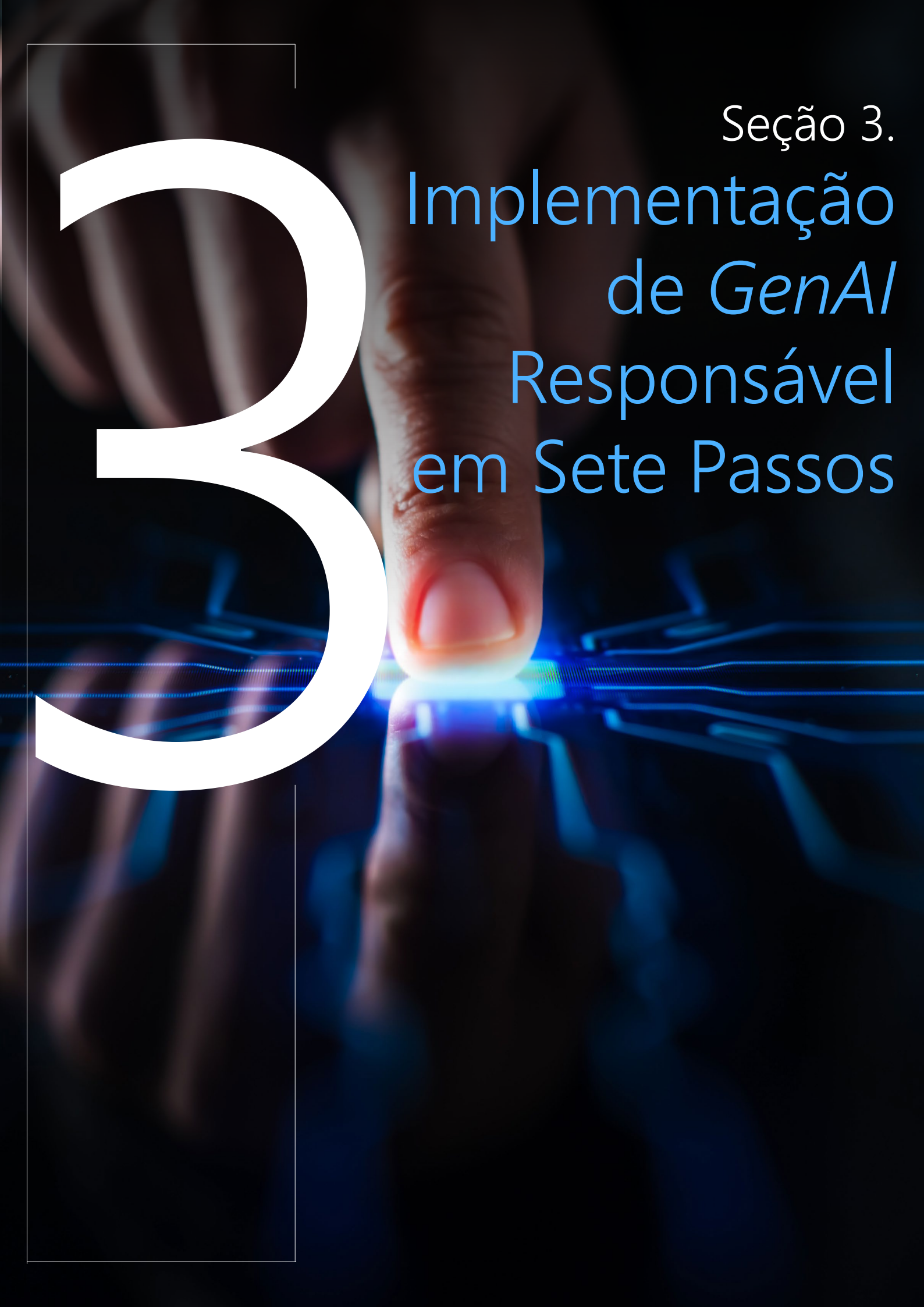
Mesmo com a utilização de modelos de *GenAI* adequados e *prompts* precisos, as informações geradas devem ser fundamentadas em evidências factuais e/ou conhecimento científico.

Nesse contexto, o usuário assume um papel fundamental na validação e verificação das informações produzidas pela IA, mitigando o risco de disseminação de informações falsas ou imprecisas. A responsabilidade pela qualidade dos resultados é, portanto, compartilhada entre os desenvolvedores dos modelos e os usuários finais.

Para viabilizar a supervisão humana efetiva, os modelos de *GenAI* são concebidos para funcionar como um copiloto tecnológico, oferecendo suas análises e resultados na forma de sugestões. Isso coloca o usuário no papel de piloto em comando, que deve revisar e validar as informações antes de decidir sobre sua aplicação. Essa abordagem assegura que o usuário permaneça como a autoridade final, com pleno controle sobre a adoção ou rejeição do conteúdo gerado. Diferentemente da IA tradicional, onde os resultados eram muitas vezes aplicados de forma automática aos processos, na *GenAI* o usuário e a tecnologia operam em uma dinâmica de colaboração, com o usuário guiando a trajetória e a IA fornecendo assistência qualificada.

A partir do entendimento das diferenças entre as IAs generativa e tradicional e onde e como seus controles devem ser posicionados, recomendamos uma adaptação das práticas de implementação dos princípios da IA responsável para *GenAI*, utilizando como referência a plataforma Azure da Microsoft e seus recursos. Para mais detalhes sobre como abordar os princípios da IA responsável sob a óptica da IA generativa, ver Apêndice I.

"... Isso coloca o usuário no papel de piloto em comando, que deve revisar e validar as informações antes de decidir sobre sua aplicação. Essa abordagem assegura que o usuário permaneça como a autoridade final, com pleno controle sobre a adoção ou rejeição do conteúdo gerado. "



Seção 3.

3 Implementação de *GenAI* Responsável em Sete Passos

3 Implementação de IA Generativa Responsável em Sete Passos

Implementar IA Generativa de maneira responsável e eficaz em um ambiente empresarial requer uma abordagem estruturada e bem definida. Um *framework* de implementação de IA Generativa responsável deve:

- **Garantir o uso ético da tecnologia** por meio da aplicação dos seis princípios de IA responsável em todas as etapas do ciclo de vida das soluções e aplicações de *GenAI*. Isso inclui estabelecer diretrizes e guias de boas práticas de desenvolvimento, adoção e governança da tecnologia que respeitem tais princípios.
- **Gerenciar riscos**: um *framework* estruturado deve conter processos de identificação e mitigação dos riscos, mecanismos de avaliação de seu grau de severidade e controles de prevenção e conformidade com os seis princípios de IA generativa responsável.
- **Alinhar o uso da tecnologia com os objetivos de negócio.** Esse alinhamento maximiza o retorno sobre o investimento, pois o propósito da tecnologia e as métricas de resultado para monitoramento são definidos pelos times de negócio.
- **Manter a conformidade regulatória** por meio de políticas e processos de avaliação e validação das soluções e aplicações de IA Generativa que averiguem sua conformidade com os requisitos legais e regulamentais.
- **Promover a adesão** das soluções de IA Generativa por meio dos programas de engajamento e treinamento dos usuários de *GenAI*, para que eles tenham conhecimento das melhores práticas de uso, do funcionamento e das limitações da tecnologia. Dessa forma, os usuários se sentirão estimulados e terão confiança e segurança na utilização das aplicações de IA Generativa.
- **Assegurar a transparência** das aplicações de *GenAI*, promovendo auditorias regulares que garantam que as soluções de IA Generativa mantenham seu desempenho dentro do esperado.

Antes de iniciar a jornada de implementação de IA Generativa responsável, é preciso definir a estratégia de adoção da tecnologia, pois ela influenciará no *roadmap* de implementação. As estratégias tipicamente mais aplicadas são a vertical e a horizontal.

A estratégia vertical de implementação de Inteligência Artificial (IA) direciona os investimentos para o desenvolvimento de soluções personalizadas, visando atender às demandas específicas de determinados nichos de mercado ou departamentos dentro de uma organização, onde soluções pré-existentes podem apresentar limitações. Um escritório de advocacia, por exemplo, poderia utilizar essa estratégia para desenvolver uma aplicação de IA generativa que automatizasse a análise de documentos legais, agilizando o processo e reduzindo custos. O serviço ***Azure OpenAI*** se destaca como uma opção viável nesse cenário, oferecendo flexibilidade e recursos para a construção de aplicações personalizadas de IA Generativa.

A estratégia horizontal, por outro lado, possui uma estrutura de decisão descentralizada, com o orçamento distribuído por diversas equipes e iniciativas, abrangendo o maior número possível de usuários. Essa estratégia é adotada para democratizar tecnologias, atingir escalabilidade rapidamente e gerar impacto em toda a organização. Um exemplo é a adoção do ***Microsoft Copilot***, integrado em ferramentas como o ***Microsoft 365***, para melhorar a produtividade de muitos usuários.

Os copilotos integrados aos produtos Microsoft, como o ***Microsoft 365 Copilot***,

proporcionam agilidade e escalabilidade em cenários que demandam soluções abrangentes e eficientes, sendo uma alternativa estratégica para a adoção horizontal da IA Generativa.

Uma vez escolhida a estratégia de adoção da tecnologia, podemos construir o *roadmap* de implementação de IA Generativa responsável. Independentemente da estratégia de adoção, ambos *roadmaps* devem garantir o uso ético e seguro da tecnologia. A diferença é que, enquanto o *roadmap* da estratégia vertical focará seus esforços iniciais em construir mecanismos e processos que insiram os princípios de IA Generativa responsável no ciclo de desenvolvimento das aplicações de *GenAI*, o *roadmap* da estratégia horizontal priorizará primeiro a construção de uma estrutura de governança e segurança para a utilização de aplicações de IA Generativa em escala por toda a organização.

Guia de Implementação de IA Generativa Responsável para Estratégia Vertical

Acreditamos firmemente que a IA Generativa tem potencial para criar ondas de impacto positivo em todo o globo, mas para isso é preciso ter uma estrutura que proporcione o desenvolvimento das aplicações com propósito e ética em um ambiente seguro. Para alcançar esse objetivo, criamos um guia de implementação de IA Generativa responsável em sete passos:

1. **Estabelecer uma estrutura ética** – essa estrutura tem o propósito de guiar a criação de práticas, mecanismos e políticas que garantam a aplicação de princípios éticos e regulatórios em todo o ciclo de vida da IA Generativa. Antecipando essa necessidade, a Microsoft, na qualidade de fornecedora da tecnologia, criou um *framework* de IA responsável baseado nos princípios da equidade, confiabilidade, privacidade e segurança, inclusão,

Guia de Implementação de *GenAI* Responsável para Estratégia Vertical

1. **Estabelecer uma estrutura ética** - desenvolver uma estrutura ética personalizada, com base em princípios empresariais e dos órgãos de regulação (6 princípios da IA Generativa responsável).
2. **Criar visão multifuncional** - montar uma equipe multidisciplinar para desenvolver modelos e aplicações de *GenAI* alinhados com os objetivos estratégicos e os valores éticos.
3. **Inventariar os ativos de *GenAI*** - criar um inventário de soluções de IA para catalogar e rastrear ativos de IA de forma contínua.
4. **Definir governança de *GenAI*** – definir papéis e responsabilidades ao longo do ciclo de vida da IA para gerenciar e governar seus riscos.
5. **Monitorar continuamente** - auditar regularmente os sistemas de IA Generativa quanto ao desempenho e conformidade ética.
6. **Estabelecer estrutura de segurança da informação** – criar uma diretriz padrão para gerenciar operações de segurança e definir gestão de identidade e acessos.
7. **Promover conscientização e treinamento** - melhorar o entendimento em toda a organização sobre o funcionamento, uso, limitações e ética da IA, garantindo que a liderança e o *staff* estejam totalmente informados.

transparência e responsabilidade para guiar os desenvolvedores de IA na adoção de práticas responsáveis. Além disso, no portfólio de produtos Microsoft, estão disponíveis uma série de funcionalidades que auxiliam na aplicação desses princípios durante o desenvolvimento das aplicações utilizando IA Generativa (ver **Apêndice I**).

