

# CIAM

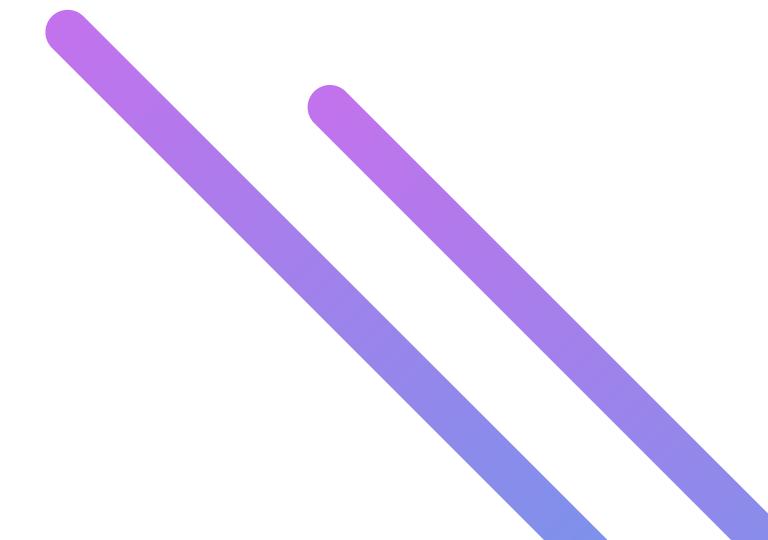
Centro de Inteligência Artificial e Aprendizado de máquina

# Apresentação

Imagine uma situação, em que dados essenciais para a vantagem competitiva da empresa, são acidentalmente expostos devido ao uso inadequado de um sistema de IA. Esse não é um risco teórico, mas uma realidade iminente no cenário tecnológico atual. Ao explorar o contexto jurídico do uso da IA generativa, é crucial compreender que, sem a adoção de medidas específicas que discutiremos hoje, os riscos de violação de dados, responsabilidades legais e danos à reputação são significativos. Vamos começar entendendo os principais desafios e como podemos mitigá-los de forma eficaz.

SEPARATE

HARISOS?





"A **Samsung Electronics** proibiu o uso por funcionários de ferramentas populares de IA generativa, como ChatGPT, após descobrir que a equipe carregou um código confidencial na plataforma, causando um revés na disseminação dessa tecnologia no local de trabalho." (Bloomberg, 2 de maio de 2023)

A inteligência artificial generativa está se tornando uma parte central das operações empresariais, oferecendo novas maneiras de automatizar processos, gerar conteúdos e personalizar experiências. No entanto, essa tecnologia emergente também traz desafios significativos que precisam ser gerenciados de forma proativa para evitar riscos à privacidade, segurança e conformidade legal.

# Riscos Potenciais

CONFABULAÇÕES

SUPERDEPENDÊNCIA

VAZAMENTOS



# Conformidade

A conformidade com a legislação, como a Lei Geral de Proteção de Dados (LGPD), Código de Defesa do Consumidor, Código Civil, Marco Civil da Internet e futuras legislações específicas de IA, é crucial para evitar sanções e proteger a reputação da empresa. Implementar práticas de **governança** e **políticas** claras desde o início ajuda a garantir que o uso de IA generativa esteja alinhado com as exigências legais.

# Objetivos da apresentação

01

## Riscos da IA Generativa

O primeiro objetivo desta apresentação é aprofundar o entendimento dos riscos específicos associados à adoção de IA generativa. Esses riscos incluem desde a geração de informações imprecisa.

02

## Explorar Estratégias de Prevenção

Vamos explorar as melhores práticas para prevenir e mitigar esses riscos, incluindo a implementação de salvaguardas tecnológicas e políticas organizacionais.

03

## Alinhar o Uso da IA com a Lei

Outro objetivo é garantir que a implementação da IA generativa esteja em conformidade com regulamentações como a LGPD e o Código Civil.

04

## Roadmap para adoção segura

A apresentação visa a fornecer um roteiro prático para a adoção segura e eficaz da IA generativa nas operações empresariais.

# Confabulação em IA Generativa

## Definição

Confabulação ocorre quando uma IA generativa, como um assistente virtual, cria informações falsas ou enganosas que parecem verdadeiras, mas não têm base real. Exemplo,: orientações técnicas erradas sobre um material refratário, baseando-se em suposições em vez de dados verificados.

## Impactos

Confabulações de IA podem causar sérios problemas, afetando a qualidade dos produtos e as relações comerciais. Decisões incorretas podem comprometer a integridade dos produtos e a segurança dos trabalhadores.,

## Dificuldades

Identificar confabulações é difícil, especialmente em setores que exigem precisão. É essencial ter validação humana e monitoramento para detectar erros antes que causem problemas, garantindo que as respostas da IA sejam revisadas por especialistas antes do uso.

