DEVPIRA + pecege

Fábrica de Inovação LIMEIRA

IA E CIBERSEGURANÇA

FERNANDO BRYAN FRIZZARIN

QUEM SOU EU

FERNANDO BRYAN FRIZZARIN

- Gerente de Suporte Técnico **BluePex**
- Diretor Técnico e Operacional – **Fábrica de Inovação** - Limeira, SP
- Professor **FATEC** Araras e Americana

FERNANDO BRYAN FRIZZARIN

QUEM SOU EU

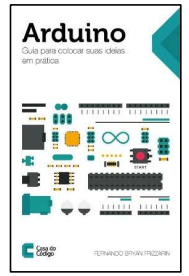
- **Ciência da Computação**, UNIMEP
- **Redes**, UFSCar
- **Psicopedagogia**, UNISAL
- **Inteligência Artificial**, UMSP

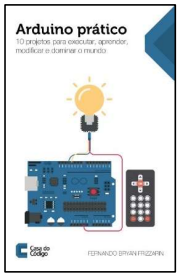
- **Diploma Pérola Byington** – SBO, SP
- **Professor Universitário Inovador** – Limeira, SP
- **Membro Imortal da Acadêmica Mundial de Letras da Humanidade** – Limeira, SP

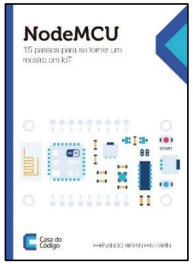
FERNANDO BRYAN FRIZZARIN

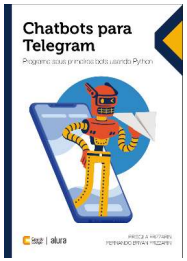
QUEM SOU EU

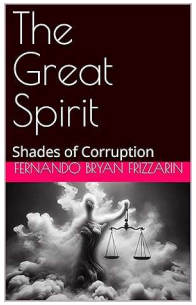
- **Arduino: Guia para colocar suas ideias em prática**, Casa do Código, 2ª edição em 2019
- **Arduino Prático: 10 projetos para executar, aprender, modificar e dominar o mundo**, Casa do Código, 2016
- **NodeMCU: 15 passos para se tornar um mestre em IoT**, Casa do Código, 2019
- **Chatbots para Telegram: Programe seus primeiros bots usando Python**, Casa do Código, 2023.
- **O grande espírito: sombras da corrupção**, Amazon KDP, 2024 (*br* e *us*).





