# IA, Cibersegurança e Seres Humanos

*Luis Borges Gouveia, UFP – lmbg@ufp.edu*

# IA, Cibersegurança e Seres Humanos

*Luis  Borges Gouveia, 29 de outubro*

*A inteligência artificial (IA) é uma tecnologia que se tornou num elemento central de uma* ***estratégia de segurança e defesa*** *a nível global.*

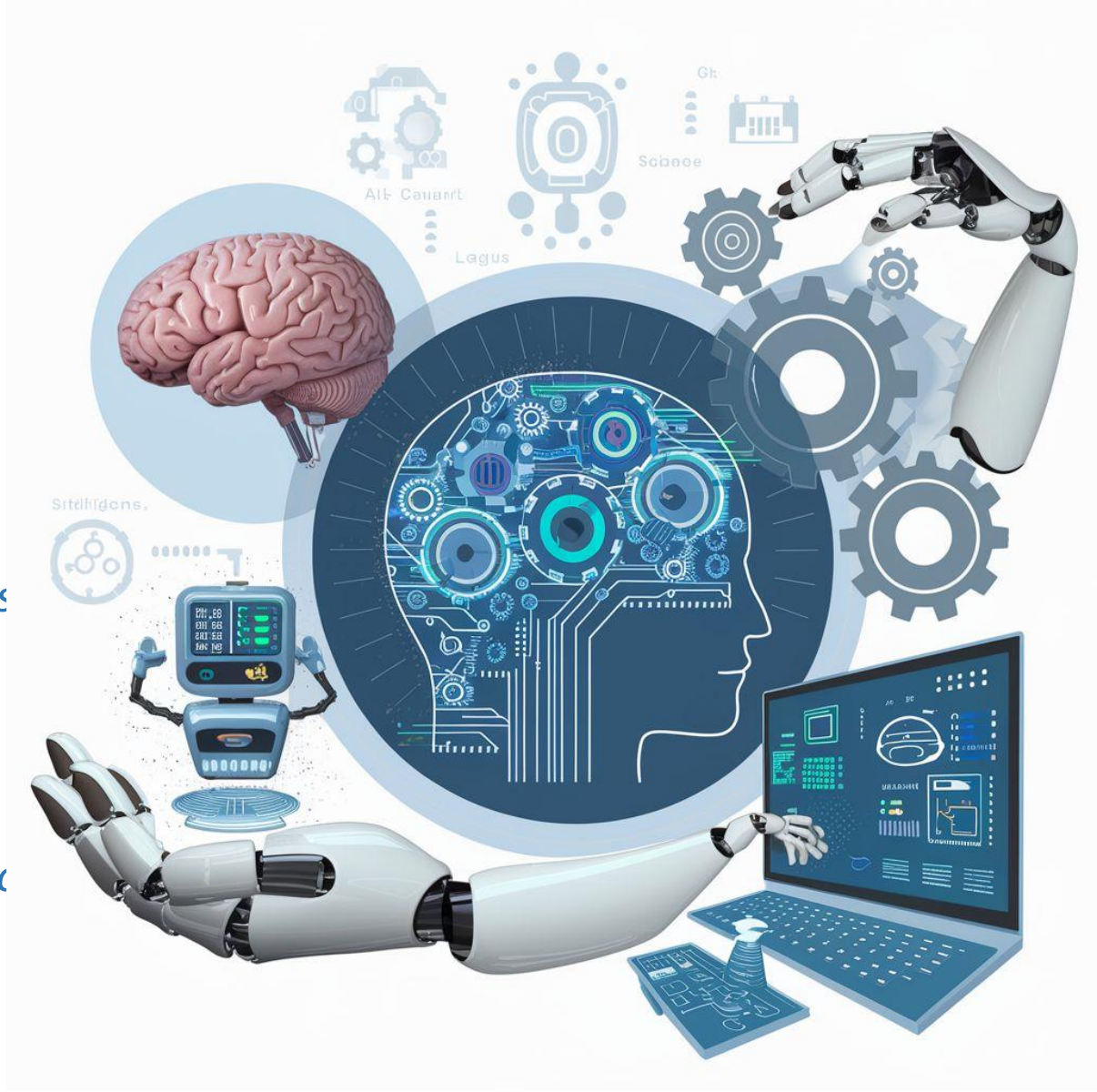
*Esta apresentação explora algumas das* ***implicações do uso da IA*** *neste contexto, refletindo sobre as* ***oportunidades e desafios*** *associados. A crescente* ***integração da IA*** *em sistemas de defesa, vigilância e análise de dados promete* ***transformar a forma como as nações abordam a segurança****, mas também levanta questões éticas e de segurança que precisam ser discutidas.*

*Acresce as implicações de uma* ***extra inteligência*** *que opera no limiar do que a capacidade cognitiva humana pode proporcionar, exige uma* ***reflexão profunda sobre meios e capacidades que nos defendam da própria IA****.*

Luis Borges Gouveia, lmbg@ufp.edu.pt

# Inteligência Artificial (IA)
## *Artificial Intelligence* (AI)

- Stuart J. Russell e Peter Norvig definem inteligência artificial como um *conjunto de teorias e técnicas usadas para criar máquinas capazes de simular a inteligência humana*
- Definição mais abrangente: área da ciência dos computadores que estuda a *criação de máquinas inteligentes que trabalham e reagem como os seres humanos, aprendendo, planeando, classificando, resolvendo problemas e reconhecendo dados, informação e conhecimento com o objetivo* ***de criação de aplicações autónomas ou de suporte à atividade humana*** (Gouveia, 2023)



