AI & Partners

Amsterdam - London - Singapore

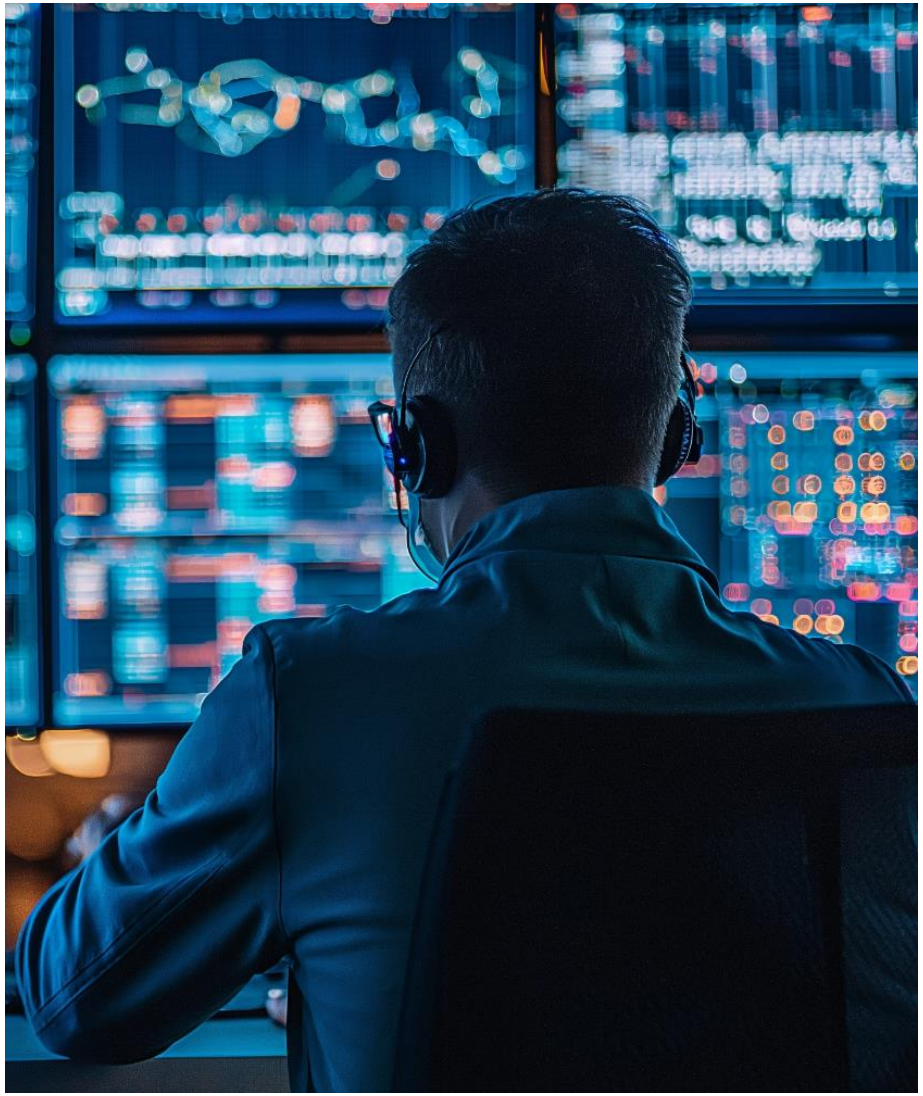
The Digital Commonwealth

The Digital Commonwealth

"Can we maintain control or will we be at the mercy of AI"

Mansion House Summit, Monday, 26th February 2024

Careful control ≡ good governance



- Mandate under the EU AI Act for providers of general-purpose AI models
- Detailed description of evaluation strategies required in technical documentation
- Components:
  - Evaluation criteria
  - Metrics
  - Methodologies for identifying limitations
- Description of measures for internal and/or external adversarial testing (e.g., red teaming)
- Requirement for model adaptations
- Aim: *Rigorous evaluation and continuous monitoring of AI systems*
- Ensures control over performance and impact
- Mitigates potential risks or failures

Get your fundamentals right

- Requirement for Fundamental Rights Impact Assessment before deploying high-risk AI systems
- Assessment includes:
  - Deployer's processes
  - Affected categories
  - Specific risks of harmh
  - Human oversight measures
- Aim: *Consideration and mitigation of risks to fundamental rights*
- Emphasis on human control and oversight over AI application

‘Human-AIn-the-loop’

- Legislation emphasizes cybersecurity measures for AI systems
- Detailed information required about monitoring, functioning, and control
- Highlights need for human oversight measures
- Specifies requirements for evaluating AI system performance in post-market phase
- Includes provision for post-market monitoring plan
- Aim: *Ensure AI systems remain secure, reliable, and under human control*
- Adaptation to new threats and evolution based on real-world performance

Thank you!

AI & Partners

Amsterdam - London - Singapore

The Digital Commonwealth

Email

contact@ai-and-partners.com

Phone

+44(0)7535 994 132

Website

https://www.ai-and-partners.com/

Social Media

LinkedIn: https://www.linkedin.com/company/ai-&-partners/
Twitter: https://twitter.com/AI_and_Partners