UM GUIA DE
INTELIGÊNCIA
ARTIFICIAL PARA
O SETOR PÚBLICO

1 01

INTELIGÊNCIA ARTIFICIAL NA GESTÃO PÚBLICA

CLP
CENTRO DE LIDERANÇA PÚBLICA

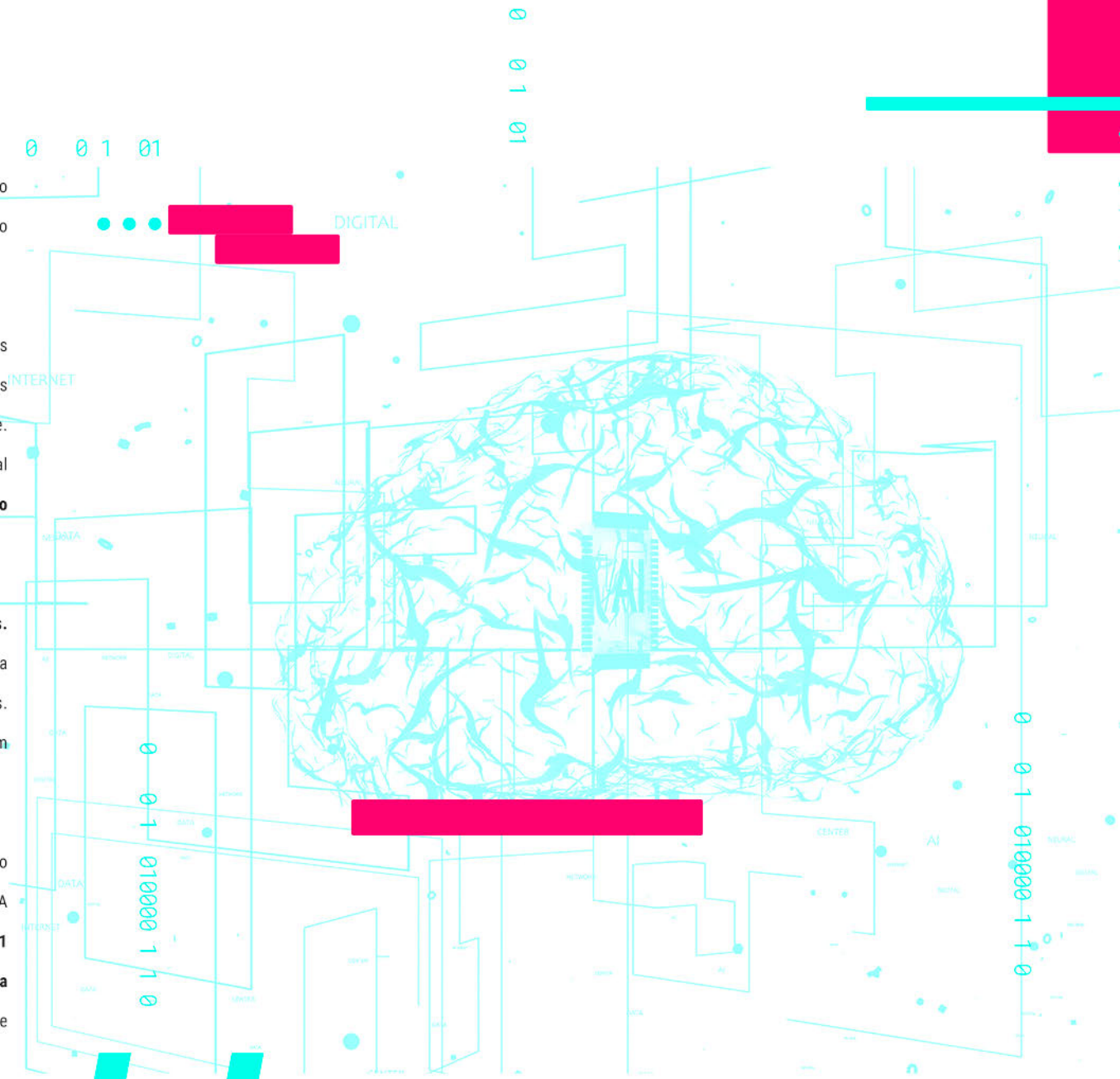
Microsoft

O aumento exponencial do uso da inteligência artificial (IA) no cenário contemporâneo levou a tecnologia além do setor privado e está impulsionando uma revolução ampla, principalmente no setor público.

A IA tem imenso potencial e vem reformulando a maneira como os governos operam e interagem com os cidadãos. Contudo, ela também apresenta desafios e riscos significativos para as organizações, a sociedade e o meio ambiente. Diante da magnitude das transformações que estão acontecendo, é fundamental que os **servidores públicos estejam adequadamente orientados para garantir o uso ético, transparente e benéfico da IA.**

No setor público, a integração dessas tecnologias envolve questões complexas. A IA oferece um vasto espectro de aplicações, desde a saúde à segurança pública, promovendo melhorias na eficiência e na tomada de decisões. A automação, por exemplo, permite que os servidores públicos se concentrem em trabalhos mais complexos e estratégicos.

