

Guia de Uso de **Inteligência Artificial**

1. INTRODUÇÃO

Bem-vindo ao Guia de Uso de Inteligência Artificial, que estabelece diretrizes para o desenvolvimento, aquisição, implantação e uso responsável de soluções de IA na Fundação Roberto Marinho.

1.1. Contexto

O uso de ferramentas de IA tem um papel fundamental nos projetos das organizações em geral, trazendo uma série de benefícios que podem melhorar processos e operações. São eles:

- **Aumento da Eficiência:** Sistemas automatizados por IA podem executar tarefas repetitivas e rotineiras com alta produtividade, permitindo que as equipes se concentrem em atividades mais estratégicas.
- **Análise de Dados:** Ferramentas equipadas com IA podem processar grandes volumes de dados rapidamente, extraindo informações valiosas que ajudam na tomada de decisões.
- **Customização:** A IA permite personalizar produtos e serviços com base nas necessidades dos projetos e usuários.
- **Redução de Custos:** Soluções de IA podem ajudar a FRM a reduzir custos operacionais ao otimizar processos e melhorar a eficiência.
- **Inovação:** Tecnologias enriquecidas com IA possibilitam o desenvolvimento de novos produtos e serviços, estimulando a inovação e a competitividade no mercado.

1.2. Objetivos

Os objetivos principais deste guia são:

- Garantir o uso de ferramentas de IA de forma segura e alinhada a princípios éticos institucionais e às leis vigentes, em especial, a Lei Geral de Proteção de Dados (LGPD – Lei nº 13.709/2018).
- Assegurar a privacidade, o consentimento do titular de dados e a segurança dos dados utilizados em serviços, plataformas e ferramentas de IA.
- Fomentar iniciativas de IA que promovam inovação e inclusão digital, sem perpetuar vieses ou discriminações.
- Contribuir para o cumprimento de diretrizes e estratégias brasileiras de IA acompanhando a evolução dos marcos regulatórios.

1.3. Panorama Legal

É importante ressaltar que existem diversas propostas legislativas que buscam regular o uso da IA no Brasil, abordando aspectos como a responsabilidade por danos causados por sistemas de IA, a necessidade de transparência no funcionamento desses sistemas e a proibição de usos discriminatórios, entre outros que coloquem em risco o exercício de direitos fundamentais.

1.4. Escopo

Este guia destina-se a:

- Equipes internas, colaboradores, estagiários e aprendizes da FRM que desenvolvem, integram ou utilizam soluções de IA.
- Terceirizados, fornecedores e parceiros que prestem serviços, desenvolvam ou utilizem soluções de IA nos projetos em colaboração ou da FRM.
- Produtos, conteúdos e serviços desenvolvidos ou oferecidos pela FRM com funcionalidades de IA.
- Projetos de pesquisa, desenvolvimento e inovação (PD&I) que envolvam IA.

1.5. Termos e Definições

- **Inteligência Artificial (IA):** Capacidade de sistemas computacionais de simular o raciocínio humano, aprendendo, resolvendo problemas e tomando decisões com base em dados. Isso inclui técnicas como aprendizado de máquina, redes neurais e processamento de linguagem natural.
- **IA Generativa:** Modelos de IA com capacidade de criar conteúdo original (texto, imagem, áudio, vídeo) que simula autoria humana (ex.: ChatGPT, Claude, Gemini, Midjourney).
- **Alucinação:** Fenômeno em que modelos de IA geram respostas falsas, sem base real, porém convincentes, que podem ser inadvertidamente aceitas como corretas.
- **Dados Pessoais:** Informações que permitam identificar uma pessoa física (conforme LGPD).
- **Dados Sensíveis:** Dados de crianças, adolescentes, saúde, raça, religião, biometria etc., que exigem cautela adicional e estão sujeitos à LGPD.