2. **Criar uma visão multifuncional** – é imprescindível que os times de tecnologia e negócio trabalhem lado a lado na elaboração de iniciativas de IA Generativa, para que elas estejam bem alinhadas com os objetivos de negócios das organizações. Na etapa inicial de planejamento da construção de soluções de *GenAI*, o time de negócio precisa definir qual problema será endereçado, quais são os resultados esperados com a solução, quais métricas devem ser medidas para monitorar o retorno do investimento e quais processos ou fluxos de trabalho serão impactados pela solução. Durante as etapas de desenvolvimento, é necessária também a participação ativa do time de negócio para validar e aprovar o desenvolvimento das soluções, garantindo que a entrega esteja em conformidade com o que foi desenhado. Essa sinergia entre os dois times é peça chave para que a tecnologia esteja contribuindo para o sucesso do negócio.

3. **Inventariar os ativos de *GenAI*** – o inventário de modelos de IA Generativa é destinado a capturar e rastrear amplamente os ativos de *GenAI* existentes pela organização, mantendo-se atualizado sobre as novas aplicações de IA Generativa e mudanças nas existentes. O inventário das aplicações de IA auxilia e facilita o processo de auditoria dos modelos, uma vez que nele são registradas informações dos atributos da modelagem (nome do modelo, breve descrição, base de dados utilizadas, variáveis do modelo, tipo de algoritmo utilizado, versão do modelo etc.) e de gestão (nome dos responsáveis por desenvolver, aprovar, implementar e monitorar o modelo, e respectivas datas dos eventos). Em caso de situações adversas com alguma aplicação de *GenAI*, é possível identificar rapidamente o responsável por endereçar o problema. No **Apêndice IV**, há um exemplo de estrutura de inventário para soluções de IA Generativa. A Microsoft dispõe de ferramenta que auxilia na criação e gerenciamento de inventários de IA, incluindo ***Azure SQL Database*** e ***Azure Blob Storage*** para armazenamento de dados estruturados e não estruturados, além de ***Power BI*** para visualização e análise.

4. **Definir governança de *GenAI*** – a governança da IA Generativa deve criar uma estrutura baseada em processos, pessoas e tecnologias capazes de estabelecer padrões éticos para gerenciar todo o ciclo de vida da construção das soluções e mitigar os riscos inerentes ao uso da tecnologia. Recomendamos que a governança tenha duas linhas de atuação contra riscos. A primeira linha de defesa tem a principal responsabilidade de administrar os riscos associados às atividades de gerenciamento do desenvolvimento das soluções de IA Generativa. Isso inclui estabelecer os responsáveis por definir o propósito de negócio da solução, desenvolver os modelos, monitorar continuamente seu desempenho e gerenciar solicitações de ajustes nos modelos. Já a segunda linha de defesa permite a identificação de riscos emergentes na execução dos modelos em operação. Isso inclui classificar o grau de risco das soluções, garantir a aplicação de controles de qualidade e segurança das aplicações, monitorar seu desempenho e realizar revisões periódicas em função do seu respectivo grau de risco do modelo. Como exemplo, a **Seção 4** detalha uma estrutura de governança responsável de *GenAI* e os artefatos e soluções Microsoft para auxiliar a criação e gerenciamento dessa estrutura.

5. Monitorar continuamente as aplicações de *GenAI* - os modelos de IA Generativa tendem a ter maior complexidade, maior consumo e dependência de dados, menor grau de *explicabilidade* e estabilidade do que os modelos tradicionais de Inteligência Artificial. Como resultado, há a necessidade de se ter um ambiente centralizado para realizar validação e automatizar o fluxo de monitoramento das aplicações, a fim de detectar qualquer mudança de comportamento no fluxo de trabalho dos modelos, falhas de segurança, queda de performance, ou algum *feedback* negativo dos usuários. O *Azure AI Platform*[9] fornece uma ferramenta de avaliação personalizada que analisa automaticamente o desempenho do fluxo de *prompt* em várias métricas específicas para *GenAI*, como fundamentação (adesão a informações factuais), relevância e coerência. Os desenvolvedores podem estabelecer alertas quando as métricas decaírem abaixo de um certo valor ou ainda aproveitar as avaliações em lote de conjuntos de dados maiores para uma análise mais aprofundada. No Apêndice VII, encontra-se uma sugestão de *framework* de monitoramento contínuo que pode ser construído com os produtos Microsoft.

6. Estabelecer estrutura de segurança da informação - uma das grandes preocupações com a adoção da IA Generativa nas empresas diz respeito à segurança da informação. É preciso implementar uma política de segurança da informação eficaz, de forma que se possa usufruir dos benefícios de eficiência que a IA Generativa proporciona sem oferecer riscos de vazamento de informações estratégicas e sigilosas. Controles de segurança sobre as informações sensíveis das empresas devem ser implementados para garantir que as respostas produzidas pela IA Generativa sejam disponibilizadas apenas àquelas pessoas que possuem permissão de acessá-las, e que ela só consuma dados que o usuário está autorizado a acessar. Nesse contexto, existem hoje ferramentas que possibilitam a implementação de rótulos de confidencialidade sobre documentos, *e-mails* e arquivos, classificando-os de acordo com suas categorias de níveis de confidencialidade, definidas pela própria empresa. Tais rótulos permitem que o processo de seleção das informações executado pela IA Generativa utilize somente documentos de determinados níveis de acesso, por exemplo, ao fazer uma pergunta a um *chatBot* desenvolvido pela empresa, ele só terá acesso a consultar documentos classificados como públicos. Ferramentas, como o *Microsoft Purview*[10], podem compor a gestão e classificação de dados em informações acessadas pelo *Microsoft Copilot*, por exemplo.

Promover a conscientização e o treinamento – a amplitude do impacto e benefícios gerados pela IA Generativa é diretamente proporcional à sua adoção pelos usuários. Para acelerar a adoção da tecnologia, é preciso que os funcionários das empresas se sintam seguros, motivados e empoderados a utilizar as aplicações de *GenAI* em suas rotinas de trabalho. Os funcionários precisam de uma visão clara não apenas para inspirar, mas também para compreender o objetivo de se adotar determinada tecnologia. Ao definir e comunicar uma visão clara para a estratégia de IA Generativa, a liderança das empresas promovem um sentimento de pertencimento e compromisso entre seus colaboradores, elevando os níveis de aceitação e envolvimento, e acelerando, consequentemente, a adoção da tecnologia. A formação de times multidisciplinares ajuda a promover a colaboração, eliminando silos e permitindo conhecimentos diversos para acelerar as iniciativas *GenAI*. Os líderes precisam abraçar, demonstrar e promover os comportamentos certos para a adoção da tecnologia. A experimentação, a tolerância a riscos e a segurança psicológica são exemplos de traços culturais que precisam estar presentes nessa transformação. É

preciso treinar e fornecer às equipes os recursos, o conhecimento e as ferramentas necessárias para integrar perfeitamente a *GenAI* em seus fluxos de trabalho de forma segura, respeitando as boas práticas e compreendendo suas limitações e restrições. Com o objetivo de acelerar o desenvolvimento das habilidades de *GenAI*, a Microsoft disponibiliza uma série de treinamentos gratuitos de capacitação em suas aplicações e ferramentas[11].

Guia de Implementação de IA Generativa Responsável para Estratégia Horizontal

Na estratégia horizontal de adoção da tecnologia, a prioridade é criar um ambiente seguro para o uso em escala das aplicações de *GenAI* pelo usuário, sem que haja nenhuma exposição dos dados da organização e nenhum impacto operacional, reputacional e ético provocado pelo uso indevido da tecnologia. Para acelerar o processo de democratização da tecnologia, adaptou-se o guia de implementação de IA Generativa responsável para focar nos seguintes passos:

1. **Estabelecer uma estrutura ética** – a estrutura ética e seu propósito são os mesmos discutidos anteriormente (ver **Seção 2 e Apêndice I**). A diferença é que agora o *framework* de IA responsável será aplicado para treinar e guiar os usuários de aplicações de *GenAI* a utilizar a tecnologia de forma responsável. A Microsoft disponibiliza uma série de treinamentos de seus produtos de IA Generativa, como Microsoft Copilot[12], para orientar seus clientes sobre como fazer um melhor uso do produto, além de esclarecer dúvidas e compartilhar informações sobre as especificações técnicas.

2. **Definir governança de GenAI** – a capacidade de democratizar os recursos de *GenAI* em toda a organização é fundamental para se alcançar a maturidade em *GenAI*. A colaboração entre as lideranças da empresa é um fator crítico. A democratização da inteligência artificial

Guia de Implementação de *GenAI* Responsável para Estratégia Horizontal

1. **Estabelecer uma estrutura ética** - desenvolver uma estrutura ética personalizada com base em princípios empresariais e dos órgãos de regulação (6 princípios da IA Generativa responsável).
2. **Definir governança de IA** – definir uma governança capaz de elevar a experiência humana, ajudando a organização a avaliar e otimizar a adoção de *GenAI* por meio de um programa acelerador que se concentra no impacto, no risco, na eficiência operacional e no retorno do investimento.
3. **Estabelecer estrutura de segurança da informação** – criar uma diretriz padrão para gerenciar operações de segurança e definir gestão de identidade e acesso.
4. **Criar visão multifuncional** – montar uma equipe multidisciplinar para definir quais aplicações de *GenAI* serão democratizadas e quais benefícios esperados e como serão capturados, como parte da estratégia de engajamento para acelerar a adoção da tecnologia em escala.
5. **Promover engajamento e treinamento** – disseminar o engajamento e o entendimento sobre o funcionamento, boas práticas de uso e limitações das aplicações, a fim de acelerar a adoção segura da ferramenta.
6. **Monitorar e melhorar continuamente** – monitorar continuamente o uso das aplicações de *GenAI*, para capturar seu benefício e coletar o *feedback* dos usuários, a fim de adaptar a estratégia de engajamento dos usuários.
7. **Portfolio de aplicativos de IA** - criar um portfólio de soluções/aplicações de *GenAI* disponíveis para os usuários.

não deveria ser vista como uma responsabilidade exclusiva do *CIO* ou *CTO*. Portanto, a estrutura de governança voltada para a democratização de IA Generativa deveria fomentar e suportar esse compartilhamento de responsabilidade entre as diferentes áreas da empresa, ao mesmo tempo em que centraliza os papéis, processos, políticas e ações de gestão e mitigação dos riscos associados à tecnologia na TI. O time de TI ainda será o único responsável por homologar as soluções, definir os guias de boas práticas, os requisitos técnicos e limitação de uso para os times de negócio. À medida em que se distribui a responsabilidade da geração de inovação por meio da adoção de tecnologia, deve-se discutir também o modelo de financiamento, para que o orçamento não inviabilize a aceleração da adoção de *GenAI*. Em vez de a TI ser responsável por arcar sozinha com os custos das licenças das aplicações de *GenAI*, muitas empresas têm optado por dividir entre os centros de custos dos respectivos usuários. A **Seção 4** detalha a estrutura da governança responsável de *GenAI*, os artefatos e soluções Microsoft para auxiliar a criação e gerenciamento dessa estrutura.