Mas a gestão responsável da IA é vital para as instituições. Por isso, países ao redor do mundo estão em busca de compreender e moldar seu uso, tornando a IA um foco de debate e colaboração global. Neste contexto, o **"NIST AI 100-1 Artificial Intelligence Risk Management Framework (AI RMF 1.0)", ou "Estrutura de Gestão de Risco de Inteligência Artificial"**, em tradução livre, destaca-se como uma ferramenta fundamental.



/ o que faz o NIST? Para que serve o AI RMF?

O NIST (National Institute of Standards and Technology) é um instituto que integra o Departamento de Comércio dos EUA e promove a inovação através do avanço da ciência e da tecnologia. Já o AI RMF - como vamos chamar o documento ao longo deste guia - é uma publicação que ultrapassa as fronteiras norte-americanas, servindo como um mapa para organizações em todo o mundo ao solidificar uma linguagem e metodologia comuns na gestão de riscos relacionados à IA.

O AI RMF traz diretrizes específicas para a gestão de riscos e reflete os esforços de diversos stakeholders, incluindo a indústria, a academia e o governo. Alinha-se a padrões e diretrizes internacionais, colaborando com uma linguagem comum no desenvolvimento responsável da IA globalmente.

Além disso, é um documento vivo, em constante atualização, flexível e adaptável às necessidades específicas de cada organização. Você pode acessar a versão original clicando aqui.

AI RMF: bússola para a implementação da IA

Ao se observar a implementação da IA no setor público em diversos países pioneiros, como Singapura, Reino Unido e Canadá, percebe-se a revolução no atendimento público e a consequente necessidade de uma gestão de riscos bem estruturada. **O AI RMF não apenas contextualiza e orienta sobre os riscos e benefícios da IA, mas também serve como uma bússola para servidores públicos** na jornada de integração responsável e eficaz da IA nos serviços públicos.

Nós do Centro de Liderança Pública (CLP) e do Unidos pelo Brasil, em parceria com a Microsoft, preparamos para você, gestor público, este guia especial. Nosso objetivo é fornecer ferramentas que facilitem a sua atuação nesse universo em constante mudança da IA, a partir da abordagem estruturada do AI RMF para a gestão de riscos, visando maximizar os benefícios desta tecnologia revolucionária.

A seguir, vamos nos aprofundar nas diretrizes, melhores práticas e nos casos de uso, sempre visando contribuir para o seu repertório no tema inteligência artificial e, consequentemente, para um serviço público mais eficaz e inovador. Boa leitura!

Os conceitos da Inteligência Artificial

Antes de tudo, é importante que você entenda como a IA funciona.

Grosso modo, é a capacidade de uma máquina para reproduzir competências humanas, entre elas, aprendizagem, raciocínio e criatividade. Neste guia, vamos além disso - nosso ponto de partida serão os conceitos da inteligência artificial. Confira.