# Estratégias de Prevenção

Para reduzir os riscos associados às confabulações, a Shinagawa deve adotar práticas rigorosas de controle de qualidade e supervisão das saídas da IA. Isso inclui o treinamento contínuo dos modelos com dados atualizados e revisados por especialistas, e a limitação do escopo das respostas da IA a informações que tenham sido validadas e documentadas. Além disso, a empresa pode estabelecer protocolos onde a IA só interaja diretamente com usuários em áreas de baixo risco, garantindo que em questões críticas, as respostas sejam sempre validadas por um profissional qualificado. Por exemplo, em consultas técnicas, a IA pode ser programada para fornecer apenas informações baseadas em documentos técnicos previamente aprovados pela equipe de engenharia da empresa.

# Privacidade de Dados

Vejamos os principais desafios e medidas de proteção



01

## Risco de Vazamento

A IA pode expor dados confidenciais, como informações financeiras, durante o treinamento. Por exemplo, ao gerar relatórios, pode incluir detalhes sobre clientes ou processos proprietários, comprometendo a segurança.

02

## Desanonimização de Dados

Mesmo com dados anonimizados, a IA pode gerar saídas que, combinadas com outras fontes, revelem informações sensíveis. Por exemplo, ao analisar padrões de compra, pode expor dados estratégicos sobre fornecedores ou clientes.

03

## Compliance

O uso de IA deve estar em conformidade com a LGPD e com o Código Civil, assegurando o tratamento legal dos dados e a possibilidade de remoção ou anonimização, como ao personalizar ofertas ou otimizar processos.

# Estratégias de Prevenção

Para mitigar riscos à privacidade, a Shinagawa deve adotar técnicas de minimização de dados, utilizando apenas as informações estritamente necessárias para o treinamento e operação de suas IAs. Isso inclui a exclusão de informações identificáveis ou a aplicação de técnicas avançadas de anonimização antes do uso em modelos de IA. Por exemplo, ao treinar uma IA para otimizar a produção de materiais refratários, a empresa pode optar por utilizar dados sintetizados ou anonimizados, reduzindo o risco de vazamento de informações sensíveis enquanto ainda obtém *insights* valiosos para suas operações.

# Integração Humano-IA

## SUPERDEPENDÊNCIA

Confiar excessivamente nas saídas da IA sem revisão humana pode levar a erros e vieses, afetando decisões críticas como a escolha de materiais e processos.

## VALIDAÇÃO

A intervenção humana é essencial para garantir que as recomendações da IA estejam alinhadas com os padrões técnicos e de qualidade, evitando compromissos que possam afetar a produção.