- **Deepfake:** Conteúdo audiovisual manipulado por IA que simula imagem ou voz de uma pessoa.
- **Viés algorítmico:** Tendência ou distorção dos conjuntos de dados usados para treinar as ferramentas de IA Generativa, que pode influenciar os resultados e provocar, ocasionalmente, tratamento discriminatório a determinados grupos.
- **Prompt:** Em IA generativa, é o texto de entrada ou instrução fornecida ao modelo para orientar a geração de respostas ou conteúdo. Funciona como interface entre usuário e IA, variando de perguntas simples a instruções complexas. A qualidade do prompt influencia diretamente a relevância e a precisão da resposta, tornando sua formulação eficaz essencial para maximizar o potencial das ferramentas de IA generativa.
- **Viés de automação:** Tendência de confiar em sugestões de sistemas automatizados de tomada de decisão, sem uma análise cuidadosa e crítica dos resultados.
- **Large Language Model (LLM):** Ou Modelo de Linguagem, é uma grande rede neural artificial, treinada em enormes conjuntos de dados, com o objetivo de compreender e reproduzir padrões humanos na produção de textos, imagens, vídeos e sons.
- **Supervisão Humana (human-in-the-loop):** Processo de monitoramento e controle humano sobre sistemas de IA garantindo que decisões importantes sejam validadas por pessoas qualificadas.
- **Princípio do menor privilégio:** Conceito de segurança em que apenas os dados necessários para executar uma tarefa sejam acessados e utilizados.
- **Alto risco:** Considera-se de alto risco o sistema de IA que, por sua natureza ou contexto de uso, apresenta probabilidade e gravidade relevantes de impactos adversos a direitos fundamentais ou à segurança. Isso inclui sistemas que decidem o acesso a serviços e vagas, que podem possuir viés discriminatório, e/ou afetar a integridade física ou mental (com foco em crianças e vulneráveis) ou cuja transparência (auditabilidade) impeça a devida auditoria e controle humano.

2. PRINCÍPIOS ORIENTADORES

2.1. Princípios Éticos no uso de IA

- **Ética e Transparência:** Todos devem garantir que ferramentas ou programas sejam utilizados de maneira ética e transparente. Isso ajuda a construir confiança entre os usuários e minimiza preconceitos e discriminação.

- **Segurança:** O uso responsável de ferramentas de IA implica buscar sistemas que ofereçam segurança contra-ataques cibernéticos, protegendo tanto a organização quanto os usuários.
- **Responsabilidade Social:** A organização tem a obrigação de considerar o impacto social do uso de ferramentas de IA para não prejudicar pessoas e comunidades, ampliar desigualdades ou reforçar vieses preconceituosos.
- **Conformidade Legal:** Devem-se cumprir regulamentos e legislações relacionados ao uso de IA para evitar sanções legais e garantir a operação dentro de padrões éticos.
- **Feedback e Melhoria Contínua:** Devem-se implementar mecanismos para receber feedback dos usuários e stakeholders, auxiliando na identificação de problemas e na melhoria contínua das diretrizes de uso dos sistemas de IA.
- **Sustentabilidade:** Adotar práticas responsáveis na implementação da IA, considerando o seu impacto ambiental e a busca por soluções mais sustentáveis.

2.2. Conformidade Legal

Todas as iniciativas de IA devem atender ao código de conduta e políticas internas da FRM, bem como à legislação aplicável. A instituição manterá acompanhamento contínuo de novas leis ou normativas sobre IA para promover as revisões necessárias.

2.3. Ética, Dignidade Humana e Respeito às Leis

A instituição repudia qualquer forma de manipulação, abuso ou discriminação potencializada por IA. Todo uso de IA deve preservar a dignidade, a liberdade de expressão e promover a inclusão digital de diferentes públicos.

2.4. Privacidade e Segurança de Dados

A LGPD e demais normas de proteção de dados devem ser integralmente observadas, mediante a implementação de medidas técnicas e organizacionais adequadas para proteção de dados pessoais.

2.5. Transparência e Explicabilidade

- Usuários devem ser informados quando estão interagindo com IA, além de serem alertados sobre seus riscos e limitações.