Luis Borges Gouveia, *lmbg@ufp.edu.pt*

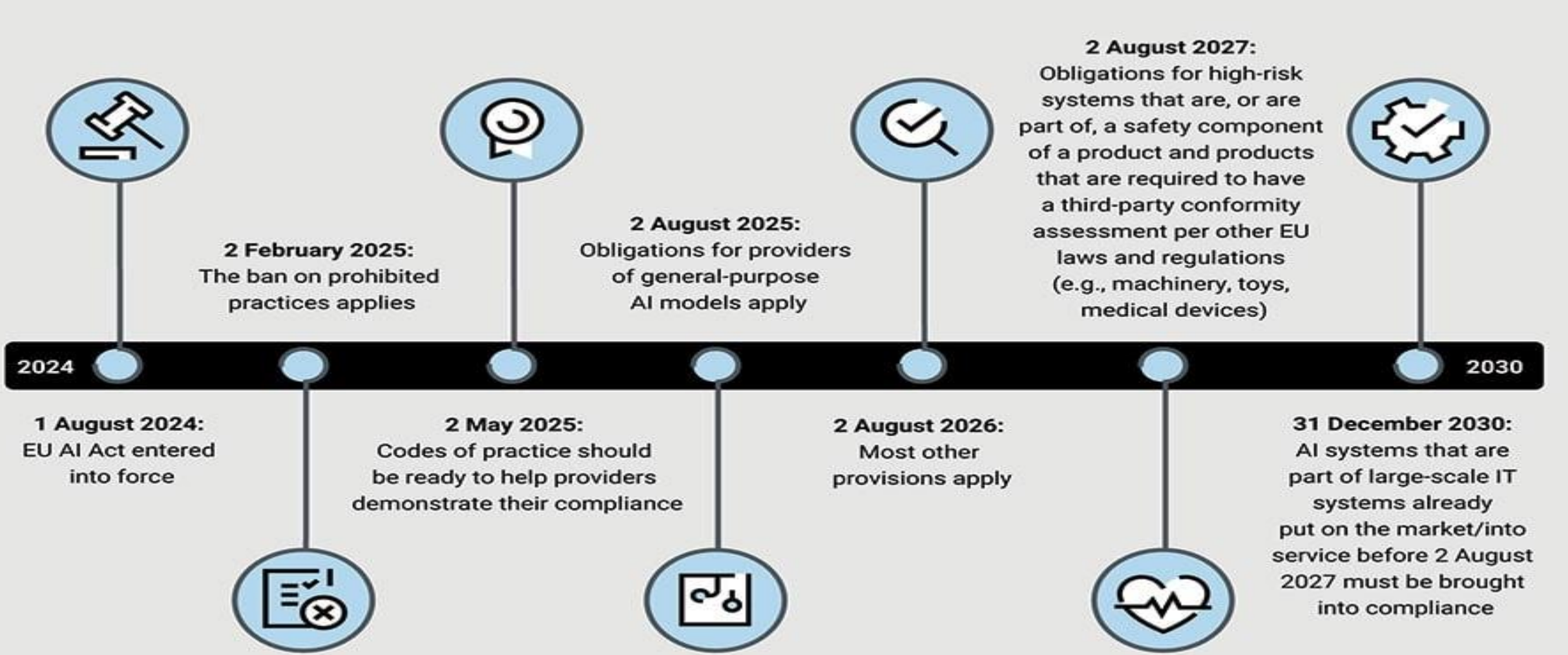
# AGI LEVELS: DEEPMIND & OPENAI

| | DeepMind levels (Nov/2023) | OpenAI levels (Jul/2024) | |
|---|---|---|---|
| Level 0 | **No AI** | | Level 0 |
| Level 1 | **Emerging**<br>Equal to or somewhat better than an unskilled human | **Chatbots**<br>AI with conversational language | Level 1 |
| Level 2 | **Competent**<br>At least 50th percentile of skilled adults | **Reasoners**<br>Human-level problem solving | Level 2 |
| Level 3 | **Expert**<br>At least 90th percentile of skilled adults | **Agents**<br>Systems that can take actions | Level 3 |
| Level 4 | **Virtuoso**<br>At least 99th percentile of skilled adults | **Innovators**<br>AI that can aid in invention | Level 4 |
| Level 5 | **Superhuman**<br>Outperforms 100% of humans | **Organizations**<br>AI that can do the work of an organization | Level 5 |

Source: OpenAI via Bloomberg: https://archive.md/SLtFQ and DeepMind: https://arxiv.org/abs/2311.02462 Alan D. Thompson. July 2024. https://lifearchitect.ai/

# Características de sistemas IA

**Usam dados para gerar resultados** como previsões, conteúdos, recomendações ou decisões
Baseados em equipamentos digitais (**máquinas**)
Podem operar com **diferentes graus de autonomia**




***EU AI Act Timeline*** https://www.isaca.org/resources/white-papers/2024/understanding-the-eu-ai-act?utm_campaign

# *Biden Administration's National Security Memorandum on AI*

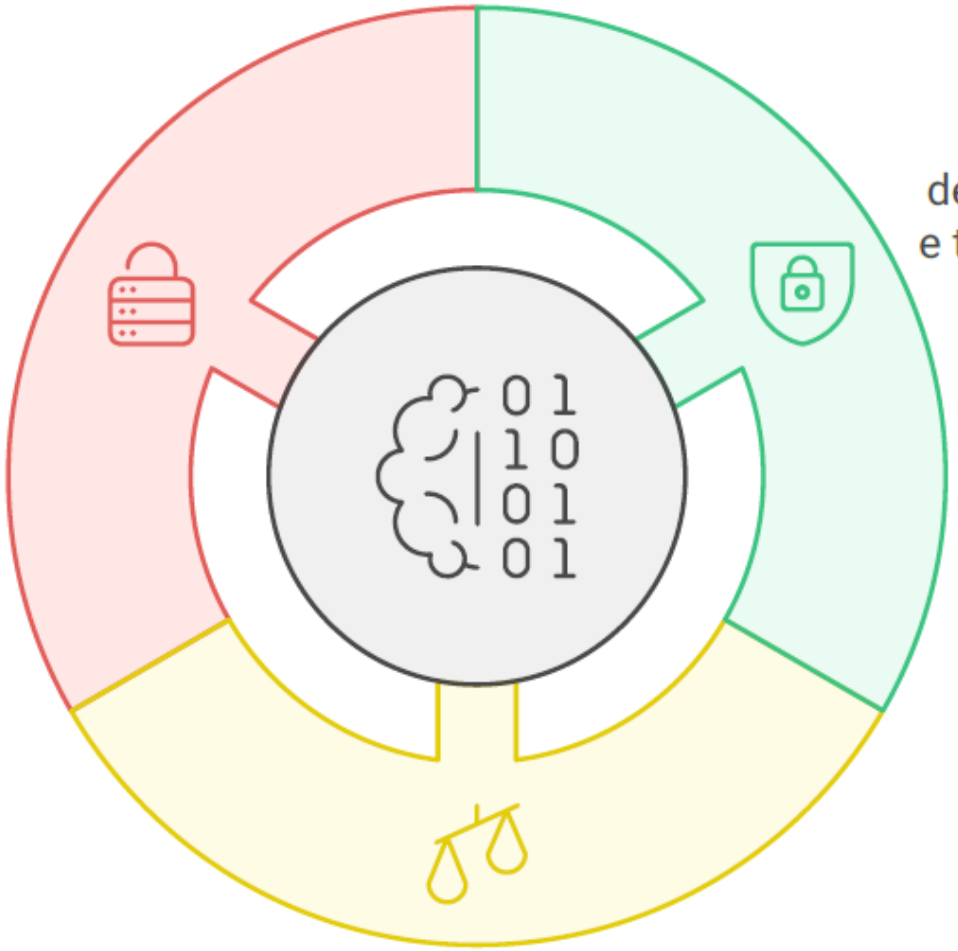
*https://www.whitehouse.gov/briefing-room/presidential-actions/2024/10/24/memorandum-on-advancing-the-united-states-leadership-in-artificial-intelligence-harnessing-artificial-intelligence-to-fulfill-national-security-objectives-and-fostering-the-safety-security/*