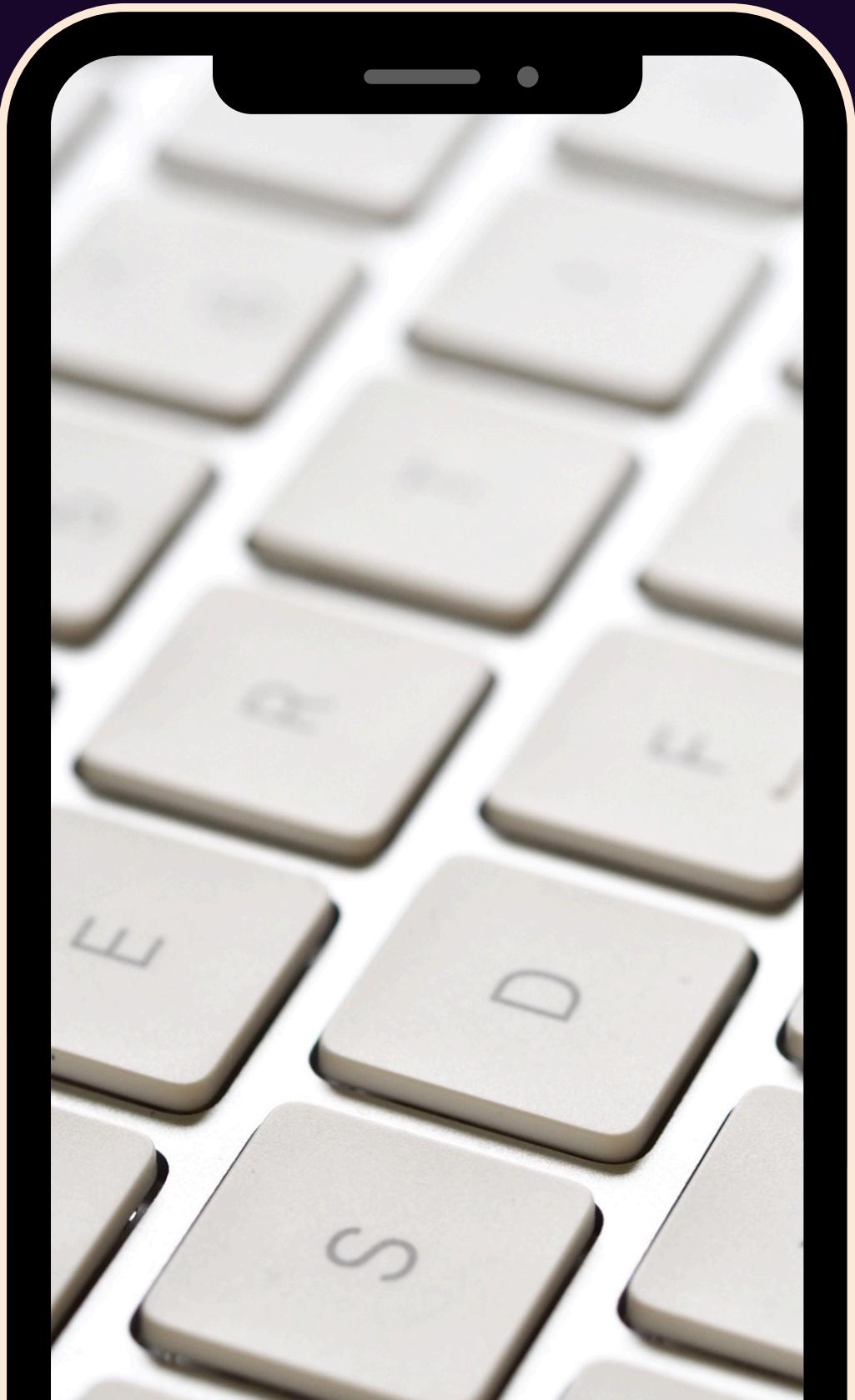
## DESAFIOS

Para integrar a IA de forma eficaz, a Shinagawa deve desenvolver processos e treinamentos que capacitem os funcionários a usar a IA como uma ferramenta complementar, garantindo a qualidade nas operações.

# Estratégias de Prevenção

Embora a intervenção humana seja necessária, ela também pode introduzir vieses, especialmente se os operadores não forem treinados para identificar e mitigar esses riscos. A integração Humano-IA na Shinagawa deve incluir mecanismos para monitorar e corrigir possíveis influências negativas decorrentes da interação humana.

Por exemplo, se um supervisor ajustar as recomendações da IA com base em preferências pessoais, pode introduzir um viés que comprometa a eficiência ou a qualidade do processo de produção.



# Segurança da Informação

- Vulnerabilidades Específicas da IA Generativa: A IA generativa utilizada na Shinagawa apresenta vulnerabilidades que podem ser exploradas por agentes maliciosos, como a injeção de prompts maliciosos. Por exemplo, um hacker poderia manipular os comandos inseridos na IA para gerar código de software, resultando em programas inseguros ou com backdoors, comprometendo a segurança dos sistemas utilizados na produção de refratários.

- Data Poisining

Um dos maiores riscos para a segurança da informação na Shinagawa é o envenenamento de dados, onde atacantes inserem dados manipulados no treinamento da IA. Isso pode levar a saídas incorretas ou prejudiciais, como a geração de parâmetros de produção inadequados, afetando a qualidade dos produtos e a segurança das operações.

- Acessos não autorizados

Devido ao manuseio de grandes volumes de dados sensíveis, é crucial que a Shinagawa implemente controles rigorosos de acesso para proteger esses dados. Medidas como criptografia, autenticação multifator e monitoramento contínuo são essenciais. Por exemplo, ao usar IA para gerar relatórios técnicos ou financeiros, apenas pessoal autorizado deve ter acesso e controle sobre esses dados, garantindo a segurança da informação.

# Estratégias de Prevenção

A Shinagawa deve realizar auditorias e revisões de segurança regulares em seus sistemas de IA generativa para identificar e corrigir vulnerabilidades. Essas auditorias garantem que práticas de segurança estão sendo seguidas e que o sistema está em conformidade com as regulamentações. Por exemplo, auditorias periódicas em sistemas de IA que suportam o desenvolvimento de novos materiais refratários ajudam a assegurar que o sistema esteja protegido contra ameaças emergentes e continue operando de forma segura.

# Propriedade Intelectual

Desafios e Estratégias em razão do uso da IA Generativa



01

## Desafios

A criação de novos conteúdos por IA gera complexidade na determinação de quem detém os direitos, especialmente quando as criações podem se assemelhar a obras já existentes.

02

## Risco de Violação

A IA pode inadvertidamente criar produtos ou marcas que infringem direitos autorais ou marcas registradas, colocando a Shinagawa em risco de disputas legais.