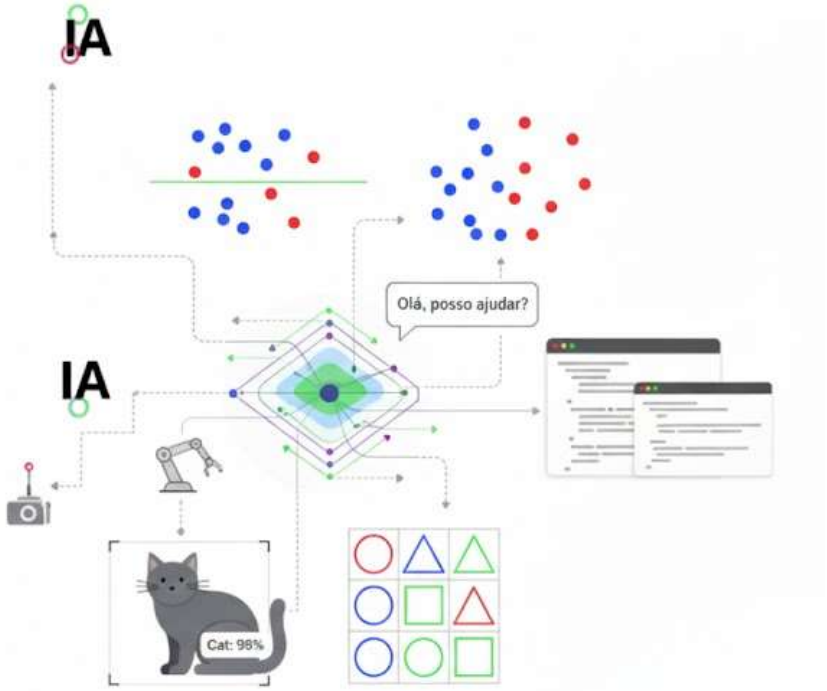






O QUE ESTAMOS CHAMANDO DE IA

- **Machine Learning - “clássico”**
  - Classificadores, detecção de anomalias;
- **LLMs e copilots**
  - ChatGPT, Gemini, GitHub Copilot, etc;
- **Modelos de visão**
  - Reconhecimento de imagem, reconhecimento de padrões;

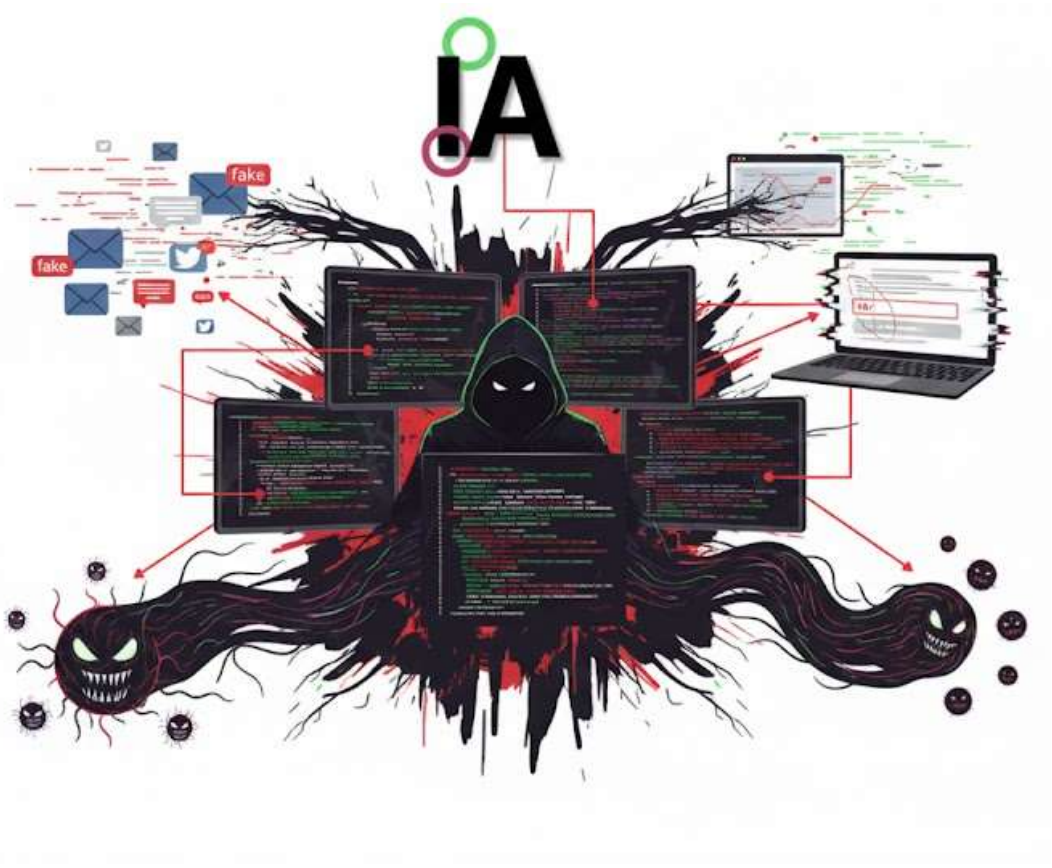


ONDE JÁ TEMOS IA

- Quase todo lugar, de forma quase imperceptível;
- Antivírus/EDR “com IA”;
- Plataformas de SOC e SIEM com correlação automática;
- Ferramentas de dev com sugestão de código;

