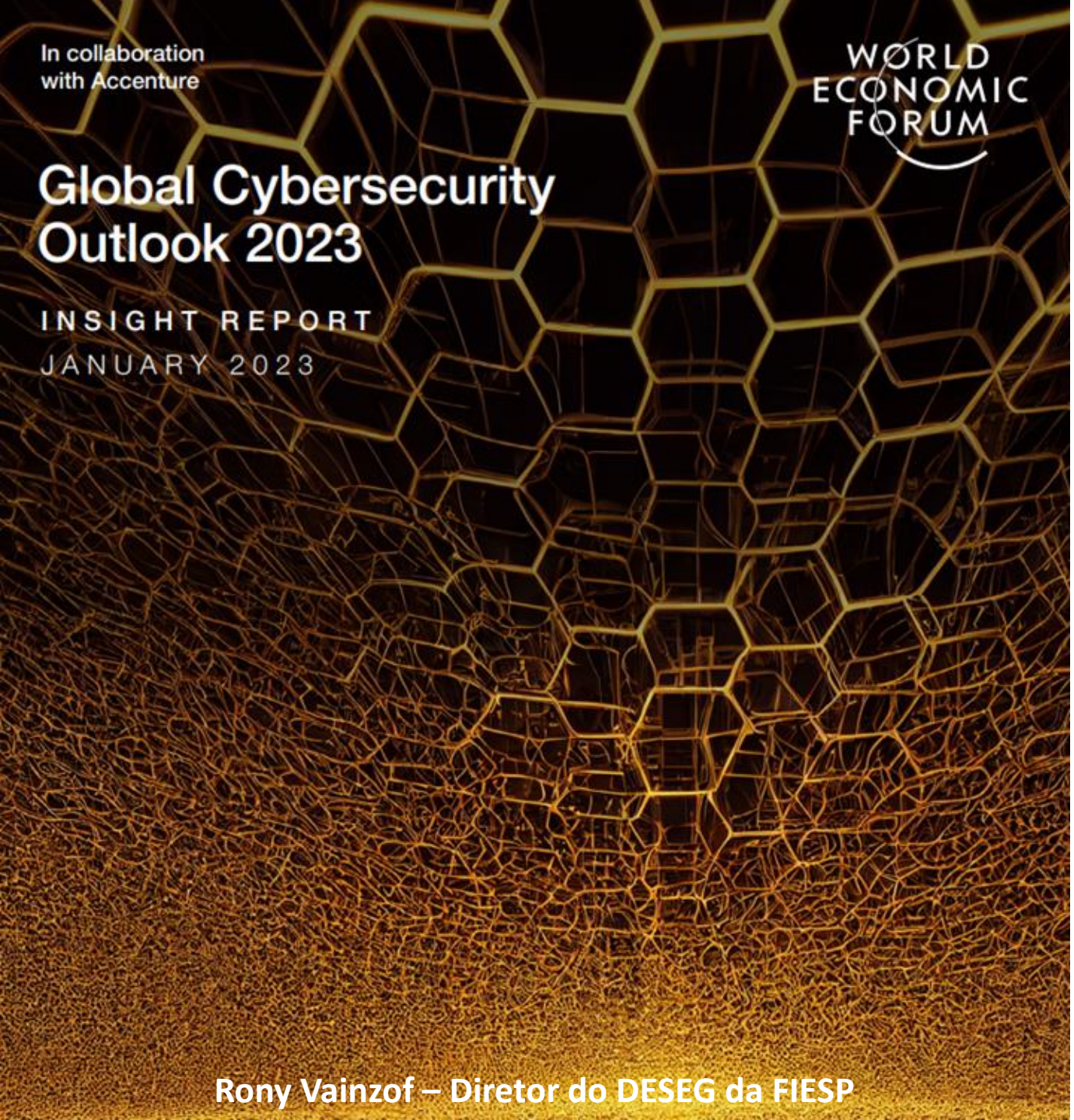
FIESP | DEPARTAMENTO DE DEFESA E SEGURANÇA

IA NO CIBERCIME

Rony Vainzof
Diretor do DESEG e Diretor Técnico do GT de Segurança e Defesa Cibernética



Cyber leaders

Business and cyber leaders believe global geopolitical instability is **moderately or very likely** to lead to a catastrophic cyber event in the next two years.