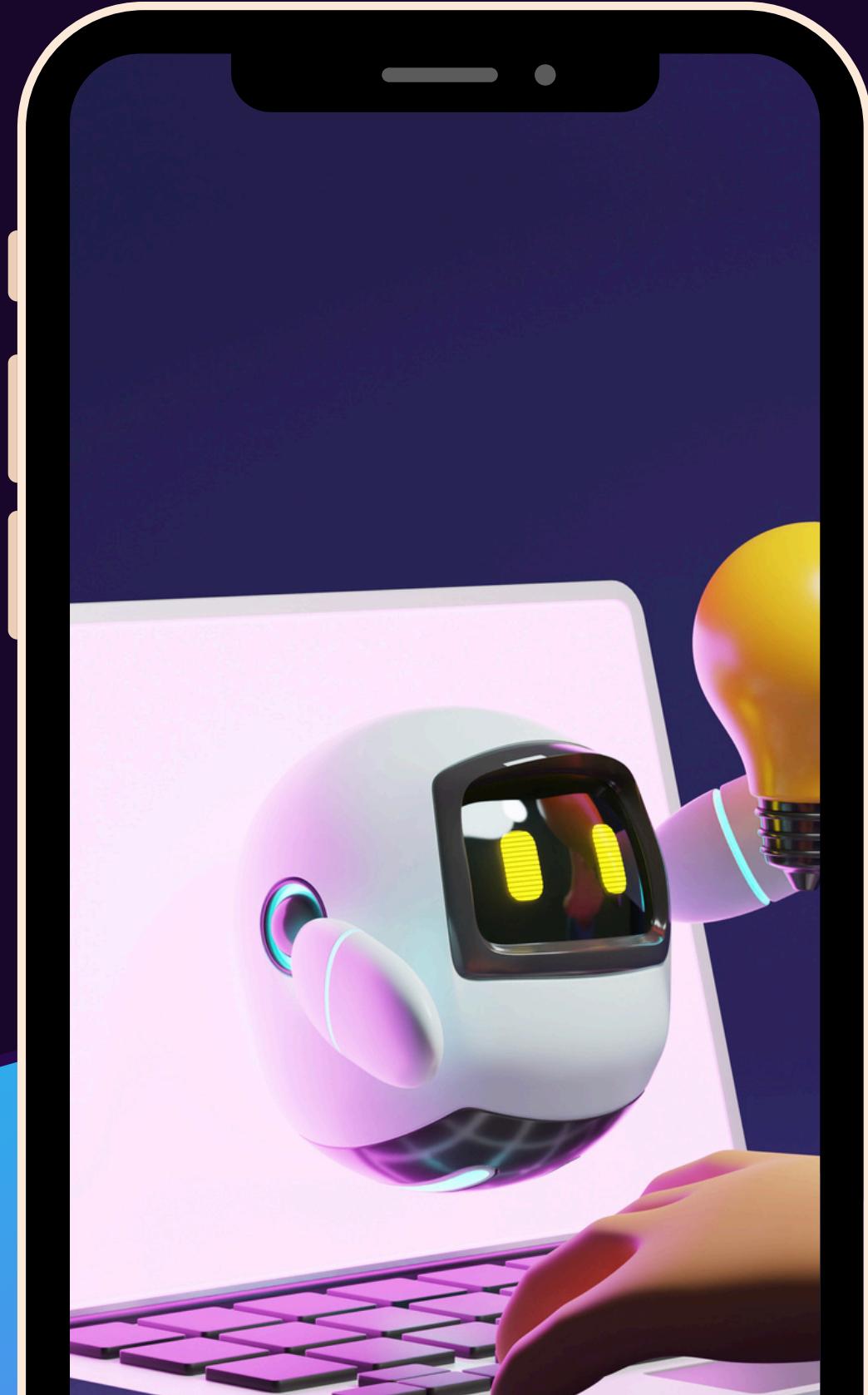
03

## Modelos Jurídicos

A evolução do uso de IA na Shinagawa exige novos modelos legais para lidar com as lacunas na proteção de propriedade intelectual, dada a ausência de regras claras em muitas jurisdições. Atualmente, está sendo discutido o PL 2338/23, PL 303/24 (Patente requerida em nome da IA) e 1473/2023 (restrição de uso de obras por IA).

# Estratégias de Prevenção

A Shinagawa deve adotar salvaguardas rigorosas e estabelecer políticas claras sobre o uso de IA em processos criativos, além de incluir cláusulas específicas em contratos para proteger as criações da empresa.



# Cadeia de Valor e Integração de Componentes

A criação e implementação de sistemas de IA generativa frequentemente dependem de componentes fornecidos por terceiros, como modelos pré-treinados e bibliotecas de código. Essa dependência pode introduzir vulnerabilidades, especialmente se os componentes não forem devidamente avaliados quanto à segurança e conformidade. Por exemplo, ao utilizar um modelo de IA pré-treinado de um fornecedor externo, a Shinagawa pode, sem perceber, incorporar vieses ou falhas de segurança em seus sistemas de produção de refratários.

- **Riscos**

A complexidade da cadeia de valor de IA na Shinagawa pode dificultar a rastreabilidade da origem e qualidade dos componentes utilizados. A falta de transparência pode levar a problemas de conformidade e aumentar o risco de integrar componentes inseguros ou não éticos. Por exemplo, se a Shinagawa não conseguir verificar a origem dos dados usados para treinar um modelo de IA, pode estar utilizando dados obtidos de forma ilegal ou não ética.

