GenAI na Cibersegurança: Desafios, Oportunidades e a Necessidade de Regulamentação no Contexto do AI Act

Stéfane Katarine Rodrigues da Silva[1], **Hugo Henrique Parreão Silva**[2]

[1]Instituto Federal de Educação, Ciência e Tecnologia do Maranhão – (IFMA)
CEP 65.906-335 – Imperatriz – MA – Brasil

[2]Departamento de Ensino Superior e Tecnologia – (IFMA)

stefanekatarine@acad.ifma.edu.br, hugoparreao@acad.ifma.edu.br

Abstract. *The article discusses the impact of Generative Artificial Intelligence (GenAI) on cybersecurity, highlighting its potential to improve threat detection and strengthen defenses. However, it emphasizes the associated risks, such as data breaches and vulnerabilities. The text addresses recent incidents involving GenAI tools that exposed privacy and security issues. It stresses the need for an adequate regulatory framework, such as the AI Act, to ensure the ethical and safe use of GenAI. Finally, it proposes a methodology to comparatively analyze the use of GenAI in cybersecurity in relation to the AI Act.*

Resumo. *O artigo discute o impacto da Inteligência Artificial Generativa (GenAI) na cibersegurança, destacando seu potencial para melhorar a detecção de ameaças e fortalecer defesas. Contudo, ressalta os riscos associados, como violações de dados e vulnerabilidades. O texto aborda incidentes recentes envolvendo ferramentas de GenAI, que expuseram problemas de privacidade e segurança. Enfatiza a necessidade de uma estrutura regulatória adequada, como o AI Act, para garantir o uso ético e seguro da GenAI. Por fim, propõe uma metodologia para analisar comparativamente o uso da GenAI na cibersegurança em relação ao AI Act.*

1. Introdução

A Inteligência Artificial (IA) tem revolucionado diversos aspectos da vida moderna, especialmente com o lançamento de ferramentas de IA Generativa (GenAI) como ChatGPT, Claude, Perplexity e Gemini [Nazareno, 2024]. No campo da cibersegurança, essas tecnologias demonstram significativo potencial, podendo detectar os ataques cibernéticos e reduzindo o tempo de resposta a incidentes, entretanto apesar da iminência das ameaças geradas pela GenAI, não há nenhuma evidência concreta de ataques cibernéticos com engenharia de GenAI até o momento. [IBM, 2024].

A GenAI, também traz desafios relacionados à ética, privacidade e regulamentação, especialmente em ambientes que lidam com dados sensíveis. Nesse contexto, surge o Ato de Inteligência Artificial (AI Act), uma legislação proposta pela Comissão Europeia para estabelecer um quadro legal abrangente que normatiza o desenvolvimento e uso da IA na União Europeia [Lantyer, 2023]. Essa regulamentação é essencial para definir padrões de segurança, garantir a proteção de dados e promover o

uso ético da IA. Desse modo, este artigo propõe uma reflexão sobre os impactos éticos e de segurança da GenAI, bem como suas contribuições para a cibersegurança, buscando equilibrar inovação e proteção.

Para isso, apresenta uma metodologia que avalia o alinhamento das práticas atuais de GenAI com os requisitos do AI Act, utilizando o modelo FAIR (Factor Analysis of Information Risk). A metodologia abrange uma análise detalhada das disposições legais, uma comparação com as práticas atuais e uma avaliação do impacto regulatório no desenvolvimento da GenAI.

2. IA Generativa: O Paradoxo entre Inovação e Vulnerabilidade de Dados Sensíveis

No âmbito técnico, a integração da GenAI na cibersegurança utiliza tecnologias fundamentais como Processamento de Linguagem Natural (NLP) e Grandes Modelos de Linguagem (LLMs). Contudo, sua eficácia depende da qualidade dos dados e do treinamento, podendo resultar em ”alucinações”ou respostas incorretas [Capodieci et al., 2024; Weise e Metz, 2023].

A popularização de ferramentas gratuitas de GenAI aumenta os riscos de exposição de dados sensíveis e vulnerabilidades a malwares, especialmente no âmbito empresarial [IPNET, 2024].

2.1. GenAI e Ética: Impactos e Riscos de Violações

Recentes incidentes de segurança evidenciaram vulnerabilidades críticas no uso de Inteligência Artificial Generativa (GenAI). Destacam-se dois casos principais: uma violação de dados no ChatGPT que expôs conversas privadas de usuários [SecurityIntelligence, ”ChatGPT confirma violação de dados levantando preocupações de segurança”, maio de 2023], e um incidente na Samsung onde funcionários inadvertidamente comprometeram dados confidenciais da empresa ao utilizar ferramentas de GenAI para desenvolvimento de código [TechRadar, ”Os funcionários da Samsung cometeram um grande erro ao usar o ChatGPT”, maio de 2023].

Estes eventos demonstram riscos significativos em duas dimensões: privacidade individual e segurança corporativa, especialmente relacionada à propriedade intelectual. Como resposta, surge a necessidade de marcos regulatórios robustos, exemplificados pelo AI Act da União Europeia, que visa estabelecer padrões para uso ético e seguro de IA. A governança em GenAI se estrutura através de diversos frameworks que abordam diferentes aspectos: análise quantitativa de riscos, “segurança por design” e conformidade regulatória. Esta multiplicidade permite que organizações selecionem estruturas apropriadas às suas necessidades específicas, possibilitando uma implementação mais segura e responsável da tecnologia.