3. **Estabelecer a estrutura de segurança da informação** – como mencionado anteriormente, uma das grandes preocupações com a adoção da IA Generativa nas empresas está relacionada com a segurança da informação. É preciso ter uma política de segurança da informação robusta capaz de identificar e endereçar de forma proativa os riscos conhecidos ou não, uma vez que, em uma adoção em larga escala da tecnologia, as consequências de um incidente podem escalonar muito rápido. Para apoiar seus clientes na gestão de risco, a Microsoft desenvolveu um *assessment* capaz de mapear a quais riscos os clientes estão expostos e orientá-los de forma eficiente a configurar seus controles de segurança[13,14].

Criar uma visão multidisciplinar – é importante definir o desenho de adoção que permitirá às organizações construir, lançar, apoiar e expandir o uso da IA Generativa por diversas funções. Essa etapa inclui atividades como: criar um consenso entre os tomadores de decisão sobre os critérios para o sucesso do programa; estabelecer métricas para a concretização dos benefícios; identificar as personas e os casos de uso mais relevantes para a captura de resultado; e elaborar estratégias de comunicação e programas de treinamento personalizados para multiplicar a adoção da tecnologia. Muito provável que já existam funcionários que entendam e usem aplicativos de *GenAI* diariamente. Por que não usá-los como embaixadores da tecnologia e criar sessões para que compartilhem sua experiência e ideias de uso para toda empresa? É possível que, dessas sessões, saiam potenciais oportunidades de melhorias operacionais e ganho de eficiência aplicável a diferentes áreas da organização.

5. **Promover engajamento e treinamento** - quais mudanças serão necessárias na organização para desbloquear os benefícios da *GenAI*? O que as pessoas precisam fazer de diferente para aceitar, adotar e escalar o uso de *GenAI*, a fim de colher mais rápido seu ROI? Entender o que engaja e motiva as pessoas a utilizarem a IA Generativa é o coração de um programa de gestão da mudança capaz de catalisar o uso de novas soluções. Reconhecendo que diferentes pessoas têm diferentes necessidades de treinamento, comunicação e engajamento, sugere-se uma abordagem além da segmentação tradicional. Para atender a essas diferenças, é essencial avaliar e agrupar os indivíduos de acordo com: suas necessidades de uso por funções; pontos de dor, medos e resistências; conhecimentos prévios em IA e outras tecnologias; motivações, desejos e interesses; preferências digitais e de trabalho; comportamentos e crenças em

relação às novas tecnologias. Uma vez identificados os diferentes perfis de usuários existentes na empresa, aplica-se esse conhecimento para adotar estratégias de engajamento mais eficazes, personalizar o plano de comunicação e treinamentos. Para fomentar a utilização completa das soluções de IA Generativa, é importante desenvolver estratégias de adoção e engajamento que vão além do uso básico de ferramentas tradicionais. Isso inclui: **a.** incentivar os usuários a explorar e utilizar todas as funcionalidades das soluções de IA Generativa, de acordo com suas necessidades específicas e a área da organização em que atuam; **b.** criar campanhas e iniciativas que mantenham o interesse e o engajamento dos usuários ao longo do tempo; e **c.** engajar os usuários para formar comunidades de uso. As comunidades são fundamentais para o sucesso contínuo da implementação das soluções de IA Generativa, pois elas promovem o compartilhamento de melhores práticas e experiências entre os usuários; identificam e apoiam influenciadores e promotores internos que possam motivar outros a adotar e utilizar a nova tecnologia; e coletam *feedbacks* regularmente para identificar áreas de melhoria e ajustar a implementação conforme necessário. Para que os treinamentos sejam mais eficazes, os conteúdos e módulos devem ser adaptados a cada nível de necessidade de uso, interesse e conhecimento prévio. Isso inclui: oferecer treinamentos que vão além do básico, explorando todas as potencialidades das soluções de IA Generativa para resolver problemas específicos e atender às necessidades e interesses dos usuários; e utilizar diferentes métodos de ensino, como *workshops* práticos, *e-learning*, e tutoriais interativos, para atender às diversas preferências de aprendizagem. Antecipando essa necessidade, a Microsoft já disponibilizou mais de 100 treinamentos[15] de ***Copilot*** para diferentes níveis de experiência e propósitos de uso, possibilitando ao usuário escolher aquele que melhor atende à sua necessidade.

6. **Monitorar e melhorar continuamente** – uma vez definidas as métricas para monitoramento dos resultados e o plano de engajamento dos usuários em prática, é necessário começar a capturar os primeiros resultados e *feedbacks*. O acompanhamento desses dados permitirá identificar se o plano de engajamento está gerando ou não os benefícios esperados e diagnosticar as razões pelas quais não se estão concretizando os resultados. A partir desse diagnóstico, é possível adaptar a estratégia de adoção e engajamento para melhorar a experiência do usuário. É natural que a abordagem de engajamento esteja em constante mudança, principalmente à medida que o usuário vai amadurecendo seus conhecimentos e proficiência com a tecnologia. O monitoramento contínuo dos dados e *feedbacks* permite o aperfeiçoamento da abordagem de engajamento conforme a evolução da experiência dos usuários.

7. **Criar portfólio de aplicativos de IA** – o programa de engajamento, apesar de ser específico para o nicho de soluções de *GenAI*, se bem sucedido, despertará o apetite dos usuários por novas tecnologias. Isso porque eles já terão rompido a barreira da resistência tecnológica. Para incentivar esse público na geração de inovação e eficiência operacional por meio da adoção de tecnologia, é fundamental manter um portfólio de aplicações de IA de fácil acesso. Assim, os usuários podem checar quais outras tecnologias estão disponíveis para uso e seus requisitos (exemplo: tecnologias de automação tipo *low code/no code*).

No **Apêndice VIII**, há uma breve exposição das principais ferramentas de IA e dados oferecidos pela Microsoft para apoiar sua jornada de Inteligência Artificial na nuvem.

4

Seção 4.
Governança Responsável de *IA Generativa*

4 Uso Seguro e Responsável da IA Generativa

No cenário atual impulsionado por dados, as organizações se deparam com desafios crescentes em relação à governança de dados, segurança, privacidade e conformidade. A proliferação de dados pessoais, o panorama regulatório em constante expansão e a crescente sofisticação das ameaças cibernéticas exigem que as empresas adotem uma abordagem abrangente e proativa para gerenciar seus dados de forma eficaz.

A Lei Geral de Proteção de Dados Pessoais (LGPD), por exemplo, introduziu requisitos para a coleta, uso e armazenamento de dados pessoais no Brasil. O não cumprimento dessas exigências pode resultar em multas e danos à reputação da organização que não as cumprir.

Aplicações de Inteligência Artificial apresentam aspectos adicionais de privacidade a serem considerados. Portanto, é preciso garantir que essas aplicações sejam empregadas de forma responsável e ética, e que os dados sejam anonimizados quando necessário.

A conformidade com as regulamentações específicas da indústria é outro fator crucial a ser considerado. As empresas que operam em setores como saúde, finanças e serviços governamentais precisam estar cientes das regras e regulamentações específicas que se aplicam a eles. O não cumprimento pode levar a sanções e outras penalidades. Dessa forma, uma governança robusta de dados pode auxiliar as organizações a:

- **Aprimorar a tomada de decisões:** com acesso a dados precisos e confiáveis, as empresas podem tomar decisões mais informadas que impulsionam o crescimento e a eficiência.

- **Melhorar a experiência do cliente:** os clientes percebem valor em produtos e serviços personalizados e recomendações assertivamente direcionadas ao seu tipo de perfil.

- **Reduzir custos:** violações de dados e outras falhas de segurança podem ser extremamente onerosas. Uma boa governança de dados pode ajudar a prevenir esses incidentes e reduzir os custos associados a eles.

- **Geração de novas linhas de negócio:** a IA Generativa e outras tecnologias de análise de dados podem impulsionar a inovação e o crescimento. Ao garantir que os dados sejam usados de forma responsável e ética, as empresas podem aproveitar todo o potencial dessas tecnologias.

Os controles adequados na adoção da Inteligência Artificial podem evitar a exposição das organizações a potenciais falhas e incidentes.

O ciclo de vida de um modelo ou aplicação de *GenAI* traz riscos inerentes ao desenho da solução, aos dados, ao algoritmo, à performance, à infraestrutura tecnológica, à conduta e ética e aos processos operacionais onde a aplicação será inserida.

A gestão desses riscos no ciclo de vida da IA Generativa é a chave para permitir uma *GenAI* confiável, tal como promovido pelos reguladores, e para isso é necessária a construção de uma governança estruturada em três pilares:

- **Pessoas**: papéis e responsabilidades claramente definidos são cruciais. Por exemplo, cientistas de dados e

desenvolvedores focam no desenvolvimento de modelos, enquanto órgãos de governança supervisionam a conformidade e a gestão de riscos. Essa divisão de trabalho garante que cada aspecto do desenvolvimento e implantação da IA seja tratado por especialistas nas respectivas áreas.

- Processos: o ciclo de vida dos modelos generativos incluem etapas como desenho, desenvolvimento, validação, implementação, monitoramento e descomissionamento. Cada estágio possui diretrizes detalhadas para garantir consistência e qualidade. Por exemplo, durante a fase de validação do modelo, é obrigatória uma revisão independente para garantir objetividade e minuciosidade.

- Tecnologia: infraestrutura e ferramentas são necessárias para suportar a governança. Isso inclui soluções seguras de armazenamento de dados, plataformas de análise e sistemas de monitoramento. Essas tecnologias facilitam o desenvolvimento, a implantação e a supervisão de modelos.

Ao estabelecer os pilares fundamentais de governança, processos e tecnologia, as empresas podem garantir que a IA Generativa seja utilizada para impulsionar a inovação, otimizar operações e alcançar seus objetivos de forma responsável e ética, construindo um futuro impulsionado pela Inteligência Artificial.