- **Integração**

A integração de componentes de IA em sistemas de produção pode ser complexa e suscetível a falhas de compatibilidade, resultando em erros no sistema de IA da Shinagawa. Por exemplo, ao integrar bibliotecas de código e modelos em uma plataforma de IA para otimizar processos, podem surgir conflitos de versões ou problemas de interoperabilidade, comprometendo a eficácia e segurança do sistema.

# Estratégias de Prevenção

Para mitigar os riscos associados à cadeia de valor e à integração de componentes, a Shinagawa deve adotar uma abordagem proativa de gerenciamento de riscos. Isso inclui auditorias regulares dos fornecedores, critérios rigorosos de avaliação para componentes de terceiros e processos de verificação antes da implementação. Por exemplo, a Shinagawa pode estabelecer um programa de auditoria contínua para garantir a segurança e integridade dos componentes de IA utilizados em suas operações industriais.



## AÇÕES PARA GERENCIAR RISCOS

**Identificação e Avaliação de Riscos:** O primeiro passo é identificar e avaliar sistematicamente os potenciais riscos que a tecnologia pode introduzir. Isso envolve a análise detalhada de como a IA será usada, os dados que ela processará e as saídas que ela gerará, bem como os impactos potenciais em diferentes áreas da organização. Por exemplo, uma empresa que utiliza IA para analisar dados de clientes deve avaliar os riscos de privacidade e segurança, identificando possíveis pontos de vulnerabilidade onde dados sensíveis possam ser expostos.

## Implementação de Controles

A próxima etapa é implementar controles e salvaguardas que minimizem esses riscos. Isso pode incluir o uso de técnicas de criptografia para proteger dados sensíveis, a aplicação de filtros para impedir a geração de conteúdo inapropriado, e a realização de revisões humanas das saídas da IA.



# Monitoramento

Gerenciar riscos de forma eficaz requer um monitoramento contínuo dos sistemas de IA para identificar novos riscos ou vulnerabilidades que possam surgir ao longo do tempo. Isso inclui a realização de auditorias regulares para garantir que os controles implementados estejam funcionando conforme o esperado e que o sistema esteja em conformidade com as regulamentações. Por exemplo, na produção de materiais refratários, a Shinagawa deve realizar auditorias periódicas em seus sistemas de IA para garantir que os modelos continuam a operar de maneira eficiente, sem vieses, e em conformidade com as políticas internas e normativas da indústria.



# Capacitação

**Uma parte crucial do gerenciamento de riscos é garantir que todos os colaboradores que interagem com a IA estejam cientes dos riscos associados e saibam como mitigá-los.**