Validação

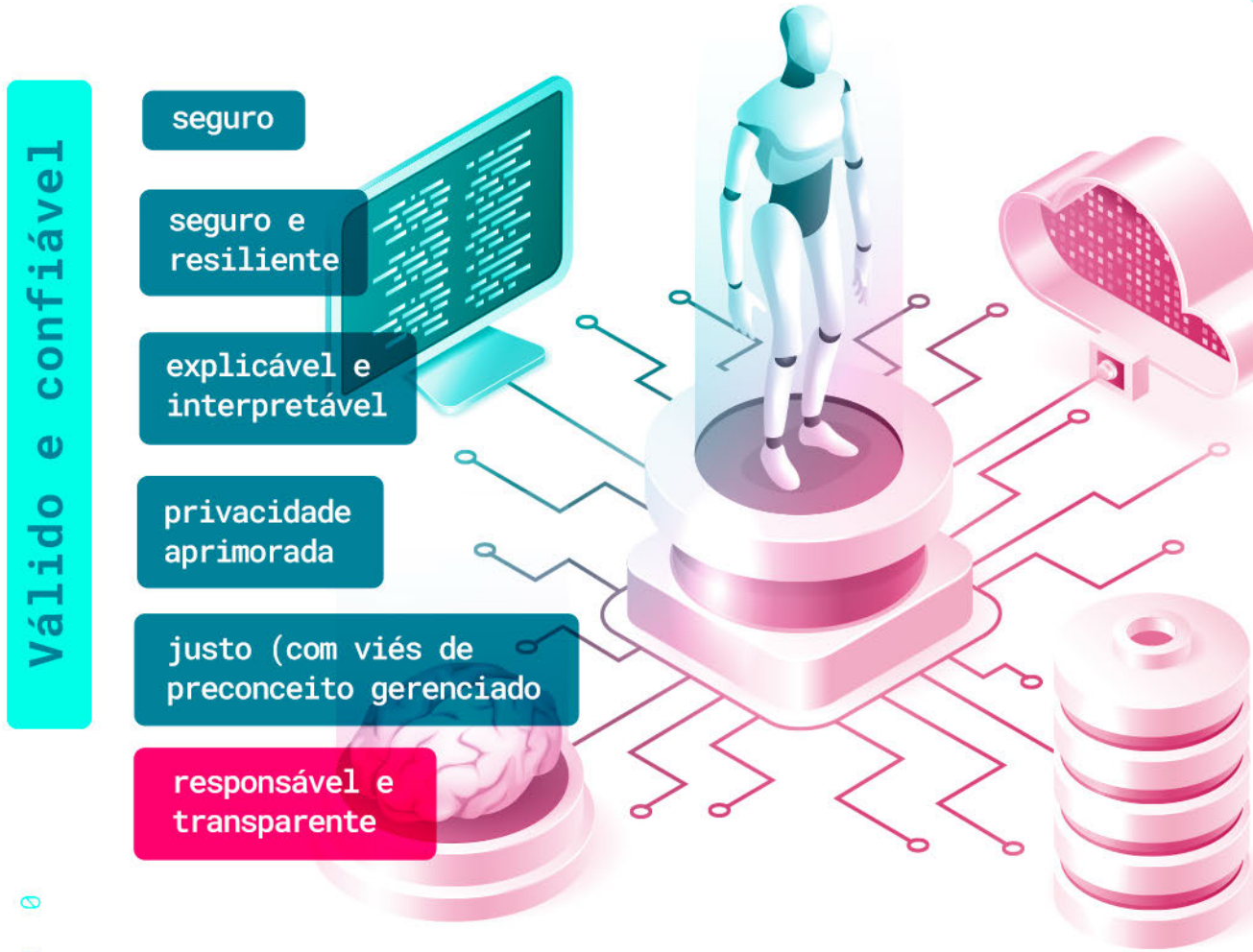
A validação é o processo de confirmar, por meio da apresentação de evidências objetivas, que os requisitos para um uso ou aplicação específicos foram atendidos. No contexto de sistemas de IA, a validação envolve garantir que o sistema seja preciso, confiável e generalizável para dados e contextos além de seu treinamento.

Confiabilidade

A confiabilidade é a capacidade de um item funcionar conforme necessário, sem falhas, durante um intervalo de tempo especificado, sob condições determinadas. No contexto de sistemas de IA, a confiabilidade é um objetivo para a correção geral da operação do sistema nas condições de uso esperadas e durante um determinado período de tempo, incluindo toda a vida útil do sistema.

Como vemos no infográfico abaixo, extraído do AI RMF, confiabilidade, responsabilidade e transparência são premissas de um sistema de IA.

/ conceitos da inteligência artificial /



Precisão

A precisão é a proximidade dos resultados de observações, cálculos ou estimativas em relação aos valores verdadeiros ou aos valores aceitos como verdadeiros. No contexto dos sistemas de IA, a precisão refere-se à capacidade do sistema de produzir saídas corretas com base nos dados de entrada e no uso pretendido do sistema.

Robustez

A robustez é a capacidade de um sistema manter seu nível de desempenho sob uma variedade de circunstâncias. No contexto dos sistemas de IA, robustez refere-se à capacidade do sistema de se desempenhar bem em um amplo conjunto de condições e circunstâncias, incluindo usos do sistema não inicialmente previstos.

Generalização

A generalização é a capacidade de um sistema de se desempenhar bem em dados e contextos além de seu treinamento. No contexto dos sistemas de IA, a generalização refere-se a essa capacidade aplicada a situações com novos dados que o sistema não tenha visto anteriormente.

Segurança

No contexto da IA, este conceito se refere ao grau em que um sistema de inteligência artificial pode operar sem causar danos aos seres humanos, à propriedade ou ao ambiente, e sem ser comprometido por intervenientes maliciosos. Os riscos de segurança que representam um risco potencial de lesões graves ou morte exigem uma priorização mais urgente e um processo de gestão de riscos mais completo.

Explicabilidade e Interpretabilidade

Explicabilidade se refere a uma representação dos mecanismos subjacentes à operação dos sistemas de IA, enquanto a interpretabilidade se refere ao significado dos resultados dos sistemas de IA no contexto dos objetivos funcionais concebidos.