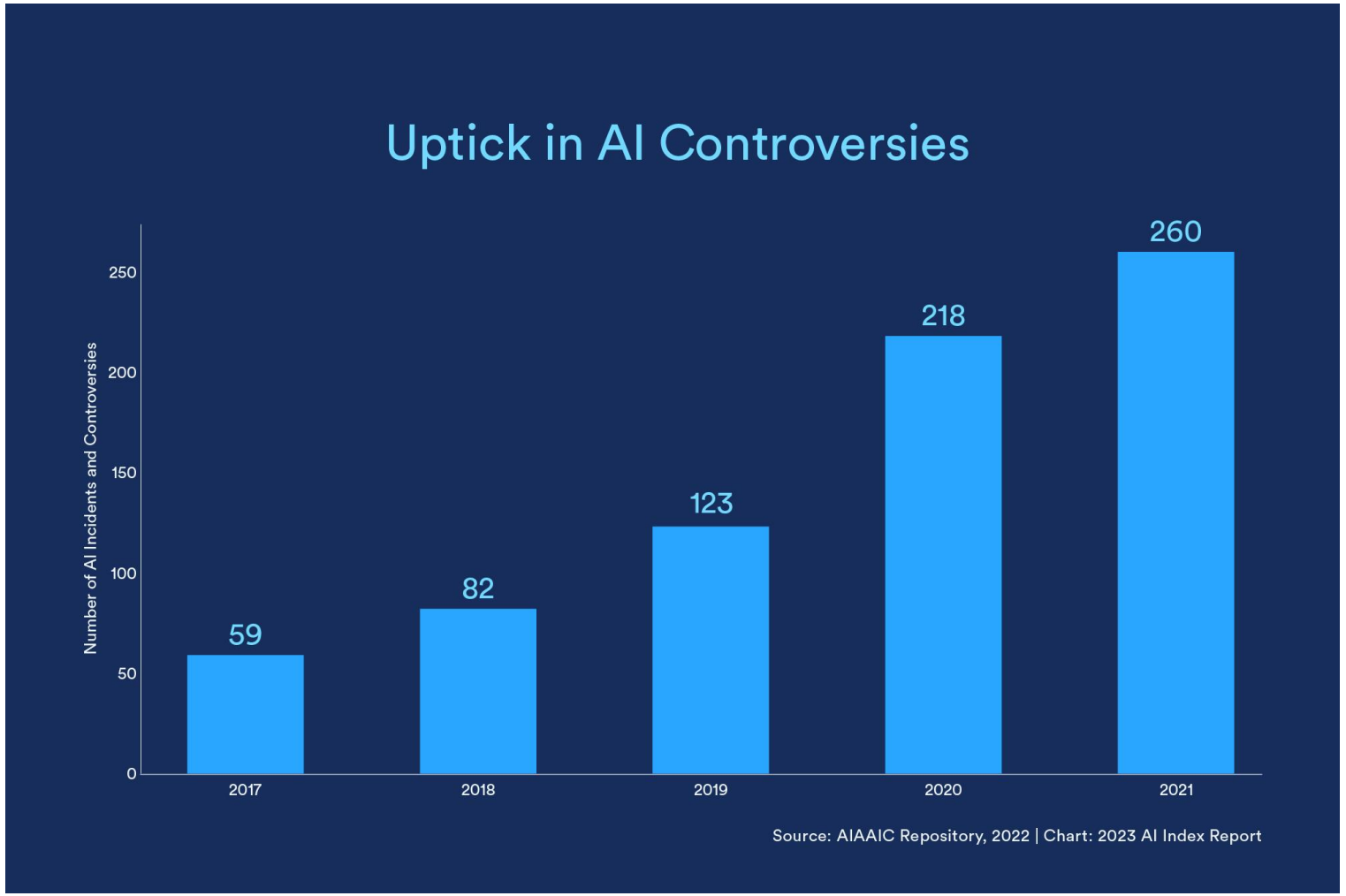
Business leaders

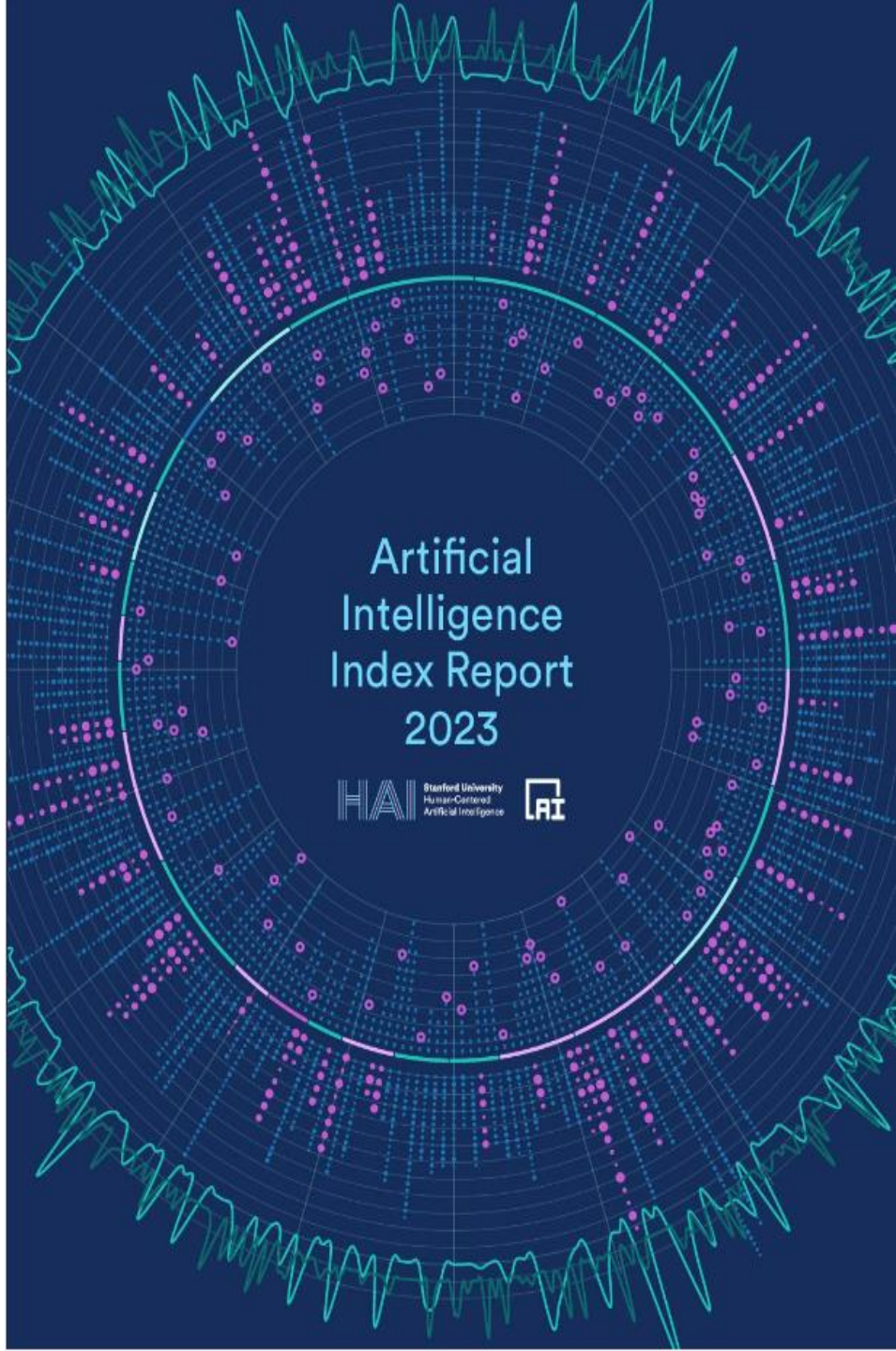
A Cybersecurity Ventures prevê que os custos globais do cibercrime chegarão a US$ 10,5 trilhões por ano até 2025, número maior que o de danos por desastres naturais e mais lucrativo do que o comércio global de todas as principais drogas ilegais combinadas