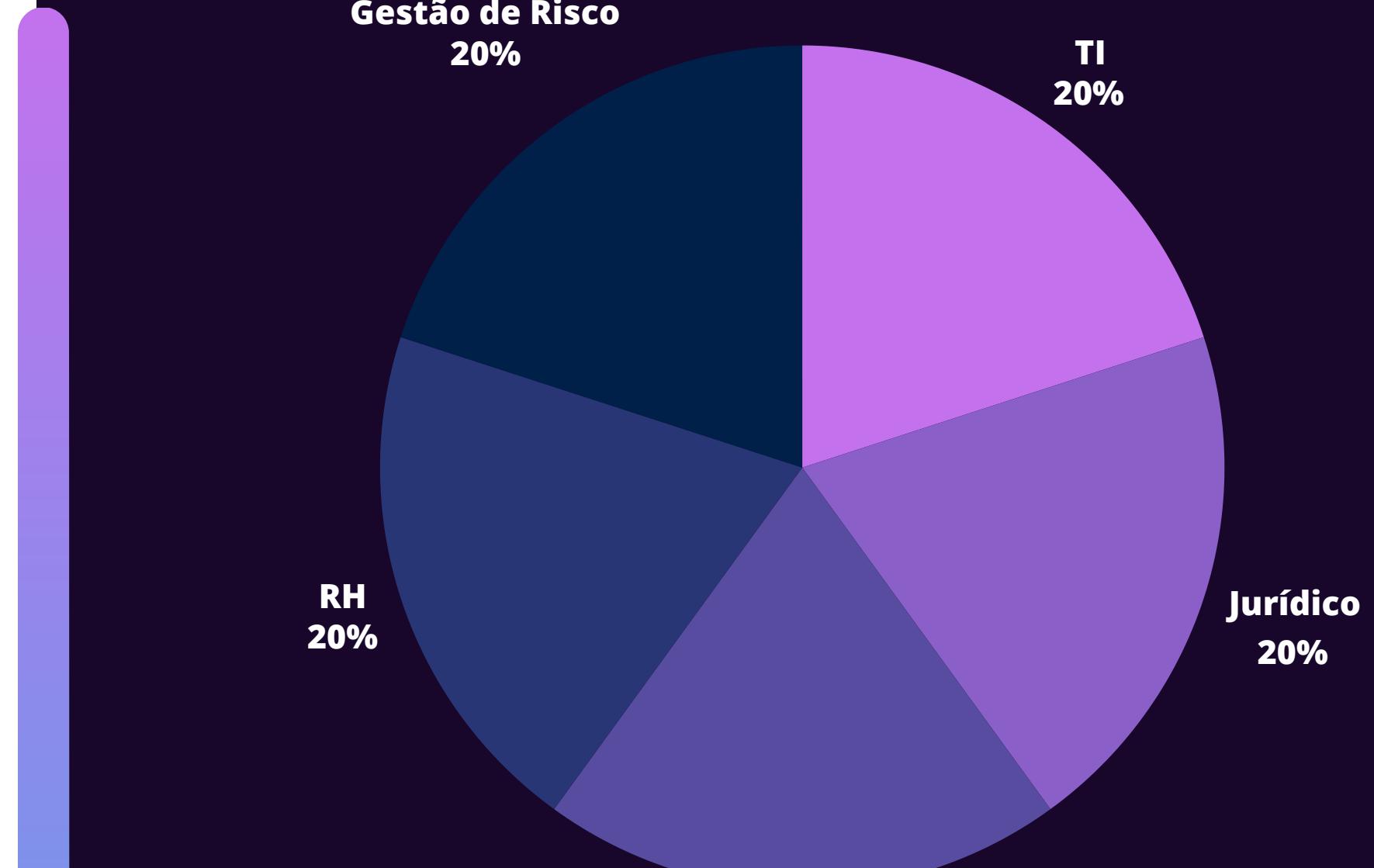


Isso envolve programas de capacitação e sensibilização que educam os funcionários sobre as melhores práticas, políticas de segurança e procedimentos a seguir em caso de incidentes relacionados à IA.

# Governança e Conformidade

## Estabelecimento de Estruturas de Governança:

Uma governança eficaz da IA generativa requer a criação de estruturas organizacionais claras que definam papéis, responsabilidades e processos decisórios. Isso inclui a formação de comitês de governança de IA que supervisionem o desenvolvimento, implementação e operação dos sistemas de IA, garantindo que todas as ações estejam alinhadas com os objetivos estratégicos da empresa e as melhores práticas éticas. Por exemplo, a Shinagawa Brasil pode estabelecer um comitê de governança de IA composto por membros de **TI, jurídico, compliance e recursos humanos** para assegurar que a IA seja utilizada de maneira responsável e segura em todas as suas operações no país.



# Garantia de Conformidade com Regulamentações

Conformidade com as regulamentações brasileiras, como a LGPD (Lei Geral de Proteção de Dados), é fundamental para evitar penalidades legais e proteger a reputação da empresa. A empresa deve implementar políticas e procedimentos que garantam que o uso da IA generativa esteja em conformidade com essas regulamentações. Por exemplo, ao utilizar IA para otimizar processos industriais, a Shinagawa Brasil deve garantir que todos os dados pessoais processados estejam de acordo com as exigências da LGPD, com o devido consentimento dos titulares e práticas rigorosas de privacidade. E, caso os dados sejam de uma PJ, é preciso verificar as cláusulas de confidencialidade em seus contratos.

# Transparência

A governança de IA exige transparência interna e externa. A empresa deve ser capaz de explicar como a IA é utilizada, como as decisões automatizadas são tomadas e como os dados são tratados. Isso inclui a documentação completa dos processos de IA, uma explicação clara dos algoritmos utilizados, e a garantia de que os sistemas de IA possam ser auditados e revisados por terceiros, se necessário. Por exemplo, ao usar IA para otimizar a cadeia de suprimentos, a Shinagawa Brasil deve fornecer explicações claras aos reguladores e stakeholders sobre as decisões tomadas e garantir que existam mecanismos de revisão.





# Responsabilidade

A governança eficaz da IA envolve a atribuição clara de responsabilidades para a supervisão contínua dos sistemas de IA. Equipes ou indivíduos específicos devem ser responsáveis por monitorar o desempenho da IA, identificar e corrigir desvios dos padrões estabelecidos, e garantir que os sistemas de IA continuem a operar de acordo com as normas legais e éticas. Por exemplo, se a IA for usada para análise de desempenho de materiais, uma equipe dedicada deve supervisionar as decisões da IA, garantindo que sejam precisas, justas e em conformidade com os padrões industriais e legais.