Juntas, a explicabilidade e a interpretabilidade ajudam aqueles que operam ou supervisionam um sistema de IA, bem como os seus usuários, a obter conhecimentos mais profundos sobre a funcionalidade e a confiabilidade do sistema, incluindo os seus resultados.

Percepções de risco resultam da falta de capacidade de compreender adequadamente os resultados do sistema. Os sistemas de IA explicáveis e interpretáveis oferecem informações que ajudarão os usuários finais a compreender os propósitos e o impacto potencial de um sistema de IA.

Ciclo de Vida da IA

A OCDE (Organização para a Cooperação e Desenvolvimento Econômico) desenvolveu uma estrutura para classificar as atividades do ciclo de vida da IA de acordo com cinco dimensões sociotécnicas: dados de entrada e conhecimento, modelos e algoritmos, avaliação e tomada de decisão, implantação e implementação, e ciclos de feedback.

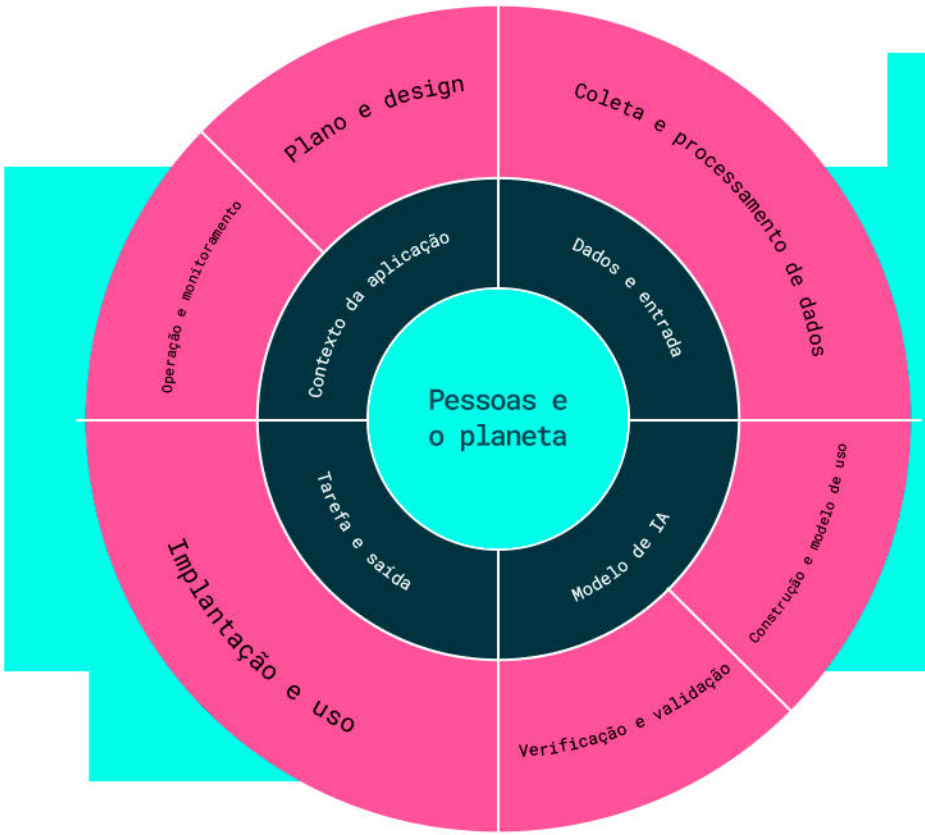
A estrutura também enfatiza a importância da transparência, responsabilidade e explicabilidade ao longo do ciclo de vida da IA, fatores cruciais para construir confiança nos sistemas e garantir que sejam usados de maneira responsável e ética.

No geral, a estrutura oferece uma ferramenta útil para os decisores políticos e partes interessadas entenderem as complexidades da IA e desenvolverem estratégias eficazes de gestão de riscos.

> **Essas dimensões visam ajudar os decisores políticos e as partes interessadas a entender as diferentes fases do ciclo de vida da IA e os riscos e benefícios potenciais associados a cada fase.**

Vale destacar que a estrutura mostra a importância de considerar todo o ciclo de vida da IA, desde a coleta de dados até os ciclos de feedback, ao gerenciar os riscos relacionados à IA. Isso ocorre porque os riscos podem surgir em qualquer estágio, e para abordar esses riscos é preciso levar em conta todo o sistema.

/ ciclo de vida e dimensões principais de um sistema de IA /



Riscos da IA

Compreender e abordar os riscos, impactos e danos associados aos sistemas de IA é fundamental. No contexto do AI RMF, risco é definido como a medida composta da probabilidade de um evento ocorrer e a magnitude ou grau das consequências do evento correspondente. **Os impactos ou consequências dos sistemas de IA podem ser positivos, negativos ou ambos, resultando em oportunidades ou ameaças.**

Ao considerar o impacto negativo de um evento, o risco é uma função do impacto negativo ou magnitude do dano que surgiria caso o evento ocorresse, e a probabilidade de ocorrência.