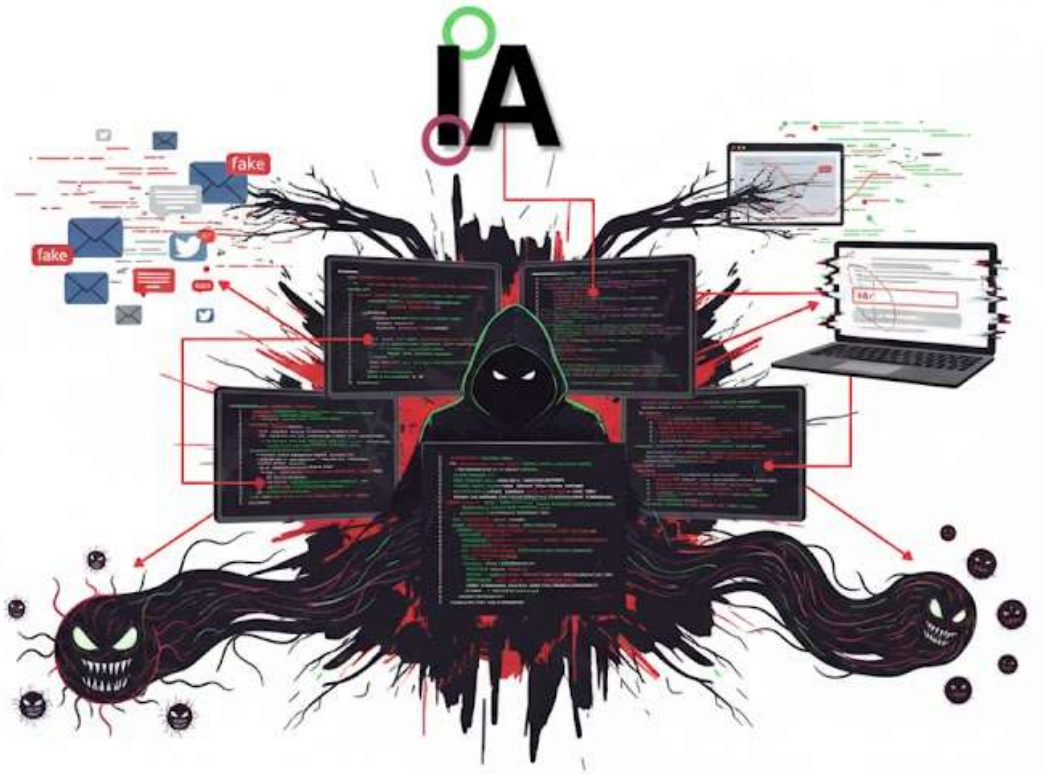
COMO OS ATACANTES USAM IA

- Aqui é a parte “divertida” e assustadora:
  - Automação de phishing e engenharia social;
  - E-mails e mensagens muito bem escritos, personalizados;
  - Geração de páginas falsas e scripts de ataque em massa;
  - Geração e melhoria de malware;
  - IA ajudando na criação de variantes de código malicioso;
  - Obfuscação de scripts e payloads;



COMO OS ATACANTES USAM IA

- Aqui é a parte “divertida” e assustadora:
  - Ataques contra modelos e dados;
  - *Data poisoning*: contaminar dados de treinamento;
  - *Model stealing*: copiar um modelo exposto via API;
  - *Prompt injection* em LLMs integrados a sistemas internos;
  - *Deepfakes* e fraude de identidade;
  - Voz e vídeo falsos usados para enganar helpdesks, suporte, financeiro;



