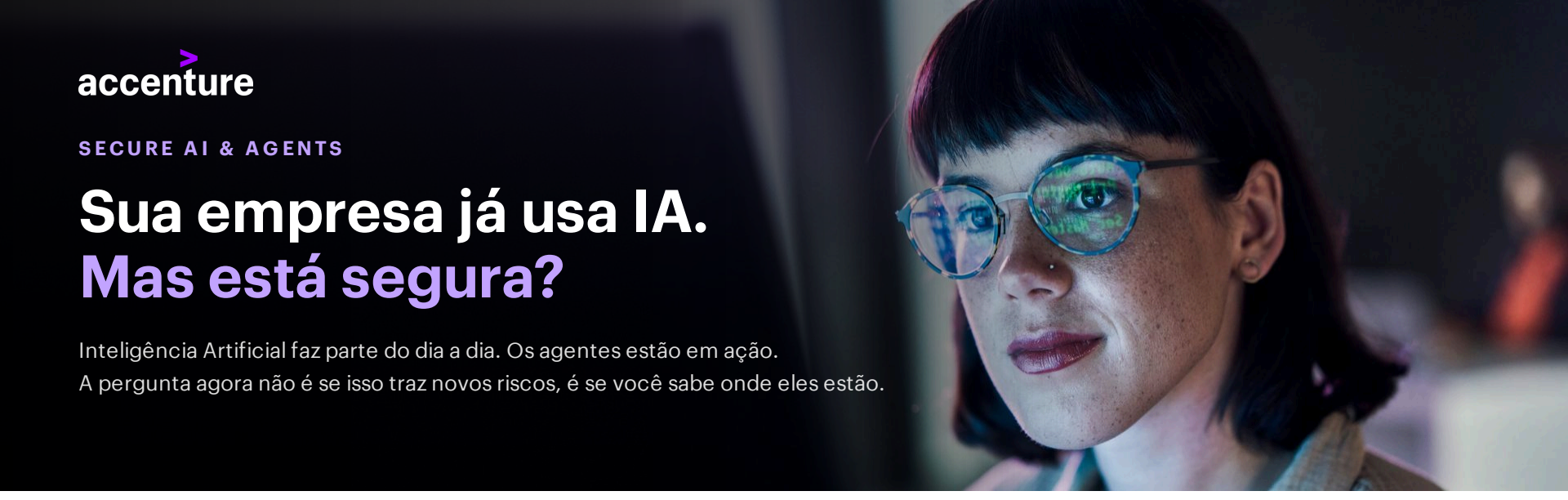

accenture

SECURE AI & AGENTS

Sua empresa já usa IA. Mas está segura?

Inteligência Artificial faz parte do dia a dia. Os agentes estão em ação.
A pergunta agora não é se isso traz novos riscos, é se você sabe onde eles estão.

A IA mudou a superfície de risco

Modelos, agentes, dados de treinamento, prompts e APIs ampliam o ambiente que a segurança precisa proteger.

E trazem riscos específicos:

PROMPT INJECTION
Instruções maliciosas podem alterar o comportamento de sistemas de IA.

EXPOSIÇÃO DE DADOS SENSÍVEIS
Informações confidenciais podem aparecer em prompts, modelos ou respostas.

AGENTES SEM GOVERNANÇA
Mais autonomia exige controle sobre acessos e ações.

SHADOW AI
Ferramentas de IA podem ficar fora da visibilidade e das políticas da empresa.

Sua segurança enxerga tudo isso?

Esses riscos são reais

Casos públicos mostram consequências que vão de exposição de propriedade intelectual a fraude e responsabilidade legal.

SAMSUNG × CHATGPT [1]
Exposição de propriedade intelectual
Shadow AI

ARUP | DEEPFAKE [2]
US$ 25 milhões em perdas
Fraude sintética

AIR CANADA [3]
Responsabilidade legal
Integridade de decisão

PYTORCH / TORCHTRITON [4]
Comprometimento
Identidades não humanas

Sua empresa usa IA ou agentes? → **Sabe quais?** → **Sabe quais dados eles acessam?** → **Sabe o que eles podem fazer?** → **Sabe quais controles os protegem?** → **Consegue identificar acessos ou comportamentos fora do esperado?** → **Se algo der errado, sua empresa sabe como responder?**

Se alguma resposta foi “não”, existe uma lacuna que merece atenção.

Firewalls, controles de acesso, monitoramento e resposta continuam essenciais.

\+

Mas a IA adiciona outros elementos à equação:

- modelos
- agentes
- dados de treinamento
- vetores e embeddings
- prompts e APIs

A IA exige uma camada adicional de proteção. Você só pode proteger o que consegue ver.

DESCUBRA
Onde IA e agentes estão presentes.

AVALIE
Quais riscos e lacunas existem.

PRIORIZE
O que exige ação.

PROTEJA
Dados, modelos, aplicações e agentes.

Secure AI & Agents ajuda a construir uma abordagem de segurança para Inteligência Artificial com foco em pessoas, políticas, processos e tecnologia.

Mais visibilidade. Mais contexto. Mais governança.

Vamos descobrir onde estão as lacunas?

Angela Maria Rosso
angela.maria.rosso@accenture.com

Fontes: (1) Bloomberg, 2 mai. 2023, sobre o memorando interno da Samsung após funcionários enviarem código-fonte ao ChatGPT. (2) CNN, 16 mai. 2024, e South China Morning Post: transferências de HK$ 200 milhões, cerca de US$ 25,6 milhões, após videochamada com executivos falsos. (3) Moffatt v. Air Canada, 2024 BCCRT 149, British Columbia Civil Resolution Tribunal, fev. 2024. (4) PyTorch, blog oficial, “Compromised PyTorch-nightly dependency chain between December 25th and December 30th, 2022”, 31 dez. 2022.