Para promover a inovação no campo dinâmico da *GenAI*, ao mesmo tempo em que se mitigam os riscos inerentes, é essencial uma abordagem estratégica que priorize ambos os aspectos. A promoção da inovação deve andar de mãos dadas com práticas de gestão de riscos para assegurar o progresso sustentável.

Um modelo de governança multifacetado, inspirado no *framework* do Instituto Nacional de Padrões e Tecnologia (NIST), categoriza os projetos com práticas de cibersegurança e requisitos de recursos baseados no ciclo de vida de IA. Essa estrutura fornece um conjunto uniforme de diretrizes e padrões para as organizações — atuando como uma ferramenta de gerenciamento de segurança de nível que ajuda a avaliar e mitigar o risco de aplicações de Inteligência Artificial em toda a organização.

O *framework* de governança (ver Apêndice V) enfatiza a importância de entender o contexto dos negócios, os recursos que suportam funções críticas e os riscos de segurança relacionados.

As atividades essenciais incluem:

- Identificar ativos de IA para estabelecer a base de um programa de gerenciamento de ativos.

- Implementar melhorias incorporando as lições aprendidas com a adoção de aplicações de modelos.

- Gerenciar as comunicações com as partes interessadas internas e externas antes, durante e após a implantação de uma solução de IA.

Recomenda-se uma estrutura de governança de IA dividida em duas linhas de defesa:

- Primeira linha de atuação: abrange e gerencia os riscos associados ao desenvolvimento, implementação e monitoramento de modelos. Por exemplo, cientistas de dados e desenvolvedores são responsáveis por garantir que seus modelos cumpram as diretrizes e padrões internos.

- Segunda linha de atuação: identifica riscos emergentes, realiza revisões independentes de modelos e eficácia contínuas. Uma equipe de validação independente da equipe de desenvolvimento

garante conformidade e alinhamento com os objetivos de negócio sem vieses.

Essa estrutura de governança garante uma independência entre a equipe que desenvolve os modelos e aplicações de *GenAI* e a equipe responsável por sustentar essas mesmas aplicações em operação.

A definição dos papéis e responsabilidades dentro da estrutura de gestão de riscos de IA é fundamental para garantir que todos os aspectos de risco sejam adequadamente gerenciados. Essa estrutura organizacional promove uma abordagem colaborativa e coordenada, assegurando que a organização esteja preparada para enfrentar os desafios e aproveitar as oportunidades apresentadas pela IA Generativa. O Apêndice V mostra nossa recomendação de divisão de papéis e responsabilidades divididas nessas duas linhas de defesa.

O *framework* de governança de IA visa criar conscientização e alinhamento sobre componentes-chave entre as equipes de análise, tecnologia, risco, ética e conformidade. Ao estabelecer diretrizes claras, papéis e procedimentos, o *framework* garante que as soluções de IA Generativa sejam desenvolvidas e implantadas de maneira responsável. Essa abordagem não apenas aumenta a eficácia das soluções de IA, mas também mitiga potenciais riscos, promovendo uma cultura de IA confiável dentro da organização.

A gestão de riscos de IA abrange todas as fases, desde a pré até a pós-implantação. Cada fase envolve atividades e participantes específicos, garantindo o desenvolvimento, a implementação e o monitoramento eficazes da IA. A conformidade com os princípios éticos e regulamentares é assegurada em todas as etapas.

A fase de pré-implantação de um projeto de IA Generativa é crucial para o sucesso e segurança da solução, pois detalha o problema a ser resolvido, escopo do projeto (limites e expectativas), métricas de sucesso a serem medidas, infraestrutura tecnológica (*hardware*, *software*, nuvem) e requisitos de dados (tipo, volume, qualidade, acesso).

O desenvolvimento e a implantação de soluções de IA Generativa demandam um processo estruturado, abarcando desde a aquisição e pré-processamento de dados até a homologação e monitoramento contínuo do modelo. Essa etapa requer uma validação rigorosa, pois aqui podem nascer os vieses e vícios do modelo.

Após a conclusão das fases de testes e homologação, a implantação bem-sucedida de uma solução de IA Generativa requer etapas adicionais para garantir sua eficácia, longevidade e gestão responsável. Um artefato muito importante para auxiliar na classificação e plano de mitigação de riscos é o *framework* de classificação de risco do modelo.

O *framework* de classificação de risco para modelos de IA é baseado em uma matriz de decisão que sintetiza respostas para uma combinação de fatores qualitativos e quantitativos, que levam em consideração a complexidade, materialidade, impacto estratégico e reputacional, confiabilidade e considerações regulatórias (ver Apêndice VI para uma sugestão de matriz de decisão). Sua importância se deve ao fato de que é ele o responsável por definir a frequência e profundidade da validação, monitoramento e revisão dos modelos de IA.

Outro artefato muito importante para a identificação e mitigação dos riscos é o *framework* de monitoramento contínuo dos modelos de *GenAI*. Recomenda-se implementar um sistema de monitoramento contínuo para rastrear o desempenho do modelo em tempo real e estabelecer mecanismos de alerta para identificar anomalias e desvios do comportamento

esperado, permitindo a intervenção rápida para correção e otimização. O Apêndice VII mostra uma recomendação de *framework* de monitoramento contínuo para *GenAI* que pode ser construído com as ferramentas Microsoft.

O monitoramento contínuo das aplicações facilitará a realização de auditorias regulares para avaliar a necessidade de ajustes ou reciclagem com novos dados, garantindo a manutenção da performance e a mitigação de vieses emergentes. Para isso, é essencial estabelecer uma estrutura de governança transparente para supervisionar o desenvolvimento, implantação e uso do modelo de IA Generativa.

Devem-se definir papéis e responsabilidades claros para todas as partes envolvidas, incluindo desenvolvedores, proprietários de dados, usuários finais e especialistas em ética e conformidade. Além disso, é fundamental criar mecanismos de comunicação e colaboração entre as partes interessadas, a fim de garantir a tomada de decisões informadas e alinhadas com os objetivos organizacionais e os princípios éticos.

A gestão de riscos em projetos de IA Generativa não se limita a uma etapa isolada do processo, mas permeia todo o ciclo de vida da tecnologia, demandando atenção constante e adaptabilidade. É importante manter a flexibilidade no plano e nas estratégias de mitigação de riscos, adaptando-os conforme o desenvolvimento do projeto. Dessa forma, garante-se que os projetos de IA continuem alinhados com as melhores práticas de governança e segurança, maximizando seu potencial enquanto minimizam os riscos associados.

Governança para PMEs com a Microsoft: Protegendo o Negócio e Impulsionando o Crescimento

A governança de TI é importante para o sucesso de qualquer empresa, especialmente para pequenas e médias empresas (PMEs) que precisam navegar em um cenário digital em constante mudança. Com recursos limitados e objetivos ambiciosos, as PMEs enfrentam desafios únicos para garantir que seus sistemas de TI estejam alinhados com os objetivos de negócios, sejam seguros e atendam aos requisitos de conformidade.

Para atender a essas demandas, está disponível no mercado um portfólio abrangente de produtos e serviços que visam fortalecer a segurança e a conformidade de dados e infraestrutura (ver Apêndice VIII).

Optar por um provedor de nuvem significa ter acesso a uma plataforma escalável que suporta o crescimento do negócio. Suas soluções de IA generativa, por exemplo, não apenas otimizam processos e melhoram a tomada de decisões, mas também impulsionam a inovação, permitindo que a empresa se mantenha competitiva em um mercado em constante evolução.

Certamente, a gestão de custos de infraestrutura é um tema relevante na implementação de soluções em nuvem, e já existem ferramentas para auxiliar nessa tarefa. O primeiro passo dessa estratégia é planejar o consumo da nuvem, permitindo que as empresas tenham controle de seus gastos. Utilizando ferramentas como o *Azure Cost Management*, é possível calcular cenários de consumo, acompanhar gastos em tempo real e configurar alertas para evitar surpresas no fim do mês. É possível também otimizar recursos na nuvem, pagando-se apenas pelo que realmente está sendo utilizado, reduzindo desperdícios.

A segurança e conformidade são essenciais para proteger os dados e cumprir as regulamentações. Ferramentas como o *Microsoft 365 Compliance Manager* ajudam a garantir que dados estejam protegidos e em conformidade com regulamentações como a LGPD. Além disso, com soluções como o *Entra ID*, pode-se controlar quem tem acesso às informações sensíveis da empresa, aumentando a segurança e reduzindo o risco de violações. O *Azure Security Center* também monitora o ambiente de TI e oferece alertas sobre possíveis ameaças.

A gestão eficaz dos dados é outra área crítica para a utilização de aplicações de *GenAI*. Por meio de implementação adequada utilizando o *Microsoft Purview*, as PMEs obtêm visibilidade completa dos seus dados, mitigando os riscos. Essa solução facilita a organização, catalogação, mapeamento de linhagem, classificação e governança dos dados, simplificando o processo de conformidade com as normas regulatórias aplicáveis. Além disso, é possível encontrar treinamentos[11] *on-line* que ajudam a implementar e utilizar esses serviços.

Para tangibilizar de forma objetiva como empresas PMEs podem alcançar escalabilidade de suas aplicações de *GenAI* de forma segura, os Apêndices IV – VIII trazem guias e exemplos práticos de artefatos para implementar a governança de *GenAI*. O Apêndice IV traz um exemplo de inventário de modelos de Inteligência Artificial; o Apêndice V exemplifica um modelo de matriz RACI de governança de *GenAI*; o Apêndice VI propõe um modelo de classificação de risco das aplicações; o Apêndice VII ilustra um *framework* de monitoramento contínuo dos modelos de IA; e o Apêndice VIII elenca as ferramentas disponíveis da Microsoft para serem utilizadas na governança dessa tecnologia.

Apêndice I – *Os Princípios da IA Generativa Responsável*

Para operacionalizar os princípios da IA responsável, a Microsoft desenvolveu o Padrão de IA Responsável[16], que traduz princípios éticos em diretrizes práticas para equipes de desenvolvimento de produtos. Componentes desse padrão incluem avaliações de impacto, supervisão humana, governança de dados e avaliação contínua. Essas práticas garantem que os sistemas de IA sejam desenvolvidos e monitorados de maneira responsável, com aprimoramentos baseados em *feedback* e avaliações contínuas.

A Microsoft adere aos princípios de equidade, confiabilidade, segurança e privacidade, inclusão, transparência e responsabilidade *by-design* em seus sistemas de IA. A seguir, seguem exemplos de como utilizar os recursos e produtos da Microsoft para aplicar os princípios da IA responsável no desenvolvimento e utilização das aplicações de *GenAI*.