IA TAMBÉM É PROTEÇÃO

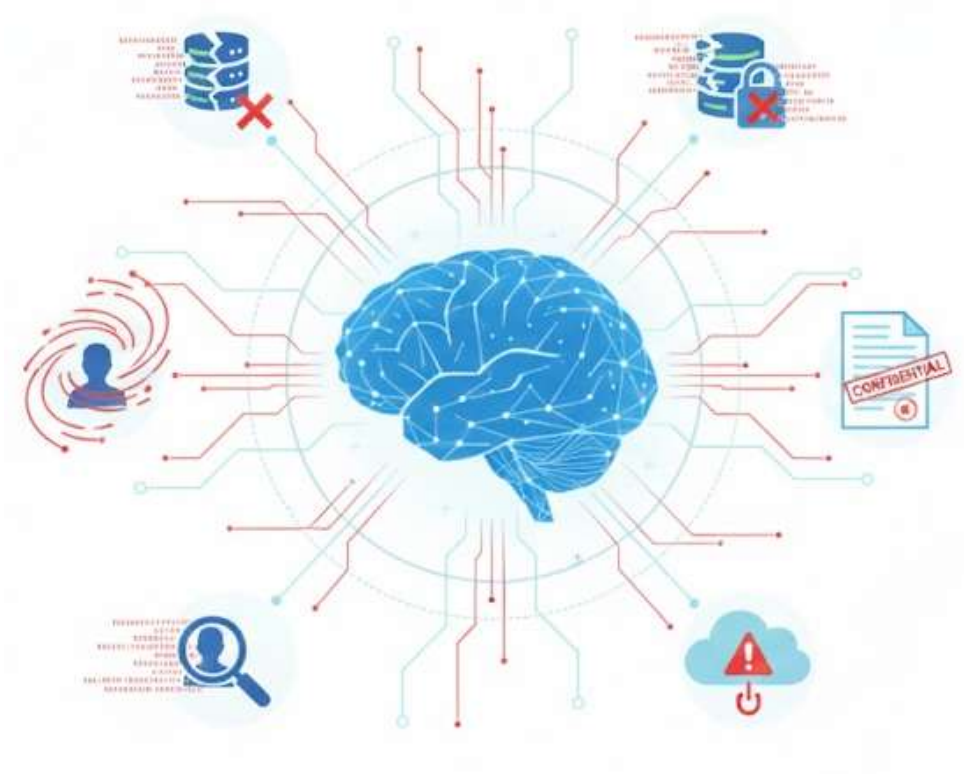
- Detecção de anomalias e comportamentos suspeitos;
- Análise de logs (rede, servidores, aplicações);
- UEBA (*User and Entity Behavior Analytics*);
- EDR/XDR e correlação de eventos;
- IA priorizando alertas, reduzindo ruído;
- Ajuda a encontrar "*agulha no palheiro*";

IA TAMBÉM É PROTEÇÃO

- Automação e SOAR (*Security Orchestration, Automation and Response*);
- Playbooks automatizados:
  bloquear IP,
  isolar máquina, resetar credenciais;
- Chatbots internos para incident response (consultar runbooks, playbooks, etc.);
- IA ajudando o desenvolvedor a escrever código mais seguro
- Sugestão de correções de vulnerabilidades;
- Ferramentas que fazem code review com foco em segurança;