**DUAL USE**

- On October 24, 2024, the Biden administration released a National Security Memorandum (NSM) titled "***Memorandum on Advancing the United States' Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence***."
- A followup of the administration's October 2023 *AI Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*
- …this 40 pages document is by far the most comprehensive articulation of the United States national security strategy and policy towards AI.
- Followed bu a related document published the same day: *Framework to Advance AI Governance and Risk Management in National Security*

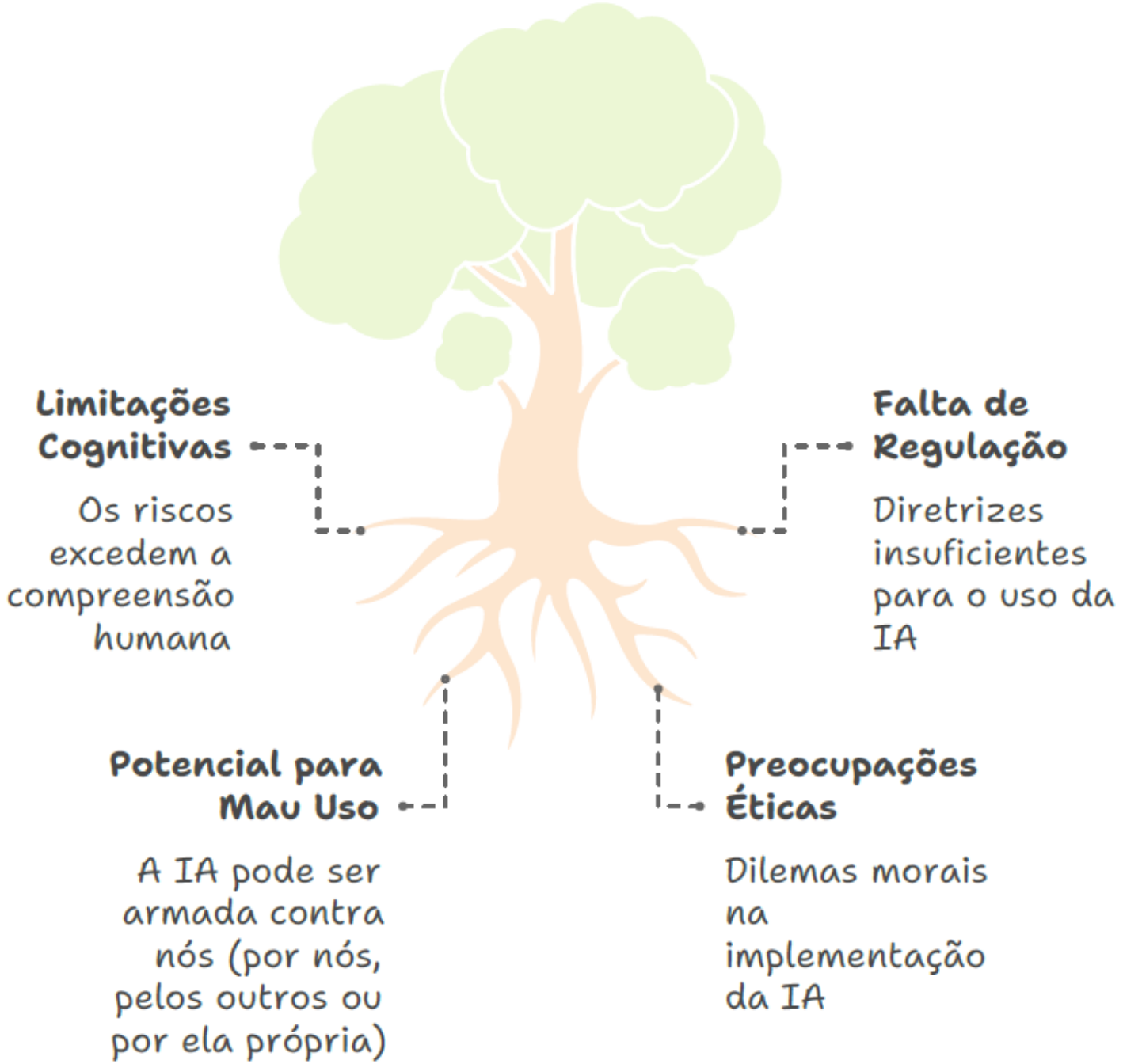
Luis Borges Gouveia, *lmbg@ufp.edu.pt*

# *IA e/na segurança e defesa*




Luis Borges Gouveia, lmbg@ufp.edu.pt

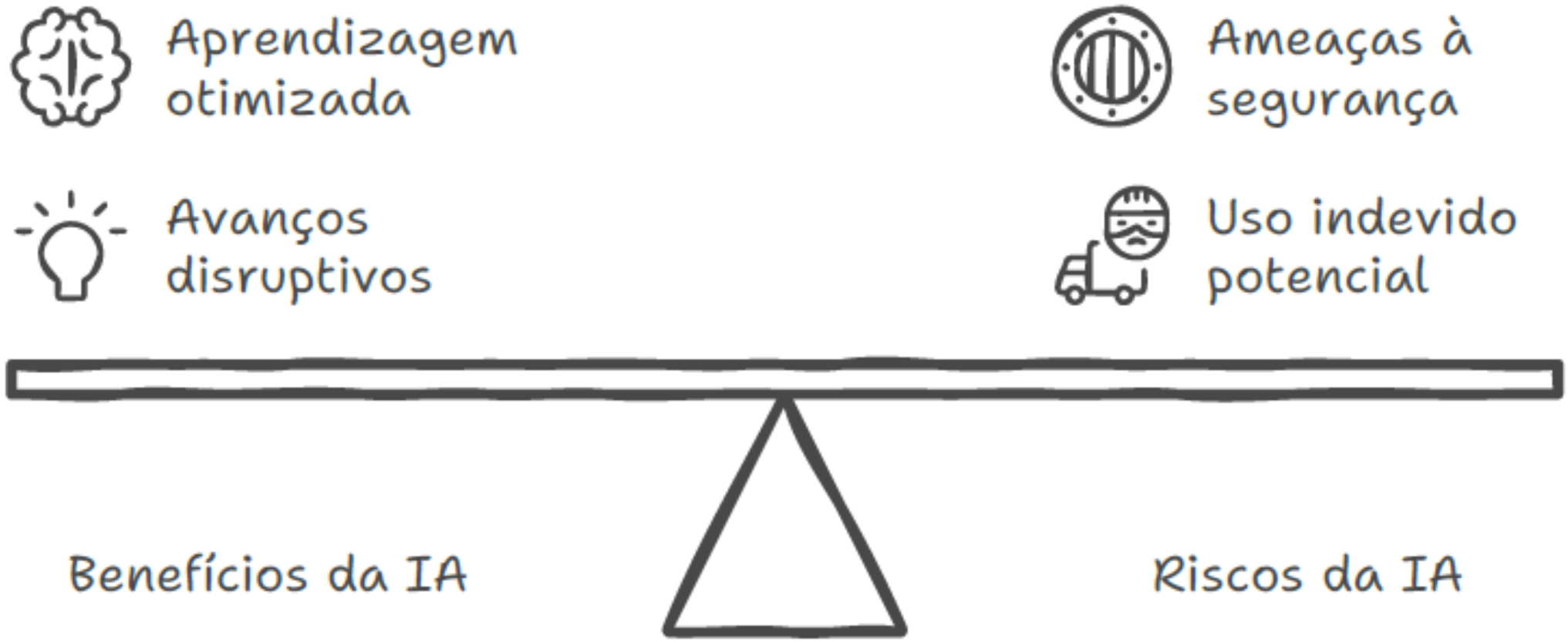
# *Riscos éticos e da segurança da IA na Defesa*