Equidade

Garantir a busca por justiça algorítmica em modelos de IA Generativa é crucial para evitar a perpetuação de vieses presentes nos dados de treinamento. Ferramentas como o painel de IA Responsável do *Azure Machine Learning* auxiliam na identificação e mitigação de possíveis vieses, enquanto a análise de erros ajuda a identificar padrões discriminatórios nos resultados. É importante que aplicações de IA, como os de recomendação de empregos, não perpetuem preconceitos sociais, como a discriminação baseada em gênero ou etnia.

Confiabilidade

O Azure oferece um conjunto abrangente de ferramentas para garantir a confiabilidade da IA Generativa. O *Azure Content Moderator* filtra conteúdo prejudicial, já o *Azure OpenAI*

Service melhora a qualidade e precisão, enquanto o *Azure Functions* permite pós-processamento. Testes de segurança e recursos de recuperação de desastres podem ser realizados com *Azure Backup* e *Azure Site Recovery*, garantindo a robustez e resiliência do sistema.

Privacidade e segurança

A segurança e privacidade são fundamentais na IA Generativa. Ferramentas como *Azure Content Moderator* e *Azure AI Content Safety* detectam e previnem conteúdo malicioso, enquanto o *Azure Security Center* e o *Azure Sentinel* protegem contra ataques. *Azure Confidential Computing, Azure Private Link* e *Azure Key Vault* garantem a privacidade dos dados, e o *Copilot para Microsoft 365* integra IA em aplicativos de produtividade com segurança.

Inclusão

A inclusão é essencial para que a IA Generativa beneficie a todos. Os serviços do *Azure AI Services*, como *Speech to Text* e *Text to Speech*, permitem a transcrição e geração de fala em diversos idiomas e formatos, tornando a IA acessível a pessoas com deficiências visuais ou auditivas. O *Azure Translator* facilita a comunicação multilíngue, quebrando barreiras e promovendo a diversidade. O *Microsoft Teams* possui legendas ao vivo e transcrição alimentadas por IA, garantindo que as reuniões sejam acessíveis para indivíduos com deficiências auditivas. O *Microsoft Seeing AI* é um aplicativo que utiliza IA para descrever o ambiente para usuários com deficiência visual, tornando as tarefas diárias mais acessíveis e inclusivas.

Transparência

A transparência é fundamental para construir confiança na IA Generativa. O *Azure OpenAI Service* oferece documentação detalhada sobre os modelos, incluindo informações sobre sua arquitetura, dados de treinamento e potenciais limitações. Essa transparência permite que os usuários compreendam como o modelo funciona e tomem decisões informadas sobre seu uso. *O painel de IA Responsável* fornece *insights* sobre o comportamento do modelo, facilitando a identificação de possíveis vieses e a tomada de medidas corretivas.

Responsabilidade

A responsabilidade é um compromisso para o uso da IA Generativa. O *Azure Policy* permite definir e aplicar regras de governança, garantindo que os modelos de IA sejam utilizados de forma ética e responsável. O *Azure Monitor* alerta sobre anomalias no comportamento do modelo, permitindo a intervenção em caso de problemas. Mecanismos de *feedback*

humano, como pesquisas de satisfação e canais de comunicação, permitem que os usuários reportem problemas e contribuam para o aprimoramento contínuo das aplicações.

Apêndice II – *AI Custommer Commitments*[17]

Em junho de 2023, a Microsoft anunciou o programa chamado *AI Customer Commitment*, delineando etapas para apoiar seus clientes em sua jornada responsável de IA. Nesse programa, a Microsoft:

- Cria o chamado *AI Assurance Program* para ajudar os clientes a garantir que os aplicativos de IA que eles estão implantando nas plataformas Microsoft atendem aos requisitos legais e regulamentares para a IA responsável. Esse programa inclui o apoio ao envolvimento com os órgãos reguladores, juntamente com a promessa de atestar como a Microsoft está implementando o *NIST AI Risk Management Framework*[18];

- Continua a interagir com o conselho dos clientes, ouvindo suas opiniões sobre como é possível fornecer tecnologias e ferramentas com IA mais relevantes e compatíveis com as demandas do mercado.

- Cria o programa *Responsible AI Partner* para o ecossistema de parceiros Microsoft, sendo que até a data de lançamento deste guia, onze parceiros já haviam aderido ao programa. Esses parceiros criaram práticas abrangentes para ajudar os clientes a avaliar, testar, adotar e comercializar soluções de IA[19].

- Anuncia e amplia o *Customer Copyright Commitment*[20], em que a Microsoft defenderá seus clientes comerciais que forem processados por terceiros por violação dos direitos autorais ao usarem o ***Azure OpenAI Service***, ***Microsoft Copilot*** ou os resultados gerados por IAs da Microsoft, em quaisquer sentenças adversas ou acordos, desde que o cliente cumpra condições básicas, como não tentar gerar conteúdo infrator e usar as proteções e filtros de conteúdo da Microsoft[21].

Nota: recomenda-se verificar as informações mais atualizadas nos *links* fornecidos, pois as tecnologias e os programas estão em constante evolução.

Apêndice III – Documentação de Transparência das Aplicações Microsoft de IA

Apêndice III

A partir de 2019, a Microsoft passou a publicar Notas de Transparência, uma documentação cobrindo tópicos de IA responsável para seus serviços de plataforma usados por clientes na construção de seus próprios aplicativos de IA.

As notas de transparência, obrigatórias para os serviços das plataformas Microsoft, fornecem informações detalhadas sobre capacidades, limitações e utilização adequada para promover integração e uso responsáveis. Alguns exemplos de notas de transparência incluem *Azure AI Vision API Face*[22], *Azure OpenAI Service*[23] e *Azure Document Intelligence*[24].

Em 2023, a Microsoft ampliou suas documentações de transparência para além das notas, exigindo que os serviços fora da plataforma *Azure*, como o *Microsoft Copilots*, publiquem perguntas frequentes (*FAQs*) de IA responsável e incluam avisos amigáveis sobre a experiência do produto para fornecer informações importantes. Por exemplo, o *Microsoft Copilot no Bing* fornece aos usuários documentação sobre IA responsável[25] e perguntas frequentes[26] que detalham os métodos de mapeamento, medição e gerenciamento de riscos da Microsoft. Além disso, no *Microsoft Copilot no Bing*, aparecem avisos informando aos usuários que eles estão interagindo com uma aplicação de IA, com citação sobre a origem do material para viabilizar a verificação das informações nas respostas. Outras comunicações importantes podem incluir isenções de responsabilidade sobre o potencial da IA de cometer erros ou produzir conteúdo inesperado.

Apêndice IV – Inventário de Modelos de IA

A utilização estratégica de sistemas de Inteligência Artificial e *Machine Learning* (IA/*ML*) beneficia diversas funções organizacionais. No entanto, gerenciar e supervisionar essa tecnologia exige uma compreensão de seu cenário de implantação. Para alcançar isso, as organizações devem priorizar a criação e manutenção de um inventário de soluções de IA/*ML*. Dentre os benefícios de um inventário, podemos citar:

- **Visibilidade:** um inventário centralizado fornece uma visão holística de todas as soluções de IA/*ML* implantadas em programas e projetos, promovendo a transparência e facilitando a tomada de decisões informadas.
- **Governança:** ao identificar o escopo e a escala do uso de IA/*ML*, as partes interessadas e a gerência obtêm uma supervisão aprimorada, permitindo-lhes estabelecer estruturas de governança eficazes e garantir um desenvolvimento e implantação responsáveis.
- **Gerenciamento proativo de riscos:** o inventário serve como uma ferramenta para identificar e mitigar proativamente os riscos potenciais associados às soluções de IA/*ML*, incluindo problemas de qualidade de dados, preocupações com a precisão do modelo, vieses, vulnerabilidades de segurança e desafios de conformidade regulatória.

Detalhes essenciais do inventário:

- **Propósito e uso do sistema:** descrição clara e concisa da função pretendida do sistema e como ele é utilizado dentro da organização.
- **Elementos de dados:** especificação detalhada dos elementos de dados que o sistema alavanca para gerar resultados ou tomar decisões.
- **Propriedade e responsabilidade:** identificação clara do proprietário do sistema (departamento, equipe etc.) para garantir responsabilidade e facilitar a comunicação.
- **Processo de desenvolvimento:** documentação do processo de desenvolvimento da aplicação, incluindo desenvolvimento interno/externo e as metodologias empregadas.
- **Datas-chave:** registro de datas críticas, como implementação inicial, atualizações subsequentes e revisões agendadas.
- **Avaliação de riscos:** avaliação dos riscos potenciais associados ao sistema, abrangendo qualidade de dados, viés algorítmico, vulnerabilidades de segurança e adesão regulatória.

Nome do atributo	Definição	Tipo de atributo
Data de aprovação do modelo de IA	Data em que a aprovação foi concedida (data de aprovação da versão)	Calendário, N/D, se não aplicável
Data de implementação do modelo de IA	Data de implementação inicial (primeira vez em que o modelo de IA foi implementado)	Calendário, N/D, se não aplicável
Data de aprovação da alteração do modelo	Data em que a alteração do modelo foi aprovada	Calendário, N/D, se não aplicável
Data de alteração do modelo de IA	Data em que o modelo foi atualizado/alterado pela última vez	Calendário, N/D, se não aplicável
Natureza da última atualização/alteração	Versão inicial / aprimoramento do modelo principal de IA / aprimoramento do modelo de IA menor / imaterial, se aplicável	*Drop-down*, N/D, se não aplicável
Data de descomissionamento planejada do modelo de IA	Data em que se espera que o modelo de IA seja desativado, se aplicável	Calendário, N/D, se não aplicável
Data de aprovação do descomissionamento do modelo IA	Data em que o descomissionamento do modelo foi aprovado	Calendário, N/D, se não aplicável
Data de descomissionamento do modelo de IA	Data em que o modelo foi desativado	Calendário, N/D, se não aplicável
Restrições de uso do modelo de IA	Restrições de uso do modelo de IA	Caixa de texto
Tipo do modelo de IA	Usado para capturar o tipo do Modelo de IA: *Machine Learning* (*ML*) / *Natural Language Processing* (*NLP*) / Computer Vision / Other (por favor, especifique)	Seleção múltipla
Nível de risco do modelo de IA	Alta / Média / Baixa	Gerado automaticamente
ID exclusivo do modelo de IA	Identificador exclusivo do modelo de IA	Gerado
Nome do modelo de IA	Nome do modelo de IA	Caixa de texto
Descrição do modelo de IA, incluindo finalidade	Descrição e finalidade do modelo de IA	Caixa de texto