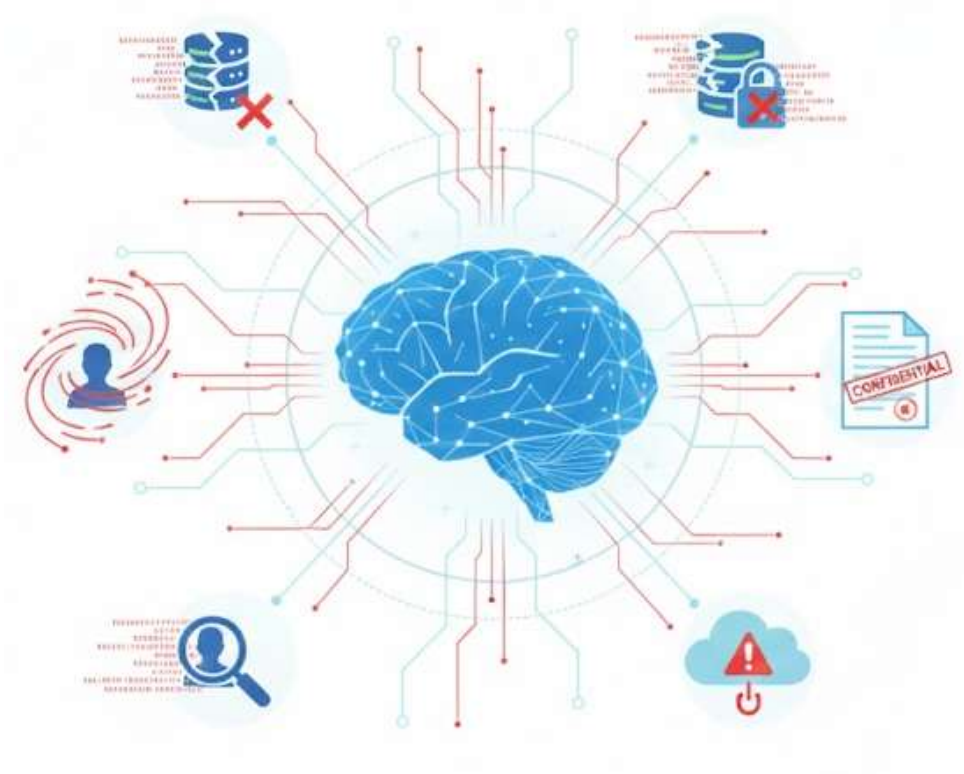
RISCOS TÉCNICOS NO USO DE IA

- Foco direto para devs e técnicos que consomem APIs de IA ou usam copilots;
- Exposição de dados sensíveis;
- Enviar dados de clientes, código proprietário, segredos para serviços externos;
- Questão de LGPD, sigilo, contratos;
- Dependência de APIs externas;



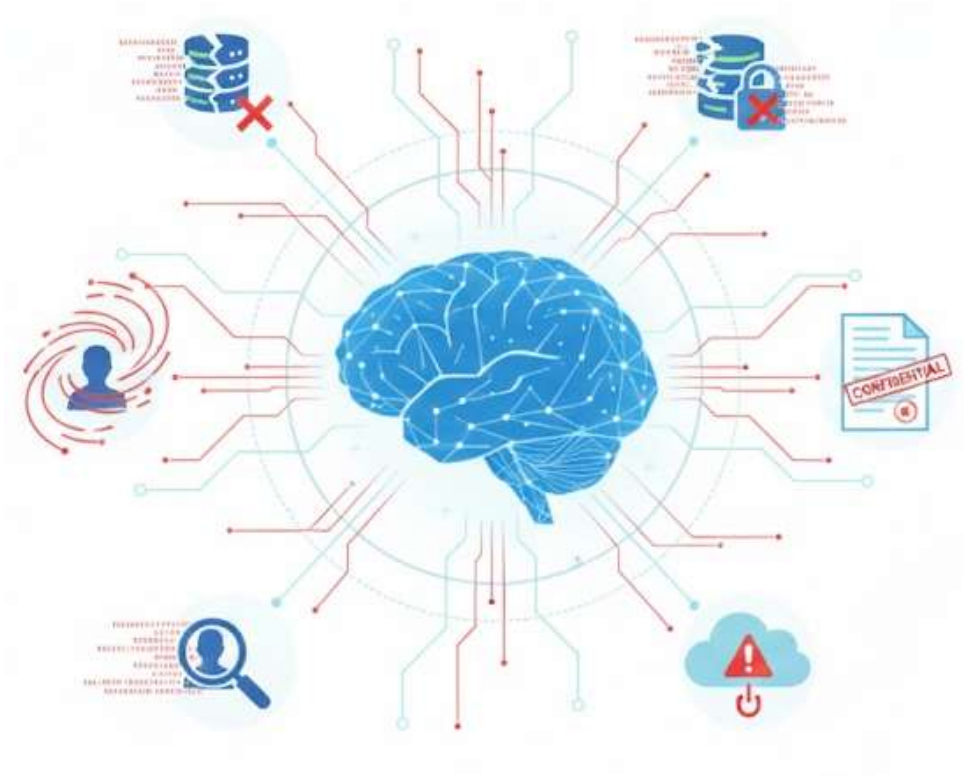
RISCOS TÉCNICOS NO USO DE IA

- Risco de *supply chain* (bibliotecas, SDKs, plugins);
- *Prompt injection* e alucinações;
- Entrada de usuário controlando o comportamento do modelo;
- Respostas "*convencidas e erradas*" sendo tratadas como verdade;
- Governança e auditoria;
- Segurança de API (chaves, *rate limiting*, validação de entrada/saída);