O documento mostra que impactos negativos ou danos podem ser experimentados por indivíduos, grupos, comunidades, organizações, sociedade, meio ambiente e o planeta. Portanto, abordar, documentar e gerenciar eficazmente os riscos e os potenciais impactos negativos pode levar a sistemas de IA mais confiáveis.

/ exemplos de danos potenciais relacionados aos sistemas de IA /

1. Consequências não intencionais ou erros nos resultados do sistema de IA
2. Preconceito ou discriminação na tomada de decisões do sistema de IA
3. Vulnerabilidades de segurança ou ataques maliciosos ao sistema de IA
4. Violações de privacidade ou acesso não autorizado a dados confidenciais
5. Impactos negativos no emprego ou na estabilidade econômica
6. Preocupações sociais ou éticas relacionadas com a utilização de sistemas de IA

Como usar a inteligência artificial?

Como vimos, os sistemas de IA não devem, em nenhuma hipótese, impactar negativamente as pessoas, a sociedade, as organizações e o meio ambiente. Um bom sistema precisa ser resiliente a eventos adversos, por isso existem várias recomendações para garantir a segurança e proteção da IA. Confira a seguir.

Considerações de segurança durante o ciclo de vida

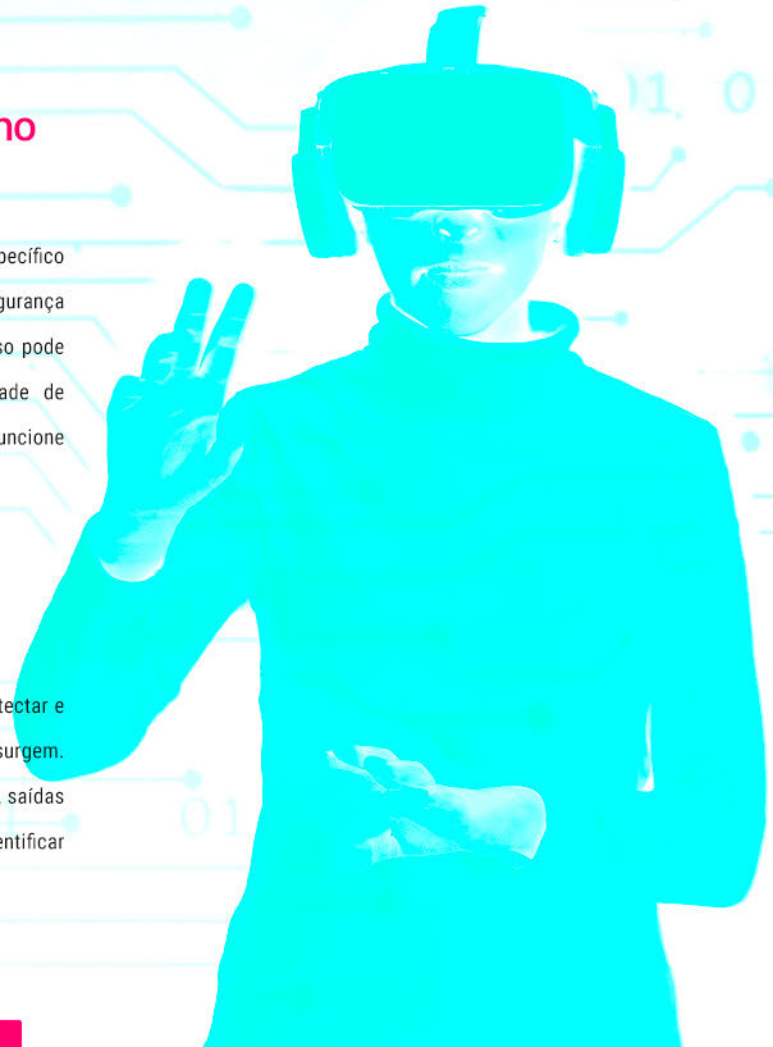
As considerações de segurança devem ser incorporadas no ciclo de vida do sistema de IA o mais cedo possível, começando com o planejamento e design. Isso pode ajudar a prevenir falhas ou condições que possam tornar um sistema perigoso.

Simulação e testes no domínio específico

A simulação rigorosa e testes no domínio específico podem ajudar a identificar potenciais riscos de segurança antes de um sistema de IA ser implementado. Isso pode envolver o teste do sistema sob uma variedade de condições e cenários para garantir que ele funcione conforme o esperado.

Monitoramento em tempo real

O monitoramento em tempo real pode ajudar a detectar e responder a questões de segurança à medida que surgem. Isso pode envolver o monitoramento das entradas, saídas e processos internos do sistema para identificar anomalias ou comportamentos inesperados.