Luis Borges Gouveia, lmbg@ufp.edu.pt

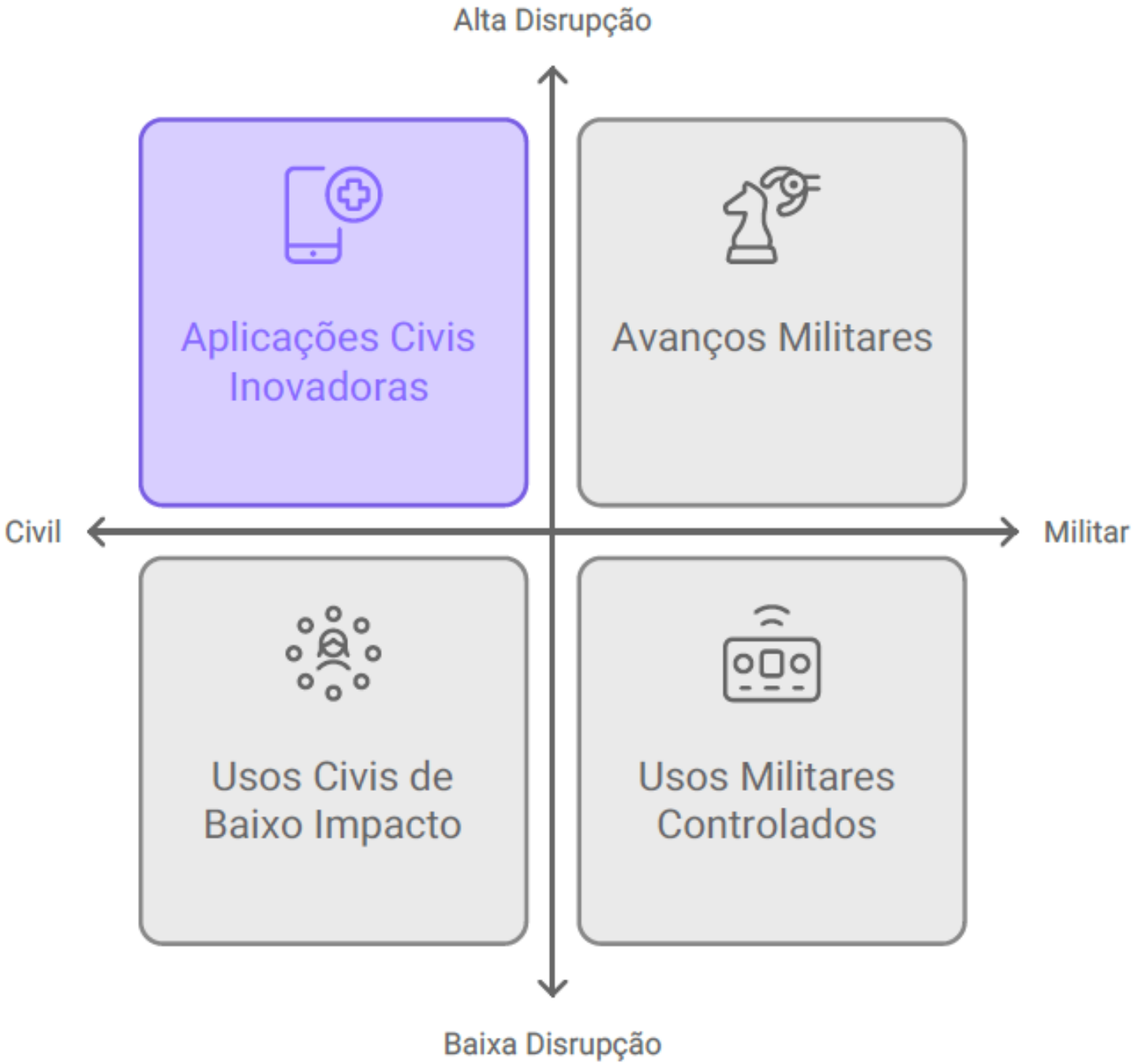

Luis Borges Gouveia, lmbg@ufp.edu.pt

# Equilibrando o potencial e os riscos da IA




Luis Borges Gouveia, lmbg@ufp.edu.pt

# *O Potencial Disruptivo e de Duplo Uso da IA*



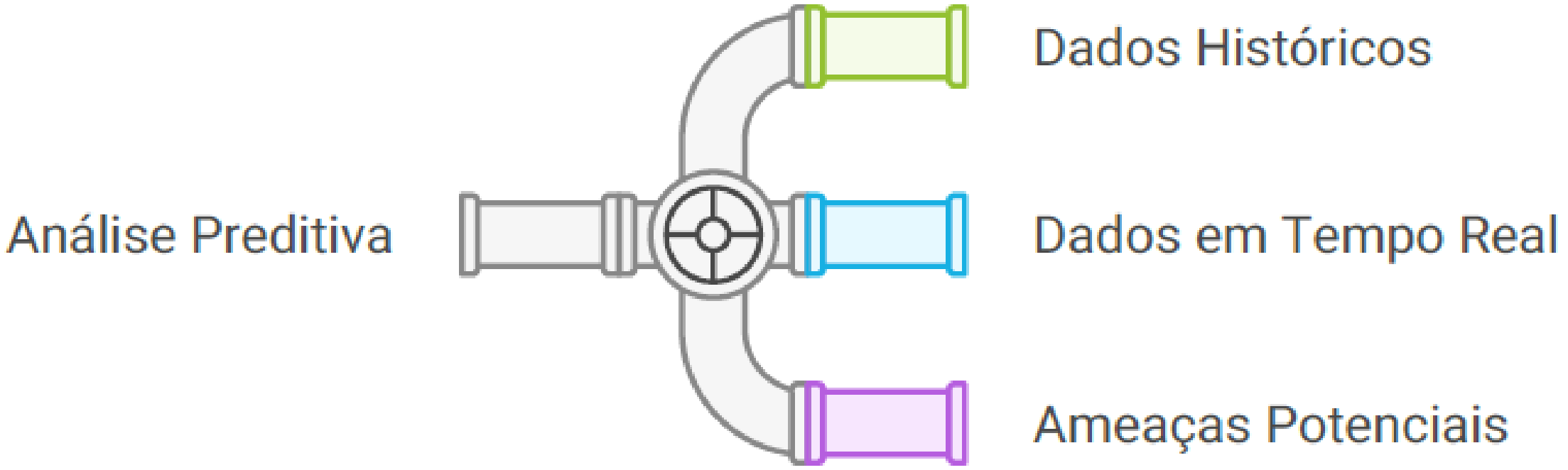