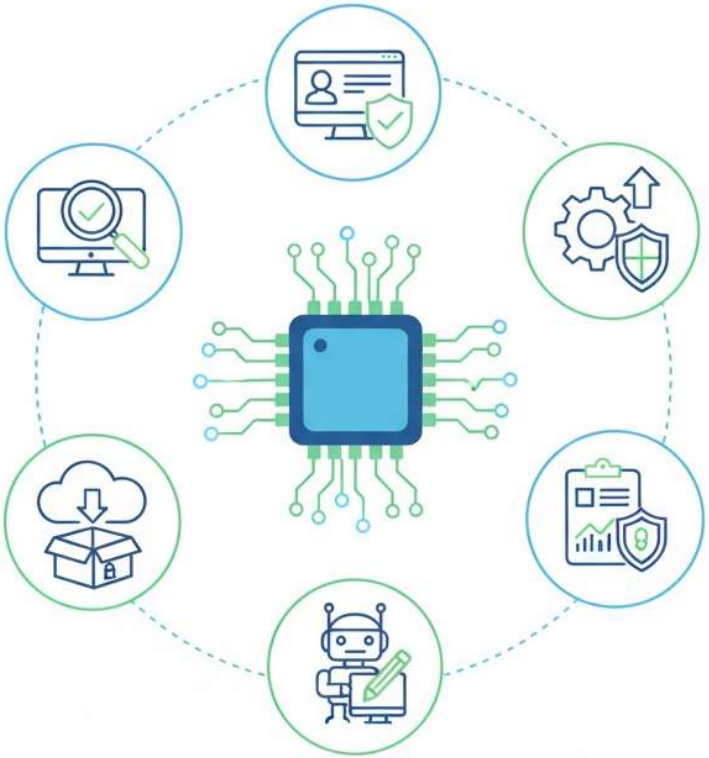
RISCOS TÉCNICOS NO USO DE IA

- Você já está usando IA no seu projeto (código)?
  - Quem aprovou?
  - Existe registro de prompts e respostas usados em decisões críticas?
  - Qual a cerca decisória ou moral usada?



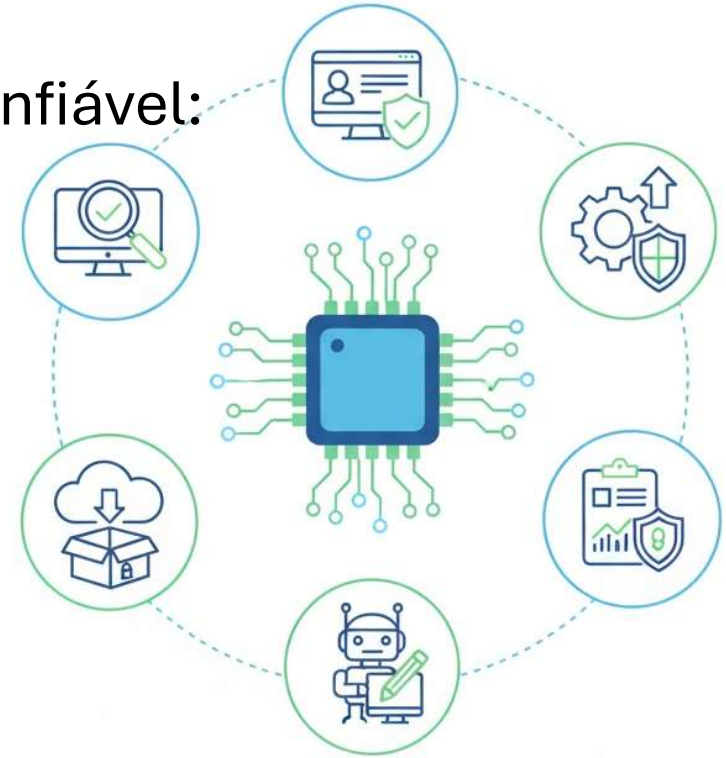
BOAS PRÁTICAS

- Transformar tudo em “*o que eu faço na segunda-feira de manhã*”;
- Uso seguro de *copilots* e assistentes de código;
- Não colar segredos, chaves, dados sensíveis;
- Sempre revisar o código sugerido (*não confiar cegamente*);
- Rodar scanners de segurança (*SAST/DAST/dep-check*) sobre o código gerado;