Capacidade de intervenção humana

Os sistemas de IA devem ser projetados com a capacidade de desligar, modificar ou ter intervenção humana caso se desviem da funcionalidade pretendida ou esperada. Isso pode ajudar a prevenir ou mitigar riscos de segurança.

Alinhamento com diretrizes ou padrões existentes

As abordagens de gestão de riscos de segurança da IA devem ser inspiradas por esforços e diretrizes para a segurança em campos como transporte e saúde, e alinhar-se com diretrizes ou padrões específicos do setor ou aplicação.

Em suma, garantir a segurança e proteção da IA requer uma abordagem abrangente que leve em conta todo o ciclo de vida do sistema de IA e incorpore as melhores práticas de campos relacionados.

Como prevenir viés e discriminação da inteligência artificial?

O viés prejudicial pode ocorrer quando sistemas de IA são treinados com dados enviesados ou quando os próprios algoritmos utilizados para tomar decisões são enviesados, gerando resultados injustos para certos grupos ou indivíduos. Para prevenir esses problemas e garantir a equidade na IA, o AI RMF traz cinco recomendações:

Utilizar dados diversos e representativos

Sistemas de IA devem ser treinados com dados diversificados e representativos para garantir que eles não perpetuem vieses ou discriminações já existentes.

Monitorar quanto ao viés e discriminação

Sistemas de IA devem ser monitorados em relação ao viés e discriminação, e ações corretivas adequadas devem ser tomadas se um viés for detectado.

Utilizar IA explicável e interpretável

A IA explicável e interpretável pode ajudar a identificar e abordar o viés e a discriminação, fornecendo insights sobre os processos de tomada de decisão utilizados pelo sistema.

Considerar a equidade no design e desenvolvimento de sistemas de IA

A equidade deve ser considerada em todo o design e desenvolvimento de sistemas de IA, incluindo a seleção de algoritmos e a definição de métricas de desempenho.

Envolver stakeholders no desenvolvimento e implantação de sistemas de IA

Os stakeholders, incluindo aqueles que podem ser afetados pelo sistema, devem ser envolvidos no desenvolvimento e implantação de sistemas de IA para garantir que suas perspectivas sejam levadas em consideração.

Prevenir a injustiça na inteligência artificial requer uma abordagem que leve em conta todo o ciclo de vida do sistema de IA e incorpore diversidade. Ao utilizar dados diversos e representativos, monitorar quanto ao viés e discriminação, usar IA explicável e interpretável, considerar a equidade no design e desenvolvimento de sistemas de IA e envolver stakeholders, os sistemas de IA ganham em eficácia e equidade.

Confiança

Existem várias características necessárias para que os sistemas de IA sejam confiáveis. Devem ser projetados e desenvolvidos com responsabilidade e transparência, com base em princípios científicos e de engenharia sólidos, e seus resultados devem ser precisos e consistentes.

Também é necessário pensar na segurança - os sistemas de IA precisam ser capazes de resistir e se recuperar de eventos inesperados ou ataques; outras necessidades são a explicabilidade e a interpretabilidade - os processos e resultados de tomada de decisão devem ser compreensíveis para os seres humanos.

E tem mais: é necessário que esses sistemas priorizem e protejam a privacidade de indivíduos e organizações. Não menos importante no desenvolvimento é o fator justiça - os sistemas de IA nunca devem discriminar indivíduos ou grupos com base em características como raça, gênero, idade, entre outras.

Criar uma inteligência artificial confiável requer equilibrar cada uma dessas características com base no contexto de uso do sistema de IA. Embora sejam atributos de sistemas sociotécnicos, a responsabilidade e a transparência também se relacionam com os processos e atividades internas a um sistema de IA e seu contexto externo. Negligenciar essas características pode aumentar a probabilidade e a magnitude de consequências negativas.

Redução de riscos

Como vimos no início deste guia, sistemas de IA explicáveis e interpretáveis oferecem informações que ajudarão os usuários finais a entenderem os seus propósitos e impactos em potencial, fator que contribui na construção de confiança no sistema e garante que ele seja usado de maneira apropriada.

O risco decorrente da falta de explicabilidade pode ser gerenciado descrevendo-se como os sistemas de IA funcionam, com atenção às diferenças individuais, como o papel, conhecimento e nível de habilidade do usuário. Isso pode ajudar a garantir que os usuários compreendam as capacidades e limitações do sistema.

Sistemas explicáveis podem ser depurados e monitorados com mais facilidade, e eles se prestam a uma documentação, auditoria e governança mais completas. Dessa forma, há mais garantias de que o sistema esteja operando conforme o pretendido e que quaisquer problemas sejam identificados e resolvidos rapidamente.