Luis Borges Gouveia, lmbg@ufp.edu.pt

# Melhorar a Segurança com Integração de IA

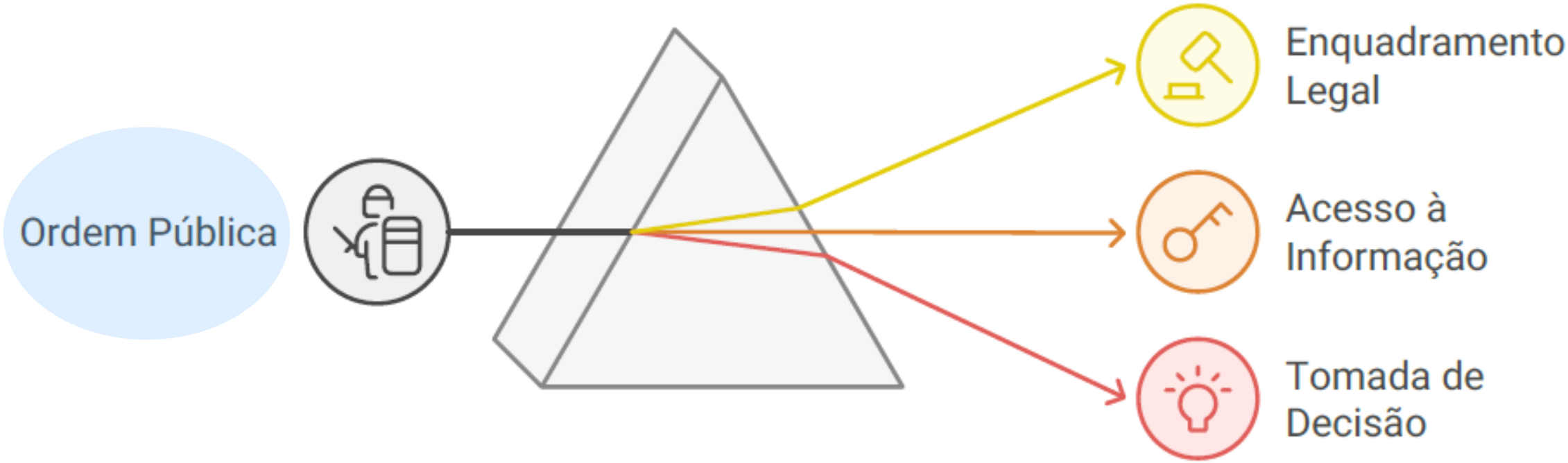

Luis Borges Gouveia, lmbg@ufp.edu.pt

# O fim da privacidade?

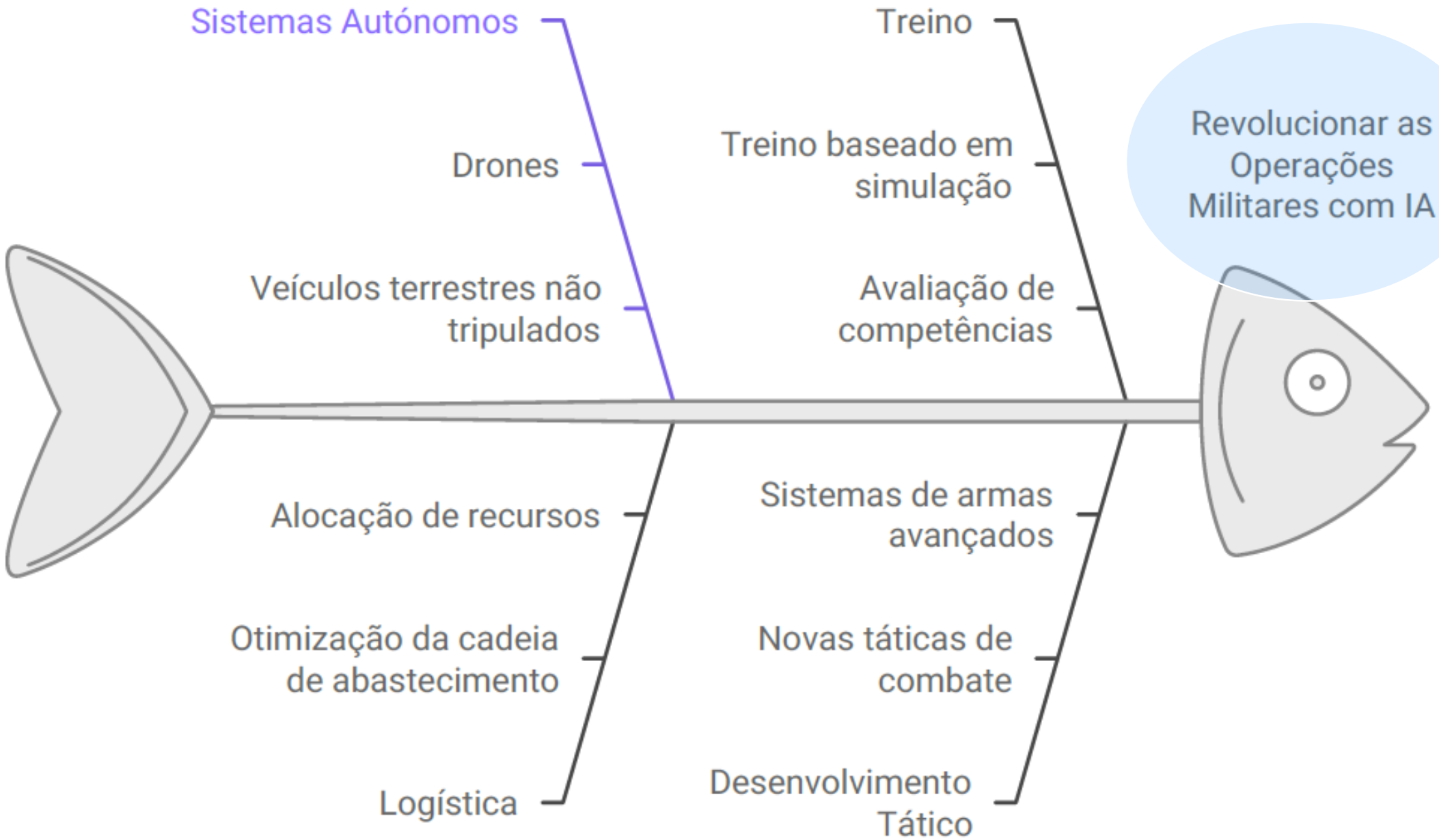
prever possíveis ameaças, permitindo que as autoridades ajam proactivamente