- Conteúdos criados para produtos e serviços que forem gerados ou assistidos por IA devem trazer menção clara (“criado/auxiliado por IA”) e passar por revisão humana para evitar informações falsas ou violações de direitos.
- Os usuários devem se certificar de que todo o material utilizado seja corretamente atribuído, inclusive em caso de citações, e de que as fontes citadas respaldem as declarações da aplicação de IA Generativa, pois não é incomum que gerem referências a obras inexistentes.
- A ocultação do uso e conteúdo de IA é uma falha ética que viola os princípios de transparência e honestidade em pesquisas e prestações de serviços.

2.6. Equidade e Inclusão

Soluções de IA não devem reforçar vieses discriminatórios de gênero, raça, classe socioeconômica ou outros, em conformidade com o código de conduta institucional e com o Guia de Diversidade e Inclusão da FRM. Caso sejam detectados vieses, a instituição adotará medidas corretivas.

2.7. Governança e Responsabilização

Será criado um comitê interno responsável pela supervisão dos projetos, alinhamento às melhores práticas e resolução de incidentes. A instituição deverá documentar formalmente o uso de IA (inventário de ferramentas, relatórios de risco) e realizar avaliações periódicas.

2.8. Inovação Responsável

O comitê interno de IA fomentará a adoção de IA em projetos de pesquisa, criação de conteúdo e formações, observando-se protocolos de teste e validação para garantir segurança e qualidade antes de implementação em larga escala.

3. CENÁRIOS DE APLICAÇÃO

3.1. Chatbots e Assistentes Virtuais

- **Finalidade:** Tirar dúvidas sobre conteúdo, guiar processos e oferecer suporte automático inicial.

- **Cuidados:** Indicar claramente que são soluções de IA, manter histórico de interações relevantes para fins de auditoria e revisões, permitir o encaminhamento para atendimento humano, no caso de suporte, e proteger dados fornecidos por usuários.

3.2. Análise de Dados

- **Finalidade:** Avaliar métricas, prever demandas, identificar padrões e tendências, detectar anomalias.
- **Cuidados:** Evitar coleta excessiva e uso indevido de dados, e revisar periodicamente os modelos para eliminar vieses.

3.3. Criação de Conteúdo com IA Generativa

- **Finalidade:** Produzir conteúdo, resumos, roteiros, imagens, vídeos, postagens em redes sociais, entre outros.
- **Cuidados:** Indicar a autoria em conteúdos gerados por IA, realizar sempre revisão humana antes da publicação ou compartilhamento. A FRM utilizará apenas dados autorizados, licenciados ou de sua própria titularidade para a geração do conteúdo, e não compartilhará com terceiros o conteúdo “imputado” ou gerado a partir da ferramenta, quando este não for permitido.

3.4. Agentes de IA

- **Finalidade:** Automatizar workflows, otimizar rotinas, simular cenários.
- **Cuidados:** Definir os limites de atuação sem supervisão humana, monitorar a execução contínua, e revisar os resultados com regularidade.

4. DIRETRIZES GERAIS

4.1. Uso de IA Generativa

- Verificar termos de uso de modelos e respeitar direitos autorais. Quando a ferramenta de IA utilizar obras de terceiros (ex.: datasets), garantir que licenças e atribuições sejam observadas.
- Submeter à revisão todo conteúdo gerado por IA destinado à publicação. Em caso de alterações de imagens e vozes reais, rotular explicitamente e obter consentimentos.

- Informar de forma clara quando o conteúdo for gerado, total ou parcialmente, por inteligência artificial e quando o usuário estiver interagindo com a tecnologia. Também esclarecer as limitações e possíveis vieses do modelo (ex.: Este texto foi gerado por IA e pode conter erros ou omissões). A instituição corrigirá ou removerá qualquer conteúdo que viole direitos ou contenha informações falsas.
- Certificar-se de que todo o material utilizado seja devidamente atribuído, incluindo citações, e que as fontes citadas comprovem as informações apresentadas pela aplicação de IA Generativa, já que a tecnologia pode criar referências a obras inexistentes.