Tabela 1: plano de inventário de modelos de IA

Nome do atributo	Definição	Tipo de atributo
Proprietário da análise de modelo de IA	Patrocinador/proprietário do modelo de IA	Menu suspenso
Proprietário de negócios do modelo de IA	Empresário do modelo de IA	Seleção múltipla
Desenvolvedor de modelos de IA	Nome do desenvolvedor do modelo de IA	Seleção múltipla
Provedor de modelo de IA	Nome do fornecedor/fornecedor do modelo de IA (pode ser interno ou fornecedor)	*Drop-down*: fornecedor interno ou externo (especifique)
Nome do fornecedor do modelo de IA (se aplicável)	Nome do fornecedor, se o modelo de IA for de origem externa (caso contrário, NA)	Caixa de texto
Validador de modelo de IA	Nome da equipe do validador de modelo de IA	Seleção múltipla
Linguagem de programação do modelo de IA	Modelos de IA desenvolvidos internamente - linguagem de codificação utilizada para desenvolver o modelo de IA; modelos de IA do fornecedor - *software* do fornecedor no qual o modelo de IA é operado	Menu suspenso
Fonte(s) de dados utilizada(s)	O repositório do qual os dados são originados (por exemplo, *data lake*). É possível vincular a fontes se as ferramentas organizacionais existentes forem usadas para implementar o inventário de modelo de IA	Menu suspenso
Domínio(s) de dados usado(s)	Domínio(s) de dados usado(s) para o modelo	Menu suspenso
Fonte(s) externa(s) de dados	Fonte(s) externa(s) de dados utilizada(s)	Caixa de texto
Sistema de implementação de modelo de IA	Descrição sobre como o modelo está sendo servido	Caixa de texto
Usuários do modelo de IA	Nome dos usuários/equipes pretendidos do modelo de IA	Menu suspenso

Tabela 1 (cont.): plano de inventário de modelos de IA

Apêndice V – Matriz *RACI* para Governança Responsável de *GenAI*

Apêndice V

A matriz *RACI* é uma ferramenta de gestão utilizada para definir e esclarecer papéis e responsabilidades em um projeto ou processo, aumentando a eficiência e a eficácia nas execuções das atividades. *RACI* é um acrônimo que representa quatro tipos de responsabilidades: *Responsible* (Responsável), *Accountable* (Aprovador), *Consulted* (Consultado) e *Informed* (Informado). Essa matriz ajuda a garantir que todos os envolvidos entendam suas funções específicas, evitando confusões e sobreposições de tarefas. O "Responsável" é quem executa a tarefa, o "Aprovador" é quem tem a autoridade final de decisão, o "Consultado" oferece conselhos e recomendações, e o "Informado" é atualizado sobre o progresso e os resultados.

R = Responsável **A** = Aprovador **C** = Consultado **I** = Informado

Fase do ciclo de vida da IA	Grupo de atividades	Atividades	1º LDD (Desenvolvimento)		2º LDD (Supervisão e Governança)			
			Gerentes e Desenvolvedores de IA	Gerentes de Negócios	Validadores da IA	Sustentação em Produção	Segurança da Informação	Comitê de Gestão de Riscos, Ética, *Compliance* e Jurídico
Planejamento	Definir objetivos e princípios	Definir e documentar o problema de negócio, propósito do projeto (escopo), métricas e requisitos de dados necessários.	R	AC			C	I
	Desenvolver plano de implementação	Criar um plano que inclua marcos, cronogramas e recursos necessários. Analisar as implicações éticas, sociais e legais do uso do modelo.	R	AC	I		C	AC
	Selecionar o modelo de IA Generativa	Avaliar diferentes modelos e selecionar o mais apropriado, com base nas necessidades do projeto.	R	I	AC			C

Tabela 2: matriz *RACI* para alinhamento de governança de IA

R = Responsável **A** = Aprovador **C** = Consultado **I** = Informado

Fase do ciclo de vida da IA	Grupo de atividades	Atividades	1º LDD (Desenvolvimento)		2º LDD (Supervisão e Governança)			
			Gerentes e Desenvolvedores de IA	Gerentes de Negócios	Validadores da IA	Sustentação em Produção	Segurança da Informação	Comitê de Gestão de Riscos, Ética, *Compliance* e Jurídico
Planejamento	Realizar avaliação de risco abrangente	Identificar potenciais riscos e desenvolver estratégias para mitigá-los.	R	CI	A	C	C	AC
Desenvolvimento	Coletar e preparar dados	Reunir dados de qualidade e realizar o pré-processamento necessário.	R	CI	AC		C	I
	Desenvolver o modelo	Construir/adotar o modelo utilizando os dados preparados.	R	CI	AC		C	
	Documentar o processo de desenvolvimento	Manter documentação detalhada de todas as etapas do desenvolvimento para referência futura e conformidade.	R	CI	A	C		
Homologação do modelo	Realizar testes	Testar o modelo para validar seu desempenho, confiabilidade, segurança e conformidade.	R	AC	AC	C	C	I
	Obter aprovação das partes interessadas	Apresentar os resultados dos testes e avaliações, e obter a aprovação formal antes da implementação em produção.	RI	AC	AC	AC	AC	AC
	Comunicar e treinar os usuários	Informar os usuários sobre o novo modelo e fornecer o treinamento necessário para sua utilização.	R	A	C	I	I	I

Tabela 2 (cont.): matriz *RACI* para alinhamento de governança de IA

R = Responsável **A** = Aprovador **C** = Consultado **I** = Informado

Fase do ciclo de vida da IA	Grupo de atividades	Atividades	1º LDD (Desenvolvimento)		2º LDD (Supervisão e Governança)			
			Gerentes e Desenvolvedores de IA	Gerentes de Negócios	Validadores da IA	Sustentação em Produção	Segurança da Informação	Comitê de Gestão de Riscos, Ética, *Compliance* e Jurídico
Implementação	Preparar o ambiente de produção	Configurar a infraestrutura necessária para suportar o modelo em produção.	CI	I	A	R	C	I
	Integrar o modelo ao sistema existente	Incorporar o modelo de IA nas aplicações ou processos atuais, garantindo compatibilidade.	CI	CI	A	R	C	I
Pós-implementação	Monitorar desempenho do modelo	Acompanhar continuamente o desempenho do modelo em produção para identificar e resolver problemas rapidamente.	I	AC	I	R	I	I
	Atualizar o modelo periodicamente	Reavaliar e atualizar o modelo conforme necessário para garantir que ele permaneça eficaz e relevante.	CI	AC	CI	R	C	C
	Gerenciar mudanças de forma eficaz	Implementar um processo de gestão de mudanças para ajustes futuros e comunicar essas mudanças aos usuários.	CI	R	I	I	C	C
	Assegurar governança contínua	Estabelecer uma estrutura de governança para supervisão contínua do uso da aplicação ou serviço, garantindo conformidade e adaptabilidade.	C	AC	C	R	I	CI

Figura 5 (cont.): matriz *RACI* para alinhamento de governança de IA

Apêndice VI – *Framework* de Classificação de Riscos

A classificação de riscos é uma forma sistemática de avaliar o grau de risco de um modelo de IA, quando sujeito a um determinado *framework* de governança. Essa ferramenta oferece consistência e transparência na hora de determinar o nível de risco da solução, de acordo com suas características de funcionamento e os impactos no funcionamento da companhia. Como qualquer solução implementada em uma empresa, o *framework* de classificação de risco de IA está sujeito a revisão regular durante todo o ciclo de vida da aplicação, e é um elemento importante para definir a frequência e a profundidade da validação/monitoramento/revisão do modelo de IA, que deve ser registrado no inventário de modelos de IA (ver **Apêndice IV**).

A seguir, são exibidas duas matrizes que compõem o *framework* de classificação de riscos de modelos de IA. A **Tabela 3** traz as frequências de monitoramento e revisão do modelo, de acordo com cada grau de risco da aplicação. Na **Tabela 3**, também é definida uma matriz de decisão que sintetiza respostas a uma combinação de perguntas qualitativas e quantitativas que auxiliam na classificação de risco para o modelo, baseada nos seguintes critérios de complexidade e materialidade do modelo, impacto estratégico e operacional, confiança no modelo, impacto reputacional e considerações regulatórias.

Nível de risco do modelo	Validação inicial	Monitoramento contínuo	Frequência de revisão
Risco Alto	Sim (alta prioridade)	Diariamente	Anual
Risco Médio	Sim	Mensalmente	A cada 2 anos
Risco Baixo	Escopo de validação reduzido	Trimestralmente	A cada 3 anos

Tabela 3: matriz de nivelamento de risco de modelos de IA

A implementação de modelos de Inteligência Artificial em processos de negócios traz uma série de benefícios, como a automação de tarefas, a melhoria na tomada de decisões e a otimização de operações. No entanto, esses modelos também introduzem riscos que precisam ser avaliados e gerenciados de forma eficaz para garantir que seu impacto seja positivo e alinhado aos objetivos estratégicos da organização.

Para abordar esses riscos, desenvolvemos um quadro de qualificação (ver Tabela 4) que ajuda a identificar, avaliar e categorizar os possíveis impactos dos modelos de IA. Este quadro considera diversos fatores de risco, como a complexidade do modelo, a materialidade do impacto financeiro, o impacto estratégico e operacional, a confiança no modelo, o impacto reputacional e as considerações regulatórias. Cada fator é analisado por meio de perguntas específicas, e as respostas a essas perguntas determinam o nível de risco associado (Baixo, Médio ou Alto).

O processo de qualificação é estruturado considerando:

- Complexidade do modelo: avalia a necessidade de grandes conjuntos de dados, dados desbalanceados, múltiplas fontes de dados, uso de dados não estruturados, complexidade computacional e qualidade dos dados.

- Materialidade do modelo: examina o impacto financeiro do modelo em termos de receita, custo e lucro.

- Impacto estratégico e operacional: determina como o modelo de IA influencia a tomada de decisões e operações estratégicas.

- Confiança no modelo: avalia a frequência de uso, abrangência geográfica e finalidade do modelo (produção de produtos/serviços *versus* relatórios/suporte).