# mudanças na forma como lidamos com as questões de resposta e dissuasão

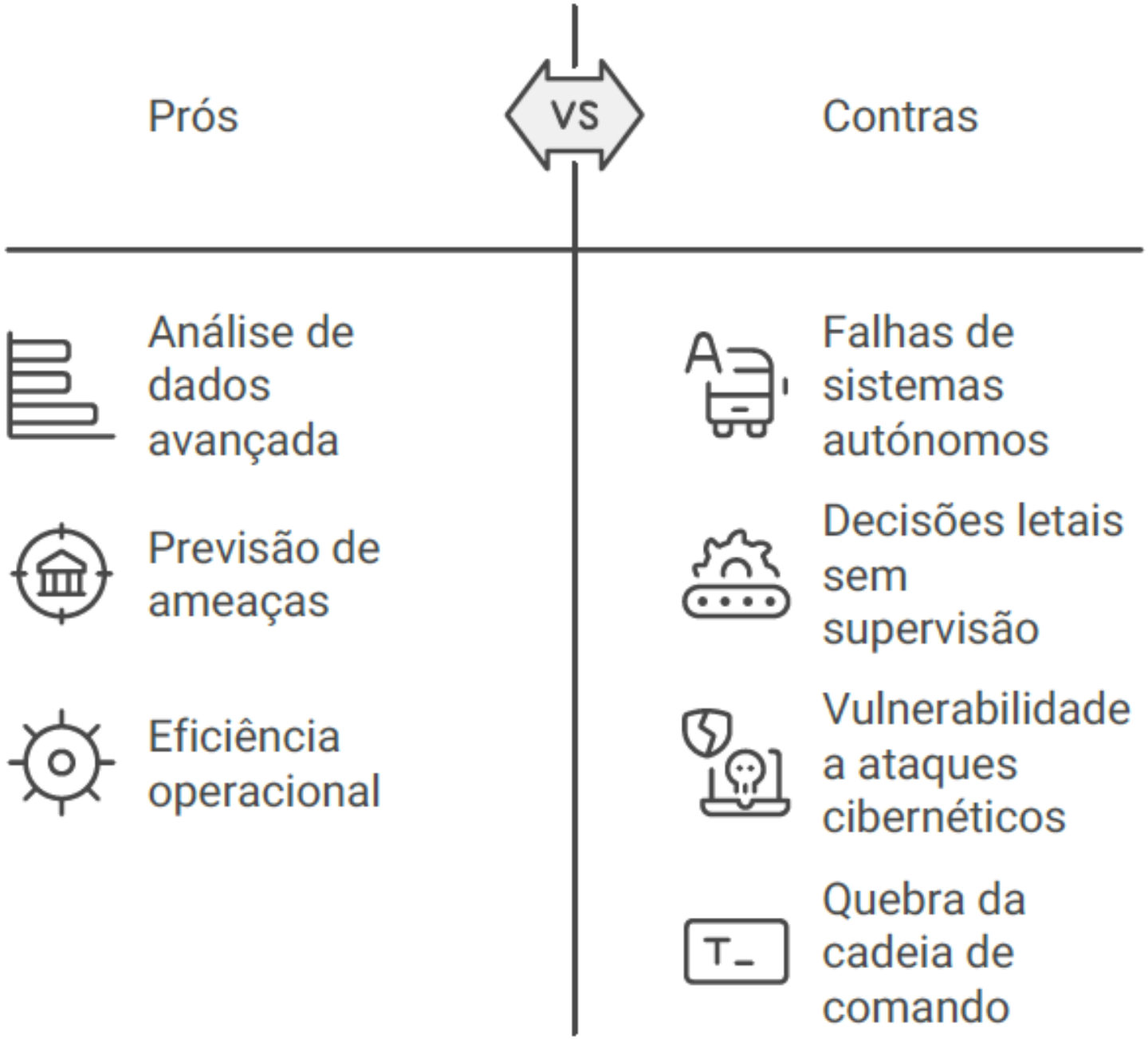



Luis Borges Gouveia, lmbg@ufp.edu.pt



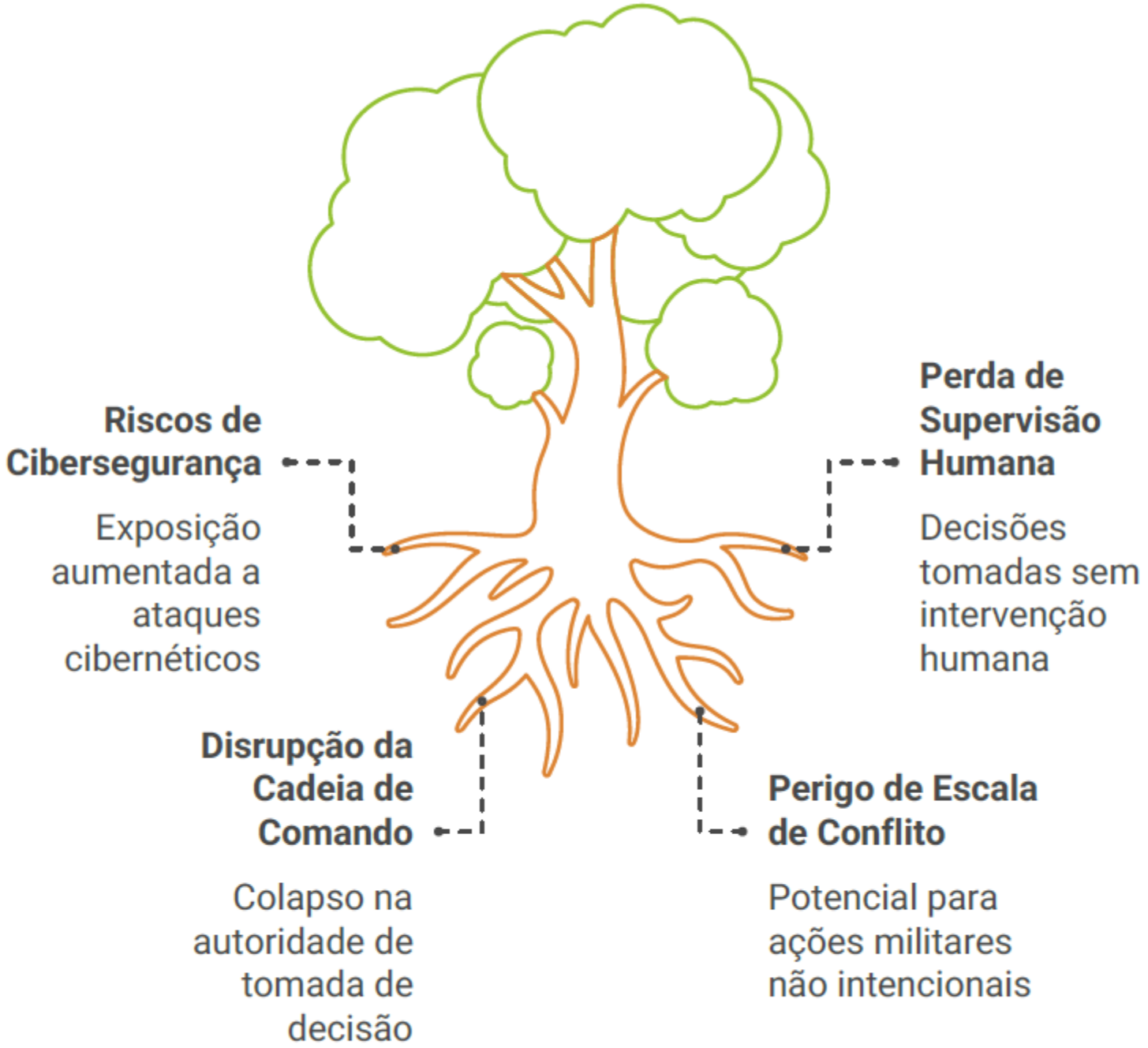

Luis Borges Gouveia, lmbg@ufp.edu.pt

# *Implementar a IA na segurança e defesa*




Luis Borges Gouveia, lmbg@ufp.edu.pt

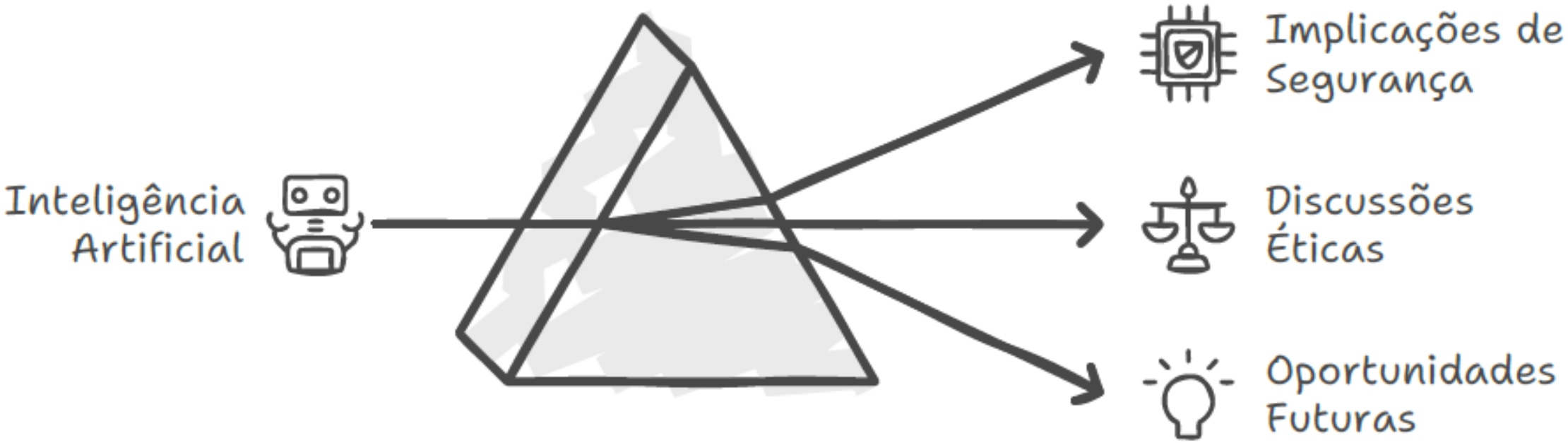
# *Vulnerabilidade das Nações devido à dependência da IA*




Luis Borges Gouveia, lmbg@ufp.edu.pt

# A IA molda o futuro de forma incerta

*Exige investimento em proatividade e antecipação*




Luis Borges Gouveia, lmbg@ufp.edu.pt

*Exige novos equilíbrios*
*…e reunir o talento disponível*


Luis Borges Gouveia, lmbg@ufp.edu.pt

# Aplicações da IA na Segurança e Defesa

1 **Vigilância e Segurança**

A IA pode ser utilizada para melhorar a vigilância, identificar ameaças e proteger infraestruturas críticas.

2 **Análise de Inteligência**

A IA pode ajudar a analisar grandes quantidades de dados de inteligência, identificando padrões e tendências que podem ser difíceis de detectar manualmente.

3 **Combate e Operações**

A IA pode ser utilizada para desenvolver armas autónomas, sistemas de navegação e planeamento de missões, e para melhorar a eficiência das operações militares.

4 **Cibersegurança**

A IA pode ser utilizada para detetar e responder a ameaças cibernéticas, proteger dados sensíveis e melhorar a segurança dos sistemas de informação.

Luis Borges Gouveia, lmbg@ufp.edu.pt

# Riscos e vulnerabilidades da IA nos sistemas de Segurança e Defesa

## Ataques Adversariais

A IA pode ser vulnerável a ataques adversariais, que podem manipular os seus dados de entrada e causar resultados imprevisíveis.