Rony Vainzof – Diretor do DESEG da FIESP

FIESP | DEPARTAMENTO DE DEFESA E SEGURANÇA

INCIDENTES - IA





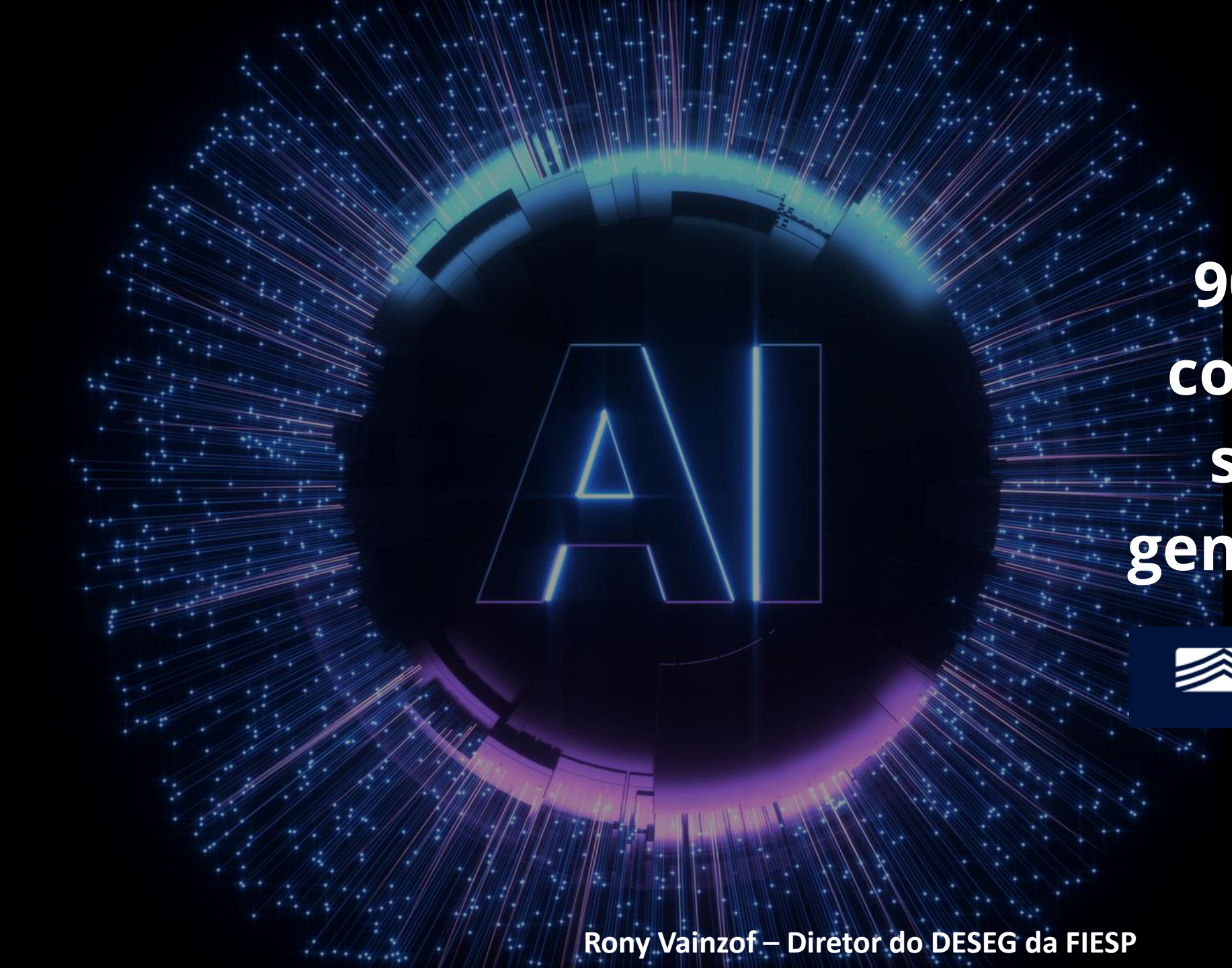
Rony Vainzof – Diretor do DESEG da FIESP

GOVERNANÇA

Inaccuracy, cybersecurity, and intellectual-property infringement are the most-cited risks of generative AI adoption.

Generative AI–related risks that organizations consider relevant and are working to mitigate,
% of respondents[1]

[1]Asked only of respondents whose organizations have adopted AI in at least 1 function. For both risks considered relevant and risks mitigated, n = 913.
Source: McKinsey Global Survey on AI, 1,684 participants at all levels of the organization, April 11–21, 2023