4.2. Uso Interno de Ferramentas de IA

- Evitar a inserção de dados sensíveis ou confidenciais em ferramentas públicas de IA sem garantias contratuais de segurança e privacidade.
- Aplicar o “princípio do menor privilégio” no acesso a dados pela IA e verificar políticas de segurança antes do uso efetivo.
- Recomenda-se que só sejam inseridas informações confidenciais em soluções de IA Generativa aprovadas pelo Comitê de IA, conforme item 6.2 deste guia. O autor continua responsável pelo conteúdo do documento, mesmo que tenha utilizado ferramentas de IA em sua elaboração. O uso da tecnologia, ainda que apropriado, não substitui a revisão cuidadosa do material, nem transfere a autoria ou a responsabilidade pública pelo resultado final.
- A instituição vai oferecer treinamento contínuo aos colaboradores para informar sobre os riscos de alucinações, vieses, segurança e verificação de confiabilidade do conteúdo gerado por IA.
- Indica-se que a implementação de soluções de IA Generativa siga um processo de concordância prévia pela área demandante, em conjunto com as áreas de TI e Assessoria Jurídica.

4.3. Gestão, Análise e Segurança de Dados

- Implementar soluções de IA que permitam monitoramento e correções em processos automatizados internos (ex.: análises preditivas, correções automatizadas).
- Manter logs de atividades para auditoria e conformidade, garantindo transparência e rastreabilidade.
- Adotar a supervisão humana obrigatória em processos decisórios relevantes e de grande impacto.

5. SEGURANÇA E PRIVACIDADE

5.1. Medidas de Proteção de Dados e Gestão de Riscos

- Para manter a segurança dos dados e sistemas da FRM, sugere-se que sejam avaliados com cautela os riscos de acessar e utilizar plataformas externas de IA ao realizar as atividades institucionais.
- Na produção de conteúdo voltado para o público externo, recomenda-se usar apenas aplicativos de IA Generativa aprovados pelo Comitê de IA.
- Para manter a confidencialidade das informações sensíveis da instituição – incluindo informações protegidas por lei, dados sensíveis ou não consentidos, informações pessoais e materiais sob propriedade intelectual –, recomenda-se que sejam inseridas informações confidenciais apenas em soluções de IA previamente aprovadas, e sempre com o conhecimento da Assessoria Jurídica e do Comitê de Privacidade da instituição.

5.2. Canal de Informações e Incidentes

- A instituição, junto com o Comitê Interno de IA, vai implementar canal para feedback e informações sobre incidentes relacionados ao uso indevido de IA, conteúdos ofensivos ou violações de privacidade.
- Acionar o Comitê interno de IA para avaliar cada caso e adotar as medidas corretivas aplicáveis.

6. GOVERNANÇA E RESPONSABILIDADES

6.1. Comitê de Privacidade

- Garante conformidade à LGPD e a este guia.
- Conduz avaliações de risco e apoia nos aspectos técnicos e éticos em projetos de IA de alto risco.
- Acompanha incidentes e orienta equipes sobre coleta, armazenamento e tratamento de dados pessoais.

6.2. Comitê Interno de IA

- Composto obrigatoriamente por representantes das áreas de TI, Assessoria Jurídica, Comitê de Privacidade, além da participação desejável de outras áreas da instituição.
- Avalia novos projetos de IA para definir criticidade, riscos e possíveis mitigações, identificando violações de privacidade, segurança e vieses.
- Monitora riscos, zela pela conformidade e delibera sobre incidentes.
- Propõe reflexões e projetos de PD&I envolvendo o uso de IA na FRM.
- Conduz ou designa especialistas para avaliações de impacto de soluções.
- Realiza revisões periódicas deste guia.
- Analisa as plataformas de IA Generativa e suas finalidades de usos.