Para aperfeiçoar a integração da IA no setor público, é imperativo que os sistemas não apenas operem com eficiência, mas também sejam transparentes em sua operação e decisões, assegurando uma maior confiança e responsabilidade em sua implementação.

Entenda o Core do AI RMF

Antes de seguirmos, é preciso compreender o Core do AI RMF, ou seja, a forma como se dá sua estrutura. O core do AI RMF oferece uma abordagem estruturada para gerenciar riscos associados aos sistemas de IA. Além disso, é flexível e adaptável a diferentes tipos de sistemas de IA e contextos organizacionais.

É composto por seis etapas, projetadas para auxiliar as organizações a identificar e gerenciar riscos associados aos sistemas de IA ao longo de todo o seu ciclo de vida, desde o desenvolvimento até a sua desativação. São elas:

- **Categorização dos sistemas de IA e dados**
- **Seleção e personalização de controles do AI RMF**
- **Implementação dos controles do AI RMF**
- **Avaliação dos controles do AI RMF**
- **Autorização da operação do sistema de IA**
- **Monitoramento e comunicação do status do sistema de IA**

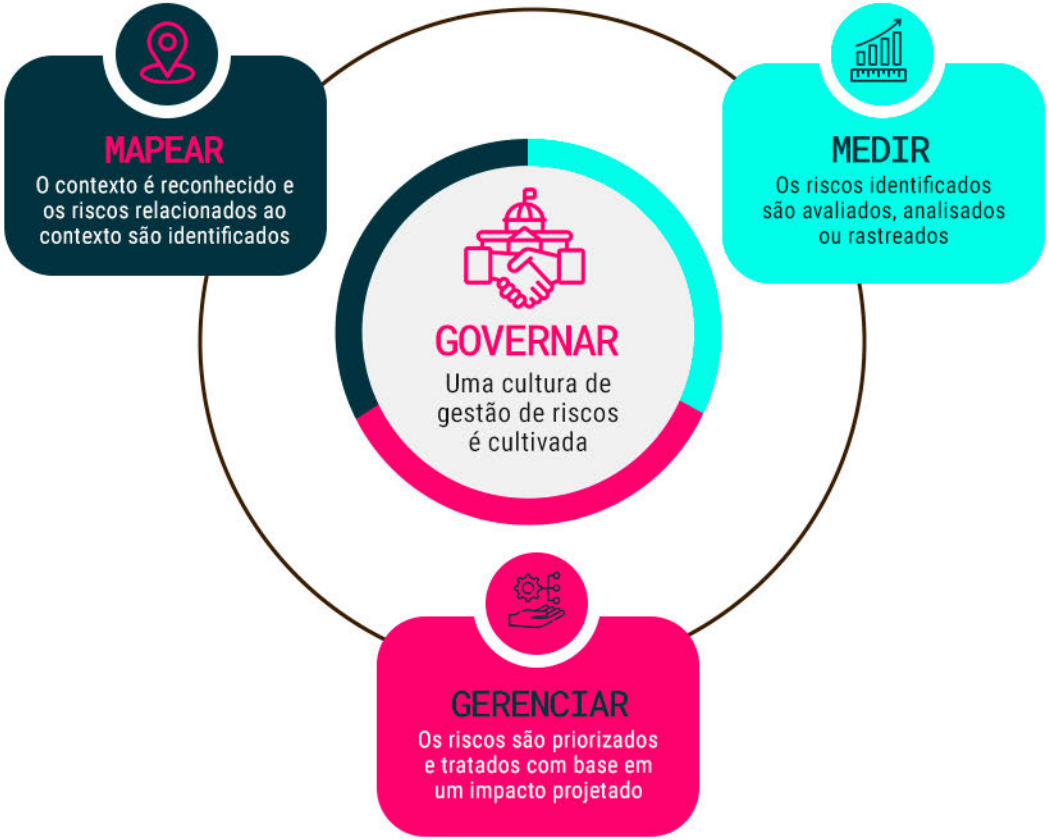
Funções do Core AI RMF

O Core do AI RMF é composto por quatro funções: GOVERNAR (GOVERN), MAPEAR (MAP), MEDIR (MEASURE) e GERENCIAR (MANAGE). Cada uma destas funções é dividida em categorias e subcategorias, por sua vez subdivididas em ações específicas e resultados.

A função GOVERNAR é transversal e está presente em todas as outras três funções. A função MAPEAR identifica e caracteriza os sistemas de IA e seus componentes, bem como os riscos associados a eles.

A função MEDIR avalia e traz uma valoração dos riscos e controles de IA, além de monitorar e relatar o desempenho do sistema de IA. A função GERENCIAR envolve a implementação e manutenção dos controles de gestão de risco da IA, bem como a resposta a incidentes e mudanças no sistema de IA ou no seu ambiente.