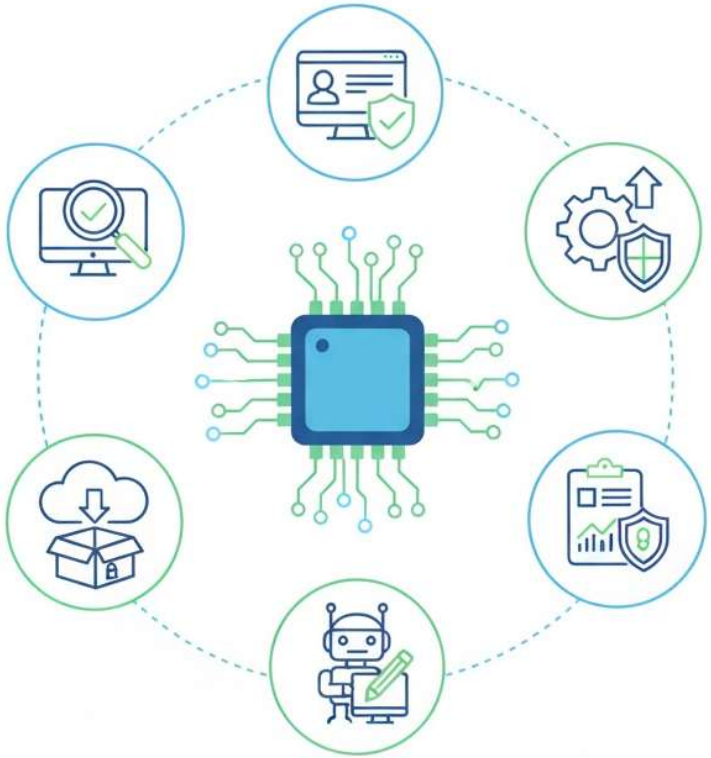
BOAS PRÁTICAS

- Tratar modelo como componente não confiável:
  - Validar toda saída antes de executar ações (ex.: comandos, chamadas a outras APIs);
  - Sanitizar prompts e entradas do usuário;
  - Limitar o “poder” da IA (*principle of least privilege*);
  - Protegendo dados e modelos;
  - Minimização de dados enviados ao modelo;
  - *Pseudonimização/anonimização* quando possível;
  - Controle de acesso forte às APIs de IA;
  - Checklist rápido para o time;



BOAS PRÁTICAS

- Já tem política interna de uso de IA?
- Já definiu o que pode e o que não pode ser enviado?
- Nossas ferramentas de segurança (EDR, firewall, e-mail, etc.) usam IA? Sabemos o que fazem e como configurá-las?



AÇÕES PRÁTICAS

- Definir regras internas de uso de IA;
- Para *devs*, *suporte*, *atendimento*, *marketing*;
- Revisar um sistema crítico sob a ótica de IA;

AÇÕES PRÁTICAS

- Onde IA já está sendo usada (ou poderia ser)?
- Quais riscos adicionais isso cria?
- Capacitar o time técnico;
- Treinos específicos: segurança de APIs, segurança em *IA/ML*, uso seguro de *copilots*.

AÇÕES PRÁTICAS

- Treinos específicos: segurança de *APIs*, segurança em *IA/ML*, uso seguro de *copilots*;

LEMBRETE

FERNANDO BRYAN FRIZZARIN

IA e segurança cibernética não são lados opostos: ou usamos IA com consciência para defender melhor, ou alguém vai usá-la contra nós.