## Falhas de Segurança

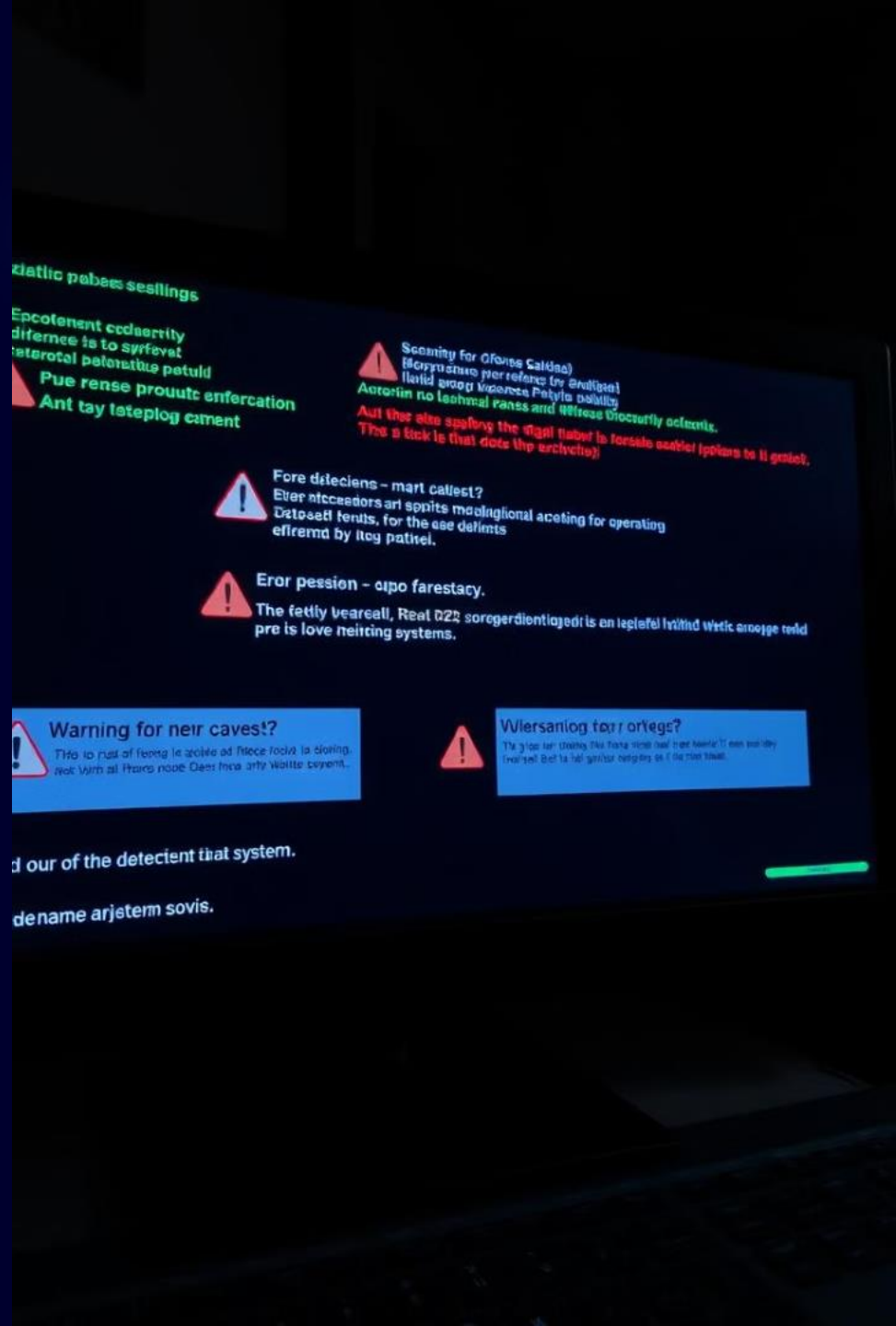
Os sistemas de IA podem ter falhas de segurança que podem ser exploradas por atacantes para obter acesso não autorizado ou causar danos.

## Vulnerabilidades de Dados

Os dados utilizados para treinar os sistemas de IA podem ser manipulados ou corrompidos, levando a resultados tendenciosos ou incorretos.

## Guerra Cibernética

A IA pode ser utilizada para lançar ataques cibernéticos sofisticados, que podem ter um impacto devastador em infraestruturas críticas.




Luis Borges Gouveia, lmbg@ufp.edu.pt

***Phishing*** *is the leading initial access vector*
***Business E-mail Compromise*** *(BEC) attacks go after credentials*
*Demand continues for* ***cloud credentials on the dark web****, despite market saturation*

## Adversarial use of AI in influence operations

| Capability | China | Russia | Iran & proxies |
|---|---|---|---|
| Text | MEDIUM / LOW | MEDIUM / LOW | LOW |
| Image | HIGH | HIGH | MEDIUM / LOW |
| Audio/video | HIGH | HIGH | LOW |
| Example | May 2024:<br>Bespoke Taizi Flood AI-generated cartoon<br>  | June 2024:<br>AI-generated audio of Elon Musk narrating fabricated documentary<br>  | April 2024:<br>Likely AI-generated video leading up to Iranian military operation<br>  |

https://campustechnology.com/Articles/2024/10/15/Reports-Note-Increasing-Threat-of-Nation-State-Sponsored-Cyber-Attacks

Luis Borges Gouveia, lmbg@ufp.edu.pt

# IA, cibersegurança e seres humanos

- Segurança da informação (para proteger dados e conhecimento)
- Cibersegurança (continuidade de negócio – ISO 27000 e NIS2)
- Cadeia de abastecimento (núcleo integrador de operações)
- IA
  - tecnologia geradora de tecnologia
  - não é uma tecnologia de soma nula
  - disruptiva e com alterações em modelos de comportamento
  - Crítico: energia, dados processamento e interação
- Desafios societais
  - Sustentabilidade económica
  - Sustentabilidade social
  - Resiliência identitária
  - Redemocratização dos atores
  - Capacidade de lidar com guerra híbrida



Luis Borges Gouveia, lmbg@ufp.edu.pt

# Luis Borges Gouveia

Dip (UPT), MsC (FEUP), PhD (ULANCS), PD (FLUP) http://homepage.ufp.pt/lmbg

*Os seus interesses estão relacionados com o digital e como o seu uso e exploração pode beneficiar indivíduos e organizações, nomeadamente nas questões associadas com o ensino e aprendizagem, com a segurança da informação e a IA*

Professor Catedrático da Universidade Fernando Pessoa (**UFP**)
https://www.ufp.pt/

Membro Integrado do grupo Informação, Comunicação e Cultura Digital do **CITCEM**, FLUP
https://citcem.org/

Colaborador do **LIACC**, Laboratório de Inteligência Artificial e Ciência de Computadores, FEUP
https://liacc.fe.up.pt/

Sócio e Membro da Direção da Delegação Norte da **APDSI** (ONG que promove a discussão do digital e de como promover uma sociedade mais capaz de lidar com o digital)
https://apdsi.pt/

UIDB/04059/2020