Estes eventos demonstram riscos significativos em duas dimensões: privacidade individual e segurança corporativa, especialmente relacionada à propriedade intelectual. Como resposta, surge a necessidade de marcos regulatórios robustos, exemplificados pelo AI Act da União Europeia, que visa estabelecer padrões para uso ético e seguro de IA. A governança em GenAI se estrutura através de diversos frameworks que abordam diferentes aspectos: análise quantitativa de riscos, “segurança por design” e conformidade

regulatória. Esta multiplicidade permite que organizações selecionem estruturas apropriadas às suas necessidades específicas, possibilitando uma implementação mais segura e responsável da tecnologia.

3. Metodologia

A estrutura metodológica proposta tem o objetivo de avaliar a implementação da GenAI na abordagem de segurança cibernética, alinhando-se aos requisitos de conformidade da AI Act da União Europeia. A metodologia se fundamenta no modelo FAIR , uma prática amplamente utilizada na cibersegurança para a avaliação e gestão de riscos. Essa abordagem estruturada permite não apenas a quantificação dos riscos, mas também a adaptação específica ao contexto da GenAI, considerando incidentes e violações de dados relevantes ao AI Act. Assim, a metodologia assegura que as organizações estejam em conformidade com as exigências regulatórias de transparência e responsabilidade nos sistemas de inteligência artificial [FAIR, 2024].

3.1. Estrutura da Framework

A estrutura opera por meio de uma abordagem tridimensional sistemática que examina ameaças, ativos e possíveis eventos de perda. A análise abrange a avaliação das vulnerabilidades do sistema (por exemplo, injeção imediata, envenenamento de modelos), a catalogação de ativos protegidos (dados confidenciais, integridade do sistema) e a avaliação de possíveis resultados negativos (exfiltração de dados, erosão da confiança). Essa abordagem em várias camadas garante uma cobertura abrangente dos aspectos técnicos e operacionais da implementação do GenAI.

3.2. Coleta e Análise de Dados

A metodologia emprega a triangulação por meio de três fontes de dados primários:

1. Revisão sistemática da literatura de publicações acadêmicas e relatórios técnicos;
2. Análise de estudos de caso de implementações bem-sucedidas e incidentes de segurança;
3. Consultas a especialistas por meio de entrevistas semiestruturadas com profissionais de segurança cibernética, desenvolvedores de IA e especialistas jurídicos.

Essa abordagem abrangente garante a coleta e a validação de dados robustos em domínios teóricos e práticos.

3.3. Avaliação de Conformidade

A avaliação de conformidade integra quatro dimensões principais:

1. Avaliação da classificação de risco com base na estrutura do AI Act.
2. Avaliação de transparência e explicabilidade dos processos de tomada de decisão.
3. Avaliação da conformidade da proteção de dados e da privacidade.
4. Avaliação da implementação da supervisão humana.

Cada dimensão passa por uma avaliação rigorosa usando critérios predeterminados alinhados com os requisitos regulatórios, mantendo a aplicabilidade prática.

3.4. Validação e Recomendações

A metodologia é concluída com um processo de validação que utiliza análises de painéis de especialistas e feedback das partes interessadas. Esse processo leva ao desenvolvimento de recomendações concretas que abrangem diretrizes de implementação técnica, estruturas de governança e procedimentos de monitoramento de conformidade.

4. Considerações Finais

Conforme exposto, a metodologia inspirada no AI Act constitui uma estrutura adaptável que harmoniza implementação ética e segura de GenAI na América Latina, com ênfase em cibersegurança. Sua arquitetura modular permite customizações para diferentes contextos setoriais e regionais, mantendo alinhamento com diretrizes globais e respondendo às crescentes demandas regulatórias de proteção de dados e privacidade.

Regulamentações emergentes, como as brasileiras, reforçam princípios de ”segurança por design”e transparência, fundamentando um ecossistema de GenAI que equilibra proteção e inovação, capacitando a região para navegar as complexidades do ambiente digital contemporâneo enquanto preserva particularidades locais e padrões internacionais de excelência.

Referências

ChatGPT confirma violação de dados levantando preocupações de segurança. Security Intelligence, maio de 2023. Disponível em: https://securityintelligence.com/articles/chatgpt-confirms-data-breach/. Acesso em: 31 out. 2024.

FAIR Institute. FAIR Risk Management. Disponível em: https://www.fairinstitute.org/fair-risk-management. Acesso em: 31 out. 2024.

IBM. X-Force Threat Intelligence Index 2024. São Paulo: IBM Brasil Ltda; Armonk: IBM Corporation, fevereiro de 2024. Produzido nos Estados Unidos da América.

IPNET. Como manter sua organização segura ao usar inteligência generativa. 2024.

Lantyer, Victor Habib. Entendendo o EU AI Act: uma nova era na regulamentação da IA na Europa. Migalhas, 2023. Disponível em: https://www.migalhas.com.br/arquivos/2023/12/B6D06B89351862Artigo-RegulamentoEuropeudeIA-.pdf. Acesso em: 28 out. 2024.

Nazareno, Claudio. Regulação da inteligência artificial: experiências internacionais e desafios para o Brasil. Brasília: Câmara dos Deputados, 2024. (Estudo da Consultoria Legislativa).

N. Capodieci, C. Sanchez-Adames, J. Harris e U. Tatar, ”O impacto da IA generativa e LLMs na profissão de segurança cibernética”, Simpósio de Design de Engenharia de Sistemas e Informação de 2024 (SIEDS), Charlottesville, VA, EUA, 2024, pp. 448-453, doi: 10.1109/SIEDS61124.2024.10534674.

Os funcionários da Samsung cometeram um grande erro ao usar o ChatGPT, maio de 2023, [online] Disponível em: https://www.techradar.com/news/samsung-workers-leaked-company-secrets-by-using-chatgpt.