Para enfrentar os desafios e potenciais riscos da implementação de sistemas de IA, **é crucial que os servidores públicos sejam capazes de compreender com clareza os processos envolvidos.** Isso vai garantir que os riscos sejam minimizados, além de amplificar os benefícios dessa tecnologia com responsabilidade e eficiência. Saiba mais sobre cada uma das funções:



Governar

A função GOVERNAR está presente em todo o gerenciamento de riscos de IA e habilita as outras funções do processo. Inclui monitoramento contínuo e revisão periódica do processo de gestão de riscos e dos seus resultados. O objetivo da função GOVERNAR é garantir que a gestão de riscos de IA seja contínua, oportuna e realizada em todas as dimensões do ciclo de vida do sistema de IA.

Mapear

A função MAPEAR foi projetada para aprimorar a capacidade de uma organização de identificar riscos. Essa função envolve a coleta de conhecimento contextual e consciência dos riscos dentro dos contextos identificados, o que permite a prevenção de riscos.

A implementação dessa função evolui ao incorporar perspectivas de uma equipe interna diversificada, integrada à equipe externa que desenvolveu e implementou o sistema de IA. Essa função também se destina a ajudar as organizações a prevenir riscos proativamente e a desenvolver sistemas de IA mais confiáveis.

Medir

A função MEDIR calcula a eficácia dos processos e controles de gestão de riscos implementados durante a função MAPEAR. Envolve a identificação de métodos e métricas apropriados para medir os riscos de IA e atualiza regularmente essas métricas, por meio da contribuição de especialistas internos e externos.

Essa função ajuda as organizações a melhorar continuamente os seus processos e controles de gestão de riscos e a garantir que os seus sistemas de IA permanecem confiáveis.

Gerenciar

A função GERENCIAR ajuda as organizações a gerenciar os riscos dos sistemas de IA implantados e a alocar recursos de gerenciamento de riscos, e também inclui processos para avaliar riscos emergentes e mecanismos para melhoria contínua.

Utiliza informações contextuais recolhidas a partir de consultas especializadas e contribuições da IA estabelecidas na função GOVERNAR e realizadas em MAPEAR para diminuir a probabilidade de falhas e de impactos negativos.

/ resumo das funções do Core AI RMF /

Govern (Governança)					
1. Políticas e práticas de risco de IA da organização estão estabelecidas, claras e eficazes	2. Equipes e indivíduos estão designados, treinados e responsáveis pelo tratamento de riscos de IA	3. Diversidade e inclusão são enfatizadas na gestão de riscos de IA durante todo o ciclo de vida	4. Equipes priorizam uma cultura voltada para a conscientização de riscos de IA	5. Processos garantem forte envolvimento com os principais atores de IA	6. Políticas abordam riscos de IA provenientes de softwares de terceiros e questões da cadeia de fornecimento
Map (Mapear)					
1. Contexto é compreendido	2. Sistema de IA é categorizado	3. Capacidades e objetivos de IA são comparados com padrões	4. Riscos e benefícios, incluindo de terceiros, são mapeados	5. Impactos de IA em várias entidades são definidos	
Measure (Medir)					
1. Métodos e métricas são aplicados	2. Sistemas de IA são avaliados quanto à confiabilidade	3. Mecanismos rastreiam riscos de IA ao longo do tempo	4. Feedback sobre a eficácia da medição é avaliado		
Manage (Gerenciar)					
1. Riscos de IA das funções MAP e MEASURE são gerenciados	2. Estratégias maximizam benefícios da IA e minimizam negativos com input de atores de IA	3. Riscos e benefícios de IA de terceiros são gerenciados	4. Tratamentos de risco e planos de comunicação para riscos de IA são documentados e monitorados		

0 1 01

Gestor público, prepare-se para liderar com a IA como aliada!

A emergência e rápida evolução da Inteligência Artificial (IA) nos últimos anos trouxe avanços sem precedentes em diversas áreas, desde saúde e finanças até governança e planejamento urbanístico. No entanto, com as inúmeras oportunidades, surgem também desafios significativos, especialmente quando consideramos a implementação e gestão desta tecnologia em contextos públicos.

Portanto, é essencial que você, servidor público, tenha conhecimentos e domine ferramentas para navegar neste cenário em constante mudança. É imperativo adotar uma abordagem estruturada e informada para gerenciar os riscos associados à IA.

Como vimos, o AI RMF oferece uma estrutura robusta e flexível que pode ser adaptada às necessidades específicas de diferentes setores e aplicações, garantindo que os benefícios desta tecnologia revolucionária sejam maximizados, enquanto os riscos são adequadamente gerenciados.

É essencial o investimento contínuo na educação e formação dos servidores públicos, garantindo que estejam preparados para liderar a próxima fase da revolução da IA.

<<<

1 01

CENTRO DE LIDERANÇA PÚBLICA

Microsoft