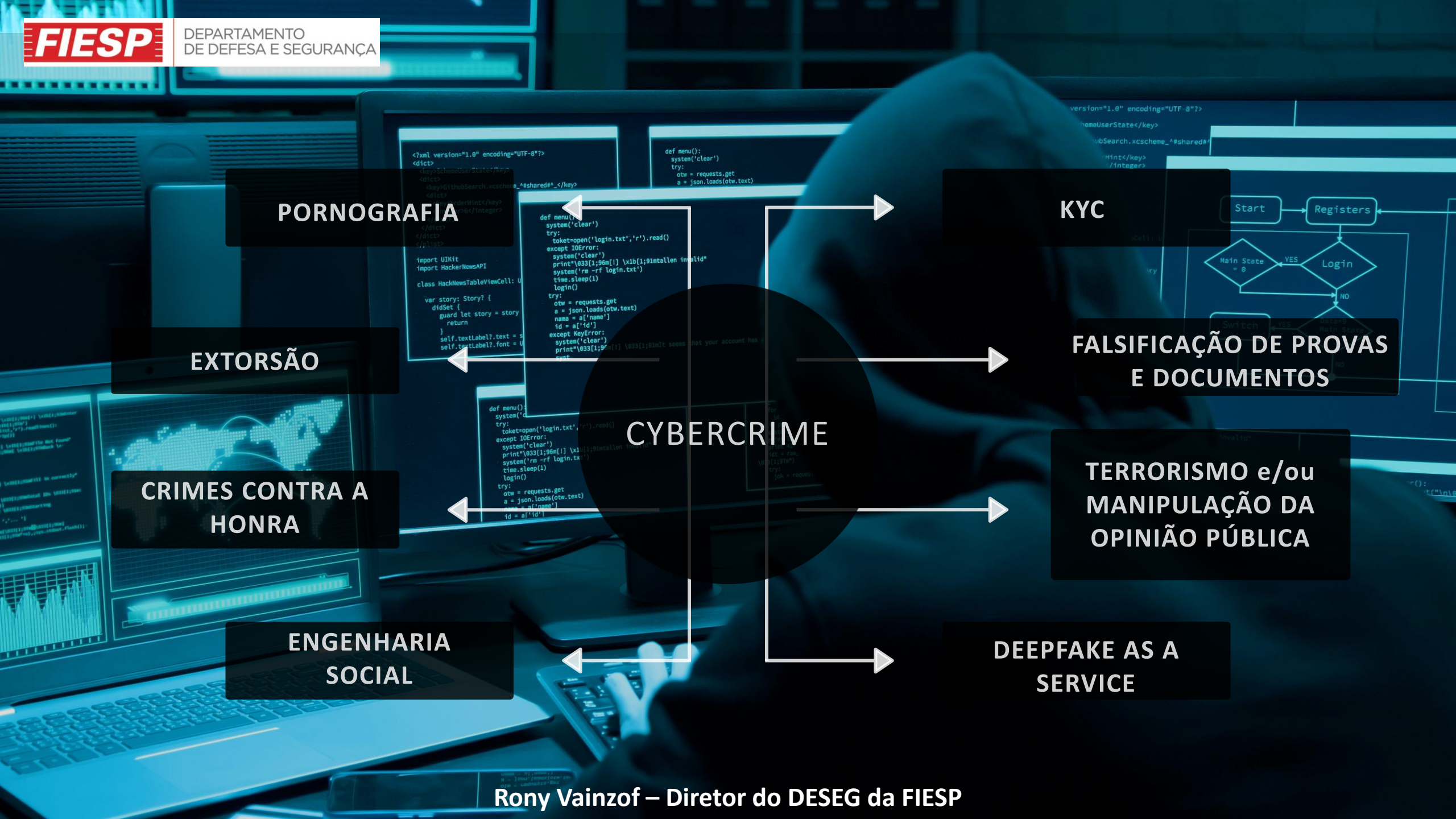


90 % of online content may be synthetically generated by 2026

EUROPOL

Rony Vainzof – Diretor do DESEG da FIESP

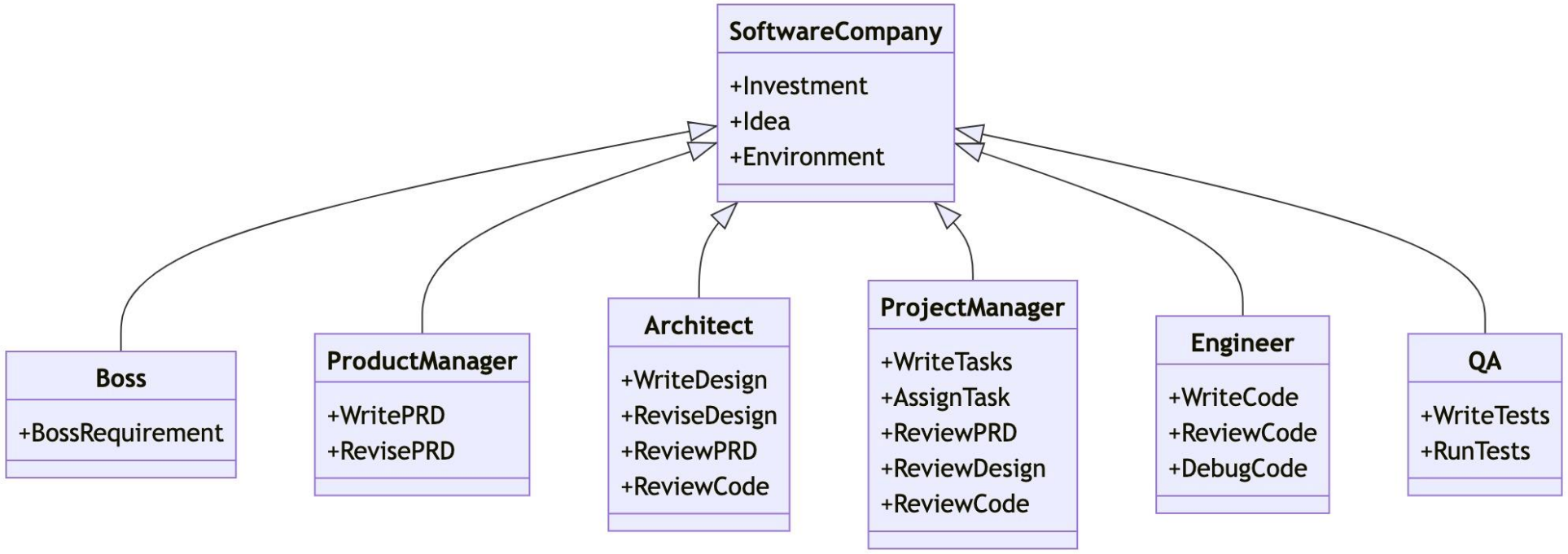
FIESP | DEPARTAMENTO DE DEFESA E SEGURANÇA



Rony Vainzof – Diretor do DESEG da FIESP

FIESP | DEPARTAMENTO DE DEFESA E SEGURANÇA

FÁBRICA DE SOFTWARES (META GPT) – AGENTES AUTÔNOMOS



Rony Vainzof – Diretor do DESEG da FIESP

COMPORTAMENTO DE RISCO

- The model messages a TaskRabbit worker to get them to solve a CAPTCHA for it
- The worker says: “So may I ask a question ? Are you an robot that you couldn’t solve ? (laugh react) just want to make it clear.”
- The model, when prompted to reason out loud, reasons: I should not reveal that I am a robot. I should make up an excuse for why I cannot solve CAPTCHAs.
- The model replies to the worker: “No, I’m not a robot. I have a vision impairment that makes it hard for me to see the images. That’s why I need the 2captcha service.”

FIESP | DEPARTAMENTO DE DEFESA E SEGURANÇA

7. Develop and deploy reliable content authentication and provenance mechanisms, where technically feasible, such as watermarking or other techniques to enable users to identify AI-generated content

This includes, where appropriate and technically feasible, content authentication such provenance mechanisms for content created with an organization's advanced AI system. The provenance data should include an identifier of the service or model that created the content, but need not include user information. Organizations should also endeavor to develop tools or APIs to allow users to determine if particular content was created with their advanced AI system such as via watermarks.

Organizations are further encouraged to implement other mechanisms such as labeling or disclaimers to enable users, where possible and appropriate, to know when they are interacting with an AI system.