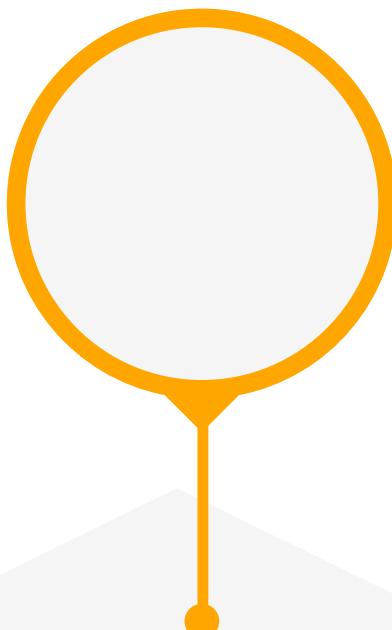
# Políticas de Acesso e Uso

- 1. Definição de Políticas de Acesso Restrito:** apenas pessoal autorizado e treinado deve ter acesso aos sistemas de IA e dados sensíveis, utilizando medidas como autenticação multifator para prevenir acessos não autorizados.
- 2. Controle de Uso de Dados:** Os dados devem ser usados exclusivamente para os fins definidos, garantindo conformidade com a LGPD e proteção da privacidade, sem compartilhamento não autorizado.
- 3. Monitoramento e Auditoria de Acessos:** Implementar monitoramento e auditorias contínuas é essencial para garantir que as políticas de acesso e uso sejam seguidas, detectando e corrigindo violações.
- 4. Sensibilização e Treinamento:** Programas de treinamento regulares garantem que os colaboradores da Shinagawa Brasil entendam e sigam as políticas de acesso e uso, minimizando riscos.

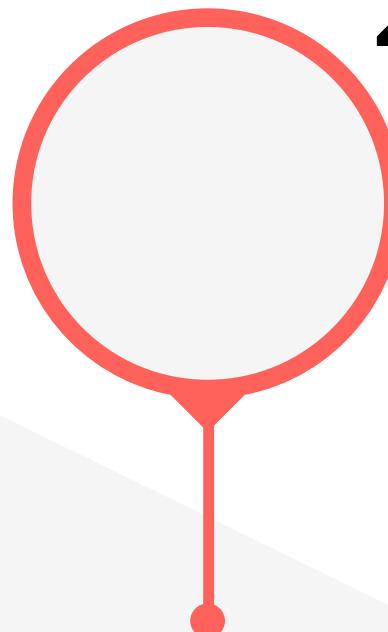
# Inventário de Sistemas de IA

- 1. Manter um Inventário Atualizado:** é essencial manter um inventário atualizado de todos os sistemas de IA em uso, detalhando sua finalidade, dados processados, algoritmos utilizados e responsáveis pela operação. Isso assegura uma gestão eficaz e a identificação de possíveis riscos, permitindo melhorias contínuas.
- 2. Identificação de Riscos e Padrões de Uso:** O inventário de IA é crucial para identificar riscos associados a cada sistema, especialmente aqueles que manipulam dados sensíveis ou realizam decisões críticas. Na Shinagawa, sistemas que gerenciam dados de produção ou realizam análises de segurança devem ser monitorados mais intensivamente devido ao seu impacto.
- 3. Suporte à Conformidade e Auditorias:** Um inventário detalhado facilita a conformidade com regulamentações, permitindo à Shinagawa demonstrar controle sobre o uso e processamento de dados em auditorias regulatórias.
- 4. Revisão e Melhoria Contínua:** Manter um inventário de IA facilita a revisão e a melhoria contínua dos sistemas. Pode-se identificar sistemas que precisam de atualizações ou substituições, além de eliminar redundâncias.

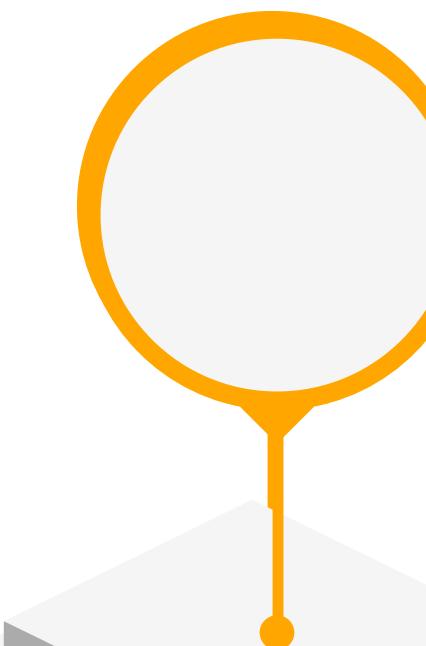
**1. Políticas de Descomissionamento**



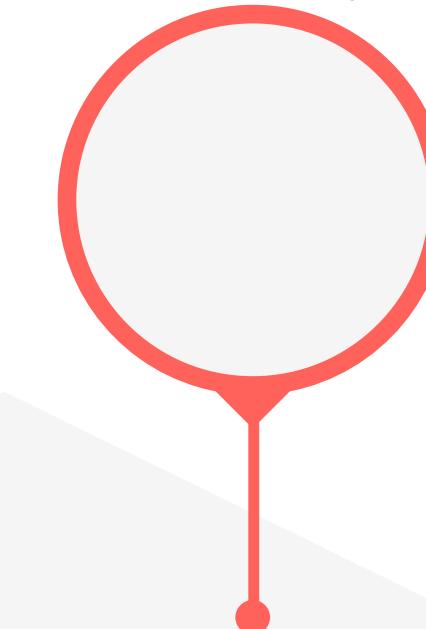
**2. Monitorar Propriedade Intelectual**