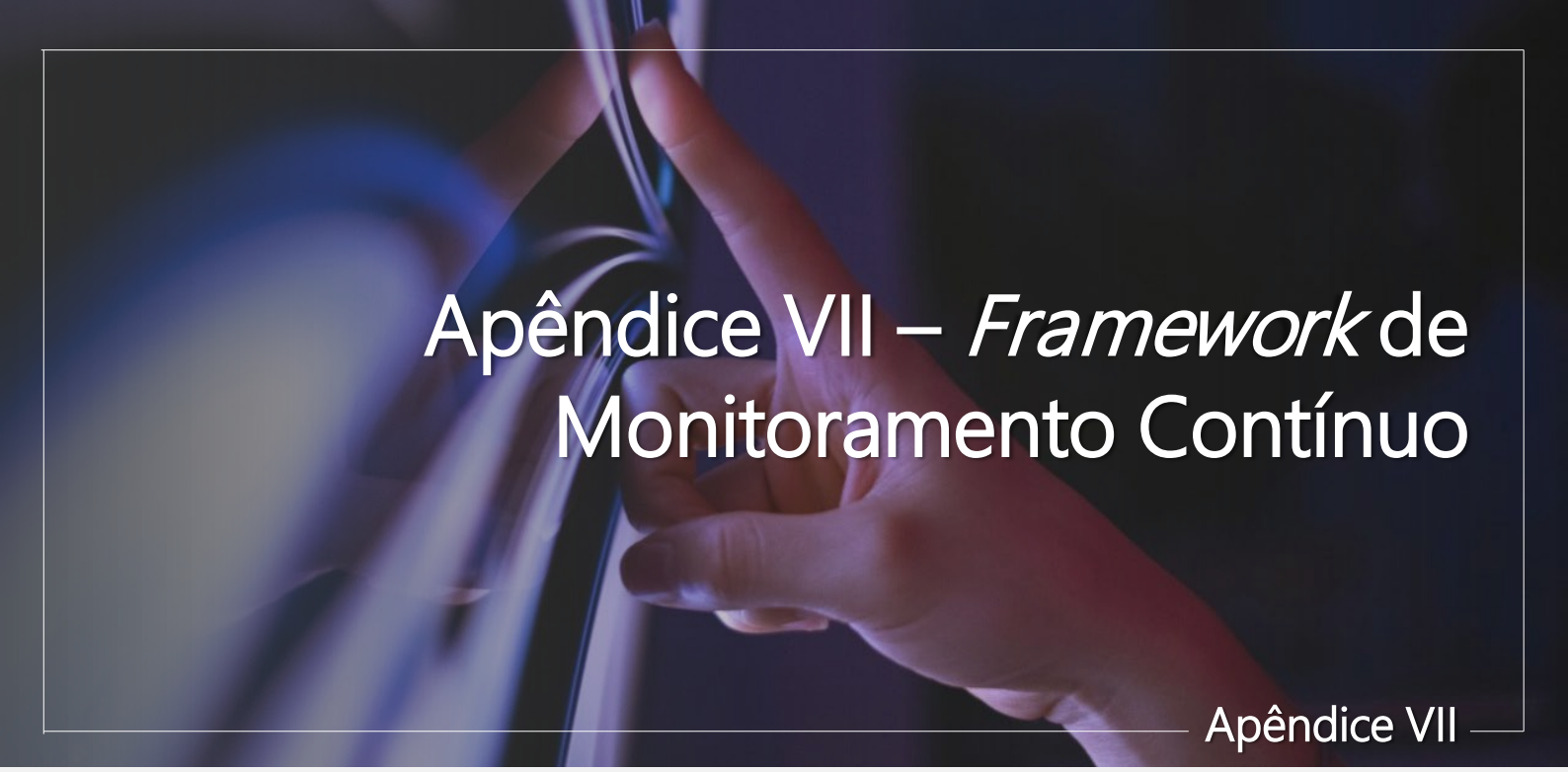
- Impacto reputacional: considera o uso de informações pessoais, a aplicação da equidade e o impacto da saída errônea do modelo na credibilidade organizacional.

- Considerações regulatórias: analisa a conformidade do modelo com requisitos regulatórios e legais.

Para cada fator de risco, são feitas perguntas específicas cuja respostas determinarão a qualificação do risco como Baixo, Médio ou Alto. A qualificação agregada é, então, calculada com base nos resultados dos fatores individuais, proporcionando uma visão clara dos riscos envolvidos na implementação e operação do modelo de IA.

Fator de risco	Questões para qualificação	Qualificação (Baixa/Média/Alta)
Complexidade do modelo	1. O modelo de IA requer um conjunto de dados de entrada muito grande (mais de 20 recursos)? 2. O modelo de IA foi treinado com conjunto de dados desbalanceado? 3. O modelo de IA requer dados extraídos de mais de três sistemas de fontes de dados? 4. O modelo de IA usa dados não estruturados ou dados externos? 5. O modelo de IA tem alta complexidade computacional? 6. Há problemas significativos de qualidade de dados que exijam correção?	Alta, se a resposta for "SIM" para duas perguntas ou mais; Baixa, se a resposta for "NÃO" para todas as perguntas; Média, para demais casos.
Materialidade do modelo	1. Qual o impacto do modelo de IA na receita/custo/lucro?	Alta, se a materialidade estimada for maior que XX; Média, se a materialidade estimada for maior que YY; Baixa, caso contrário; N/A, se não aplicável.
Impacto estratégico e operacional	1. Qual o impacto do modelo de IA na tomada de decisões estratégicas e operacionais?	Alta, se a saída do modelo informar diretamente as decisões e operações de negócios; Média, se for utilizado como entrada; Baixa, se for usado como informação de contexto.
Confiança no modelo	1. O modelo de IA será usado ao vivo ou diariamente? 2. O modelo de IA será usado em vários países? 3. O modelo de IA será usado para o fornecimento de produtos/serviços (*vs* relatórios, suporte)?	Alta, se a resposta for "SIM" para todas as perguntas; Baixa, se a resposta for "NÃO" para todas as perguntas; Média, para demais casos.
Impacto reputacional	1. O modelo de IA usa informações de identificação pessoal (PII) como dados de entrada? 2. A equidade é aplicável para o caso de uso? O impacto é baixo, médio ou alto? 3. A saída errônea do modelo de IA impactaria a credibilidade da organização com partes externas?	Alta, se PII for usada ou se a equidade for aplicável com alto impacto; Média, se não for utilizada PII e se a equidade for aplicável com baixo/médio impacto ou se a credibilidade da organização estiver em jogo; Baixa, para demais casos.
Considerações regulatórias	1. Alguma parte do modelo de IA está sujeita a requisitos de conformidade regulamentares ou legais? Se sim, é regulado direta ou indiretamente?	Alta, se o modelo for regulado; Média, se for regulado indiretamente; Baixa, para demais casos.
	Qualificação agregada (Baixa/Média/Alta)	Alta, se a qualificação for "ALTA" para dois ou mais fatores de risco; Baixa, se a qualificação for "MÉDIA" para, no máximo, um fator de risco; Média, para demais casos.

Tabela 4: matriz de decisão para classificação de risco de modelos de IA



Apêndice VII – *Framework* de Monitoramento Contínuo

A Monitoramento de Métricas

- As métricas para monitorar modelos de IA Generativa se dividem em:
- **Infraestrutura:** avaliam a eficiência e o custo do modelo, incluindo o custo por chamada de *API*, consumo de *tokens*, latência (tempo de resposta e de inferência), escalabilidade (uso de recursos e número de instâncias).
- **Conteúdo:** medem a qualidade e o desempenho das respostas, incluindo a taxa de alucinação (informações falsas), viés, coerência, toxicidade, relevância e precisão em tarefas específicas.

B Frequência de Monitoramento

- Modelos com autoaprendizagem e maior impacto nos negócios devem ser mais frequentemente monitorados do que modelos estáticos.
- Mais amplamente, definição de níveis de risco, incluindo materialidade, é fundamental para determinar frequência.

C Análise de Alterações

- Qualquer substituição ou nova versão de modelo deve ser documentada e devidamente justificada.
- Substituições substanciais em um modelo podem sinalizar a necessidade de um maior refinamento.

D Monitoramento de Entrada

- Verificações de qualidade e validade dos dados.
- Verificações adicionais para dados não estruturados.
- Utilize as ferramentas de avaliação do fluxo de *prompt* para analisar a eficácia do *prompt*. Identifique *prompts* que levam a respostas de baixa qualidade ou indesejadas.

E Monitoramento de Saída

- Filtragem de conteúdo e verificações de precisão e viés:
- Fundamentadas na verdade (sem alucinações);
- Relevantes, coerentes e
- em conformidade com as permissões de acesso a dados.

F Monitoramento de Modelos

- Monitoramento de desempenho: diretamente medindo o desempenho quando o alvo está disponível; e estimar o desempenho quando o alvo real está comprometido.
- Estabilidade do modelo: analisar a estabilidade do modelo (detecção de desvio de conceito, ou seja, se a relação entre entrada e saída Alterações).
- Comparação de importância de características.
- *Benchmarking* de modelos.

G Alertas de Revisão

- Gatilhos quantitativos - teste a precisão e a estabilidade do modelo no resultado recente e em relação a um modelo de *benchmark* (por exemplo, utilize a análise de sentimento e a detecção de entidades nomeadas para identificar linguagem imprópria nas respostas do modelo).
- Gatilhos qualitativos – mudanças na estratégia de negócios / requisitos regulatórios (por exemplo, características do produto, oferta de produtos condições).
- Gatilhos de *fallback* -mude para o modelo de referência/legado, se houver uma deterioração grave no desempenho do modelo (por exemplo, *marketing* de *fallback*).
- Gatilhos de processamento e capacidade - captura de aumento do uso do modelo ou aumento do consumo de dados (por exemplo, aumento do consumo de dados Demandas de E/S/Memória/*CPU*).

H Plano de Ação

- Diagnóstico e aprofundamento para *insights* adicionais.
- Plano de remediação para o modelo de reciclagem/recalibração/revisão.

Figura 2: *framework* de monitoramento contínuo para modelos de IA

O monitoramento de um modelo deve ser refinado, minucioso, com visão aprofundada e um mecanismo de acionamento automatizado. Um plano de monitoramento abrangente deve estar em vigor para monitorar quaisquer mudanças graduais ou repentinas que afetem a qualidade da previsão e a adequação do modelo. A Figura 2 traz uma proposta de *framework* para monitoramento contínuo de um modelo.

Monitoramento de Modelos de IA Generativa em Produção com o *Azure*

I. Integração com o *Azure Prompt Flow*

Aproveite os modelos pré-construídos: comece com um modelo pré-construído para tarefas comuns (por exemplo, perguntas e respostas, geração de texto) para acelerar o desenvolvimento.

Habilite o monitoramento durante a implantação: ao implantar seu fluxo de *prompt*, certifique-se de habilitar inferência, coleta de dados e integração com o *Application Insights*.

II. Coleta e Armazenamento de Dados

Azure *AI/ML Data Asset*: coleta automaticamente *prompts* de entrada e saídas do modelo em um ativo de dados.

ID de correlação: o *Azure AI/ML* gera automaticamente um ID de correlação para vincular *prompts* de entrada e respostas do modelo para análise contínua.

III. Monitoramento de Modelos

Criar um monitor: no estúdio do *Azure AI/ML*, crie um monitor vinculado à sua implantação de fluxo de *prompt*.

Especificar tipo de tarefa: escolha o tipo de tarefa relevante (por exemplo, *prompt* e conclusão para cenários de perguntas e respostas).

Conectar ativo de dados: o monitor detectará automaticamente o ativo de dados que contém os dados de entrada e saída.

Configurar métricas de avaliação: escolha as métricas relevantes entre as opções disponíveis (*grounding*, relevância, coerência, fluência, toxicidade). Defina o contexto para avaliar o *grounding* e a relevância.

Definir limiares e alertas: estabeleça limites aceitáveis para cada métrica e configure alertas para notificá-lo quando o desempenho do modelo estiver abaixo do esperado.

Agendar o monitoramento: defina a frequência do monitoramento (por exemplo, diário, por hora, contínuo).

IV. Painel de Monitoramento

Acessar *insights*: o *Azure AI/ML* fornece um painel para visualizar os resultados do monitoramento ao longo do tempo.

Analisar tendências: acompanhe as taxas de aprovação/reprovação para cada métrica, identifique

problemas em potencial e investigue as causas-raiz.