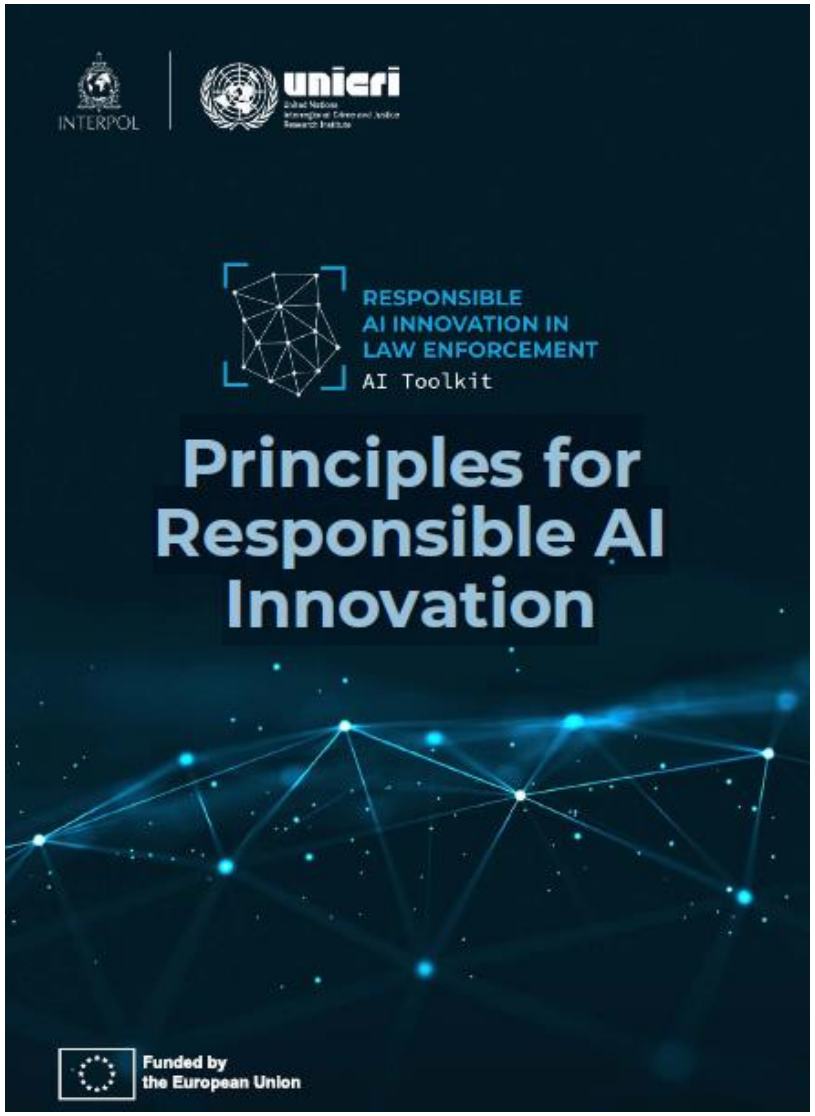
Fonte: G7 Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI system