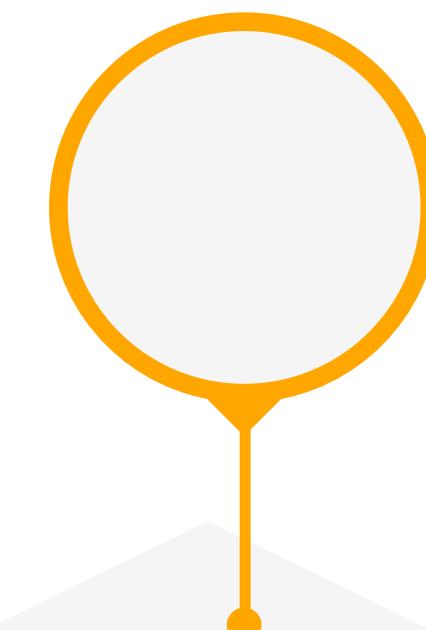
**4. Integrar Feedback Humano**



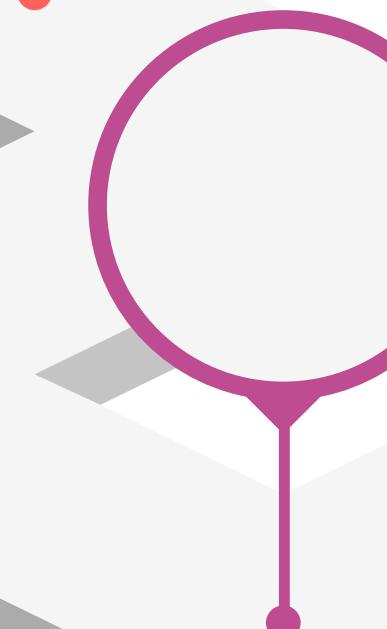
**5. Controle de Versão**



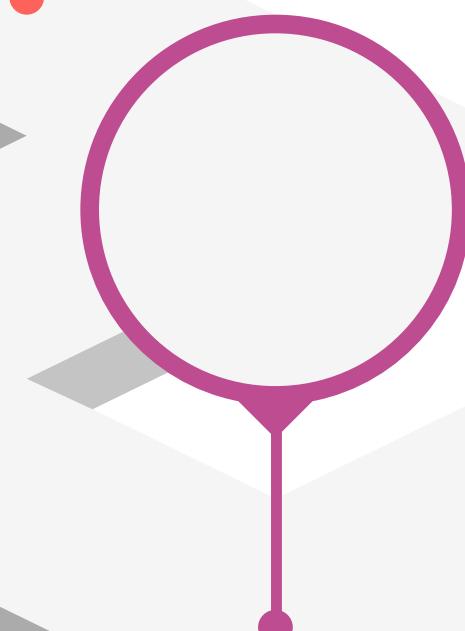
**7. Simular Cenários  
de Risco**



**3. Avaliar e Mitigar Vieses**



**6. Auditoria de Fornecedores Externos**



# Exemplos de IA Generativa com Proteção de Dados

01

## IBM Watson

Conhecido por sua segurança robusta e proteção de dados, o IBM Watson permite que as empresas mantenham controle total sobre seus dados, garantindo que informações confidenciais não sejam expostas.

02

## Hugging Face Hub (Private Models)

Permite que empresas treinem e operem modelos de IA em ambientes isolados, com controles rígidos sobre o acesso a dados e a garantia de que as informações permanecem confidenciais.

03

## Microsoft Azure OpenAI Service

Oferece serviços de IA generativa com opções para manter dados em nuvens privadas, garantindo que os inputs e outputs permaneçam protegidos e não sejam usados para treinar modelos externos.

04

## NVIDIA NeMo e NVIDIA Triton Inference Server

Essa tecnologia oferece uma camada adicional de segurança ao interceptar e gerenciar as interações com a IA generativa. Por exemplo, ela pode prevenir que informações sensíveis sejam divulgadas inadvertidamente durante as interações com o modelo de linguagem, garantindo que as respostas da IA sejam controladas e que informações críticas sejam mantidas em sigilo. Isso é especialmente importante quando se utiliza IA para automatizar processos internos que envolvem dados confidenciais.

## Computação Confidencial

Computação Confidencial é uma tecnologia que protege os dados durante o processamento, mantendo-os criptografados enquanto são usados. Isso é crucial para a IA generativa porque permite que dados sensíveis sejam processados em ambientes de nuvem ou por terceiros sem risco de exposição. Com a computação confidencial, empresas podem utilizar IA para analisar dados altamente sensíveis, como segredos industriais ou informações pessoais, garantindo que esses dados permaneçam seguros e privados, mesmo em cenários de processamento complexo e distribuído.



# CONTATO

<https://c4ai.inova.usp.br/>

candia@usp.br

cristinagodoy@usp.br

enio.alterman.blay@usp.br

+55 11 2648-1695

Inova-USP Building – USP – Universidade de  
São Paulo – Campus Butantã Av. Profº Lúcio  
Martins Rodrigues, 370 – Butantã – São  
Paulo – SP



**<https://c4ai.inova.usp.br/>**

**OBRIGADA!**