Entradas e saídas do modelo: aprofunde-se em instâncias específicas em que o desempenho do modelo foi problemático.

V. *Pipelines* do *Azure*

Monitoramento automatizado: o *Azure AI/ML* utiliza *pipelines* para automatizar o processo de monitoramento, calculando métricas e as enviando para o painel.

Extensibilidade: personalize o *pipeline* para incorporar métricas adicionais ou realizar análises mais aprofundadas, se necessário.

Considerações Adicionais

Engenharia de *prompt*: projete *prompts* com um entendimento das métricas de avaliação para garantir um monitoramento eficaz em produção.

Revisão humana: considere incorporar a revisão humana para casos complexos ou ao lidar com conteúdo sensível.

Melhoria contínua: utilize os *insights* do monitoramento para refinar iterativamente seus *prompts*, modelo e estratégias de implantação.

A Microsoft oferece, ainda, um conjunto de soluções que permitem a adoção segura e responsável da IA Generativa. Ferramentas como ***Defender***, ***Entra***, ***Purview*** e ***Intune***, trabalham em conjunto para proteger dados e interações em aplicações de *GenAI*, como o ***Microsoft Copilot*** e aplicativos de terceiros. Essas soluções permitem descobrir e avaliar riscos, proteger dados sensíveis com criptografia e DLP, e gerenciar o uso da IA com auditoria e detecção de violações de políticas. Para o ***Microsoft Copilot***, a Microsoft oferece recursos adicionais, como o ***AI Hub*** no ***Microsoft Purview*** para *insights* sobre riscos, proteção de dados confidenciais, detecção de ameaças e resposta, e governança e conformidade. Para aplicativos de IA de terceiros, a Microsoft oferece recursos de descoberta, proteção de acesso e proteção de dados sensíveis.

Apêndice VIII – Portfólio de Ferramentas de Dados e IA da Microsoft

Apêndice VIII

Aqui, resumimos os tipos de banco de dados e tecnologias de IA em pontos-chave, com foco em como eles se relacionam com os serviços do *Azure*.

Tipos de Banco de Dados e Casos de Uso

- **Relacional (SQL):** ideal para dados estruturados com relações, cargas de trabalho *OLTP* (Processamento de Transações *Online*), transações de alto volume e consultas rápidas. Dados normalizados garantem a integridade dos dados e reduzem a redundância.
- **Ferramentas Microsoft**:
  - *SQL Server* (*VM*, Instância Gerenciada, Banco de Dados),
  - *Azure PostgreSQL* , *Azure MySQL* , *Azure Database for MariaDB*

- **Não Relacional (NoSQL):** adequado para esquemas flexíveis, dados não estruturados ou semi-estruturados e escalabilidade horizontal. Diferentes tipos incluem documento, chave-valor, colunar e grafo.
- **Ferramenta Microsoft:**
  - *Azure Cosmos DB* (suporta múltiplas *API*s para modelos de documento, chave-valor, colunar e grafo)

- *Data Warehouses* (*OLAP*): projetados para análise e relatórios em grandes volumes de dados históricos. Dados desnormalizados priorizam consultas mais rápidas em relação à eficiência de armazenamento.
- **Ferramenta Microsoft:**
  - Azure Synapse Analytics

Fluxo de Dados e *Pipelines*

- **ETL/ELT:** *Extract, Transform, Load* (*ETL*) e *Extract, Load, Transform* (*ELT*) são padrões de processamento de dados. O ELT está ganhando popularidade devido aos custos de armazenamento mais baixos e à capacidade de reter dados brutos para análises futuras.

- **Ferramenta Microsoft:**
  - ***Azure Data Factory:*** orquestra o fluxo de controle em *pipelines* de dados, chamando atividades que realizam a extração, transformação e carregamento de dados em bancos de dados ou *data warehouses*.

IA e *Machine Learning*

- **Inteligência Artificial (IA):** emula a inteligência humana para resolver problemas.

- **Ferramentas Microsoft:**
  - ***Azure Machine Learning***: plataforma destinada ao desenvolvimento, treinamento, implantação e gerenciamento de modelos de aprendizado de máquina em larga escala. Ela é projetada para suportar o ciclo de vida completo do *machine learning*, desde a preparação dos dados até a implementação de modelos em produção.
  - ***Azure AI Services***: modelos de IA pré-construídos para visão, linguagem, fala, tomada de decisões e muito mais.
  - ***GitHub Copilot:*** projetado para ajudar os desenvolvedores em tarefas como geração de código, documentação e migração de sistemas legados para plataformas mais modernas.
  - ***Microsoft Copilot e Dynamics***: se você precisa de assistência de IA no ***Microsoft 365*** ou ***Dynamics*** (por exemplo, resumir *e-mails*, criar documentos), use os *copilots* específicos para esses ambientes. Esses *copilots* atuam como orquestradores, puxando dados relevantes e fornecendo respostas em linguagem natural.
  - ***Copilot Studio***: uma plataforma de código baixo/sem código para criar *copilots* personalizados. Permite: integração com várias fontes de dados (por exemplo, *sites* públicos, ***SharePoint, OneDrive***); definir ações e respostas usando uma interface simples; implantar *copilots* em vários canais, como ***Teams***, *Slack* ou *sites* personalizados.
  - ***Azure AI Studio:*** oferece capacidades para criar e gerenciar modelos de IA, com: controle total sobre os parâmetros e configurações do modelo; integração com a pesquisa de IA do *Azure* para recuperação eficiente de dados; capacidade de criar fluxos de trabalho complexos e personalizar *prompts* do sistema; integração simplificada de modelos de IA conversacional, como o *GPT-x* da ***OpenAI***, em dados personalizados.
  - ***Azure OpenAI Service:*** acesso aos modelos *GPT*-x (texto), *Codex* (código) e *DALL-E 2* (imagens) da *OpenAI* para geração de conteúdo.
  - **Ferramentas de Orquestração**: para desenvolvedores que desejam construir aplicações de IA sofisticadas, ferramentas de orquestração como *Langchain* e *Semantic Kernel* podem ajudar a gerenciar interações, memória e encadeamento de tarefas. Essas ferramentas abstraem grande parte da complexidade, facilitando a construção de soluções baseadas em IA.

- **Desenvolvimento Local de IA:** para executar modelos de IA localmente no seu *PC*, o ***AI Toolkit do Visual Studio Code (VS Code)*** suporta modelos de linguagem menores, permitindo que você desenvolva e teste aplicações de IA sem precisar de recursos na nuvem.

Referências

[1] *ChatGPT Sets Record for the Fastest-Growing User Base-Analyst Note*. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

[2] *The Difference Between Generative AI And Traditional AI: An Easy Explanation For Anyone.* https://www.forbes.com/sites/bernardmarr/2023/07/24/the-difference-between-generative-ai-and-traditional-ai-an-easy-explanation-for-anyone/?sh=5cfb7cb7508a

[3] *C-Suite members are optimistic but cautious about priorities and GenAI's impact.* https://www.thomsonreuters.com/en-us/posts/corporates/future-of-professionals-c-suite-survey-2024/

[4] *EY CEO Outlook Pulse Survey, January 2023.* https://info.ey.com/Nordics-Cross-Multiple-GC-2023-01-24-CEO-Outlook-Pulse-Survey_01LP.html

[5] *What's Dividing the C-Suite on Generative AI?* https://www.bcg.com/publications/2023/c-suite-genai-concerns-challenges

[6] *What is Responsible AI?* https://learn.microsoft.com/en-us/azure/machine-learning/concept-responsible-ai?view=azureml-api-2

[7] *Responsible and trusted AI.* https://learn.microsoft.com/en-us/azure/cloud-adoption-framework/innovate/best-practices/trusted-ai.

[8] *A ambição da IA deve avaliar a viabilidade, a oportunidade e o risco.* https://www.gartner.com.br/pt-br/tecnologia-da-informacao/temas/prontidao-para-a-ia

[9] What is Azure AI Studio? https://learn.microsoft.com/en-us/azure/ai-studio/what-is-ai-studio

[10] Learn about Microsoft Purview https://learn.microsoft.com/en-us/purview/purview

[11] AI Learning Hub | Microsoft Learn. https://learn.microsoft.com/en-us/ai/

[12] *Copilot* para *Microsoft 365. https://www.microsoft.com/pt-br/microsoft-365/microsoft-copilot*

[13] *Risk Assessment Guide for Microsoft Cloud.*
https://learn.microsoft.com/en-us/compliance/assurance/assurance-risk-assessment-guide

[14] *Discover, protect, and govern AI usage with Microsoft Security*
https://techcommunity.microsoft.com/t5/security-compliance-and-identity/discover-protect-and-govern-ai-usage-with-microsoft-security/ba-p/4078883

[15] *Microsoft Learn.*
https://learn.microsoft.com/pt-br/training/browse/?terms=copilot%20ms%20365

[16] *Microsoft Responsible AI Standard, v2*
https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE5cmFl

[17] *Announcing Microsoft's AI Customer Commitments - The Official Microsoft Blog.*
https://blogs.microsoft.com/blog/2023/0608/announcing-microsofts-ai-customercommitments/

[18] *NIST AI Risk Management Framework (AI RMF). https://www.nist.gov/itl/ai-risk-management-framework*

[19] *Partner Innovation | Empowering Responsible AI Practices Through Partners (microsoft.com).*
https://partnerinnovation.microsoft.com/initiatives/empowering-responsible-aipractices-through-partners/

[20] *Microsoft announces new Copilot Copyright Commitment for customers - Microsoft On the Issues.*
https://blogs.microsoft.com/on-the-issues/2023/09/07/copilot-copyrightcommitment-ai-legal-concerns/

[21] *Azure OpenAI customers are also required to use guardrails made available by the service and follow responsible development practices. For the full list of required mitigations for Azure OpenAI customers, see: Customer Copyright Commitment Required Mitigations | Microsoft Learn.*
https://learn.microsoft.com/en-us/legal/cognitive-services/openai/customer-copyright-commitment

[22] *Transparency Note for Azure AI Faceservice | Microsoft Learn.* https://learn.microsoft.com/en-us/legal/cognitiveservices/face/transparency-note?context=/azure/ai-services/computer-vision/context/context

[23] *Transparency Note for Azure OpenAI - Azure AI services | Microsoft Learn.*
https://learn.microsoft.com/en-us/legal/cognitive-services/openai/transparencynote?context=%2Fazure%2Fai-services%2Fopenai%2Fcontext%2Fcontext&tabs=text

[24] *Transparency note for Document Intelligence - Azure AI services | Microsoft Learn.* https://learn.microsoft.com/en-us/legal/cognitive-services/document-intelligence/transparency-note

[25] *Copilot in Bing: Our approach to Responsible AI.* http://aka.ms/responsibleAI-copilotinbing

[26] *Your Everyday AI Companion | Microsoft Bing.* https://www.microsoft.com/en-us/bing?form=MG0AUO&OCID=MG0AUO#faq

Relatório encomendado pela Microsoft e produzido pela EY

Shape the future
with confidence