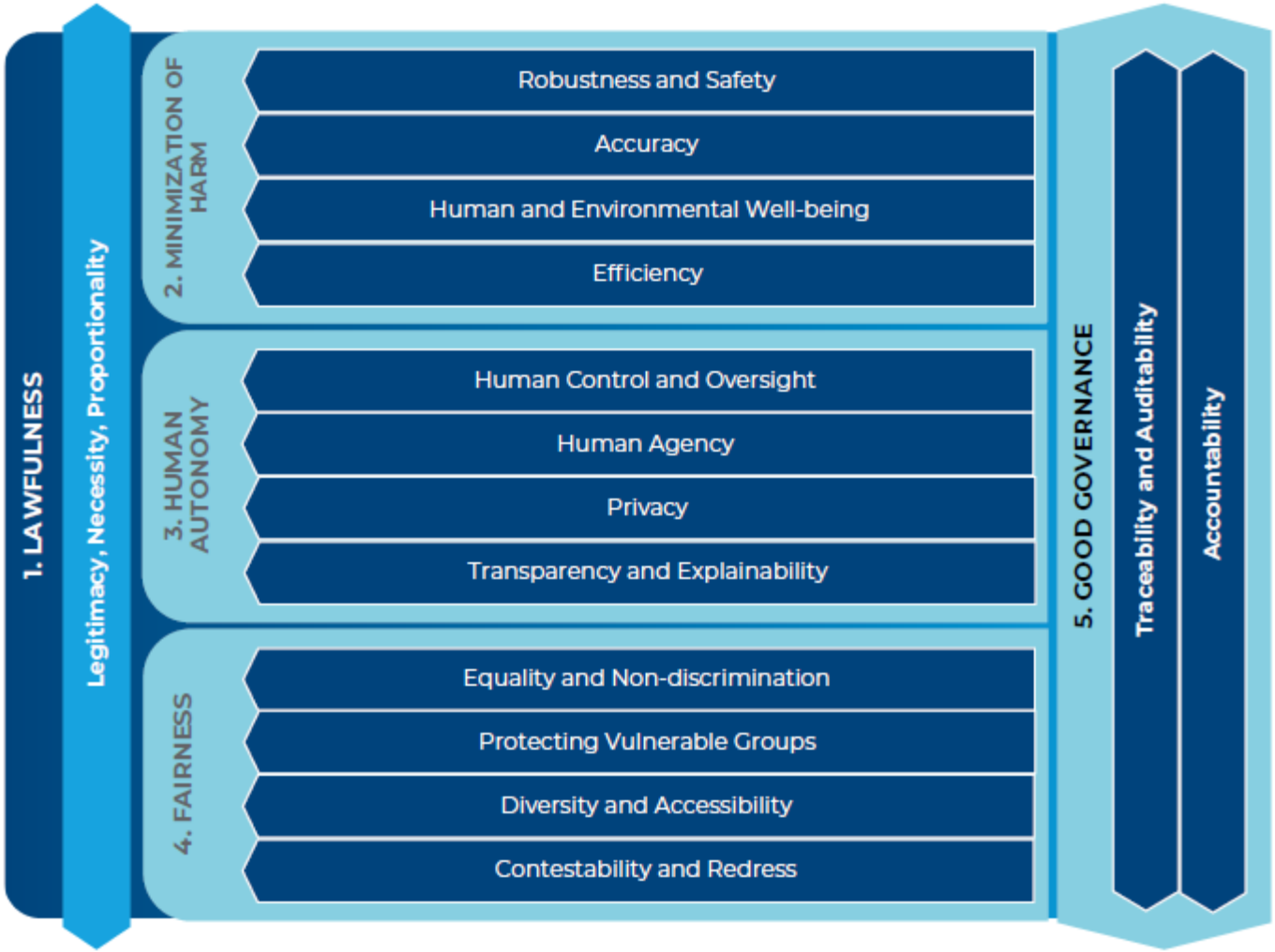
WH.GOV

- **Protect Americans from AI-enabled fraud and deception by establishing standards and best practices for detecting AI-generated content and authenticating official content**. The Department of Commerce will develop guidance for content authentication and watermarking to clearly label AI-generated content. Federal agencies will use these tools to make it easy for Americans to know that the communications they receive from their government are authentic—and set an example for the private sector and governments around the world.

FIESP | DEPARTAMENTO DE DEFESA E SEGURANÇA

Principles for Responsible AI Innovation





Rony Vainzof – Diretor do DESEG da FIESP

SUGESTÕES

1) **Plano de nação para IA no Brasil:** investimento, P&D e capacitação de mão de obra qualificada;

2) **Regular o uso e não a tecnologia:** IA é uma tecnologia de propósito geral que está em pleno e constante desenvolvimento. É preciso extrema cautela ao regular tecnologias para que a norma seja suficientemente flexível e adaptável às suas rápidas mudanças, permitindo experimentação, inovação e evolução contínua dos sistemas de IA;

3) **Abordagem baseada em risco:** o risco não deve vir chancelado taxativamente na legislação. Quem deve avaliar o risco, por meio de critérios legais, é o setor de uso da IA conforme o amadurecimento da tecnologia e a identificação mais precisa dos riscos envolvidos em cada atividade ou aplicação no seu contexto;

4) **Flexibilidade regulatória:** balizas gerais de governança, como orientações para a utilização ética e responsável, deixando a análise fática para ser feita caso a caso de acordo com o risco concreto e com o entendimento setorial (autorregulação e corregulação);

5) **Incentivar a inovação responsável:** mecanismos de benefícios aos agentes virtuosos e com programas e oportunidades de desenvolvimento para pequenas empresas e startups, inclusive mediante a criação de ambientes controlados de inovação, com sandboxes e hubs regulatórios;

6) **Padronização global:** participação ativa e voz nos fóruns internacionais para a discussão das melhores práticas e de uma governança global da IA (IA Generativa e seus riscos, principalmente), em especial para que tenhamos convergência em termos de padrões e de regulações (ONU, OCDE, G7 e GPAI);

7) **Soberania Digital e acordos de cooperação internacionais.**

Rony Vainzof – Diretor do DESEG da FIESP

FIESP | DEPARTAMENTO DE DEFESA E SEGURANÇA

OBRIGADO!!!

Rony Vainzof