6.3. Equipe de TI

- Analisa e aprova projetos de IA.
- Acompanha a implementação de soluções.
- Avalia aspectos técnicos antes de se disponibilizar soluções e ferramentas de IA em produção.
- Implementa controles técnicos (logs, testes de vulnerabilidades etc.).
- Apoia o monitoramento de performance, segurança e conformidade.

6.4. Equipe de Assessoria Jurídica

- Acompanha a implementação de soluções.
- Avalia aspectos legais antes da disponibilização de soluções e ferramentas de IA em produção.
- Avalia, se for o caso, o direito de uso dos conteúdos fornecidos e utilizados para criação de bases de conhecimento.

6.5. Equipe de Ativos

- Analisa os Termos de Uso das plataformas de criação de conteúdos, além dos materiais fornecidos e utilizados nas bases de conhecimento.

6.6. Lideranças de Projetos

- Avaliam riscos e impactos ao ciclo de vida do projeto.
- Consultam Comitê de Privacidade e/ou o Comitê Interno de IA em caso de uso de dados sensíveis ou cenários de risco.
- Realizam testes e homologações.
- Garantem a qualidade dos conteúdos fornecidos e utilizados para a criação de bases de conhecimento, e os seus direitos de uso, mediante consulta a área de ativos e/ou assessoria jurídica.
- Mantêm a responsabilidade pela supervisão humana (human-in-the-loop) em processos decisórios sensíveis.

6.7. Terceirizados, Fornecedores e Parceiros

- Cumprem este guia por cláusulas contratuais, incluindo privacidade, confidencialidade e protocolos de incidentes.
- Previamente à realização de suas atividades, informam se o conteúdo ou serviço prestado será gerado ou realizado por IA (total ou parcialmente), quais ferramentas serão utilizadas e se haverá a inserção de dados ou conteúdos FRM nas mesmas.
- Se solicitados, os prompts utilizados para produção do conteúdo ou serviço prestado devem ser compartilhados com a FRM.

7. MONITORAMENTO E MELHORIA CONTÍNUA

7.1. Avaliação Interna Periódica

Revisar soluções de IA com a periodicidade recomendada pelo Comitê Interno de IA, ou quando ocorrer mudança significativa de tecnologia ou legislação.

7.2. Auditorias Externas

Para soluções de alto risco, prever auditoria externa independente, documentar achados e implementar recomendações.

7.3. Medidas Corretivas

Em casos de denúncia, detecção de comportamento indesejado ou mau funcionamento, o Comitê Interno de IA deverá ser acionado para avaliação do problema e definição dos procedimentos de resposta.

8. VIGÊNCIA, REVISÕES E ATUALIZAÇÕES DESTE GUIA

- **Revisões e Atualizações:** Periodicidade semestral sugerida, com atualizações extraordinárias em casos de novas regulamentações, evolução tecnológica ou recomendações da ANPD.
- **Responsáveis:** Comitê Interno de IA e Comitê de Privacidade.
- **Vigência e Aprovação:** Este guia entra em vigor na data de sua divulgação interna. Situações não previstas serão avaliadas pelo Comitê Interno de IA, em conjunto com demais órgãos e/ou áreas envolvidas/interessadas.

9. DISPOSIÇÕES FINAIS

Este guia:

- Reforça o compromisso da Fundação Roberto Marinho com inovação responsável, aliando geração de conhecimento e proteção de direitos.
- Complementa as demais políticas internas existentes (Política de Privacidade, Código de Ética e Conduta, Política de Segurança da Informação etc.).
- Orienta que, em situações de riscos, conflitos ou lacunas, deve-se recorrer ao Comitê Interno de IA para orientação.
- Convida a enviar dúvidas ou sugestões ao Comitê Interno de IA.

Versão: V1.0

Observação: A implementação integral deste guia requer colaboração ativa das áreas, além de sensibilização contínua de colaboradores, usuários e parceiros quanto aos princípios de uso ético e seguro de IA.