datrix
AI SOLUTIONS GROUP

AIA GUARD

Forum Cyber 4.0
6 / 7 Jun 2023

Claudio Zamboni
Co-Founder & Chief Revenue Officer, Datrix SPA

Claudio Zamboni

Co-Founder & Chief Revenue Officer @ Datrix

claudio@datrixgroup.com

DATRIX Group

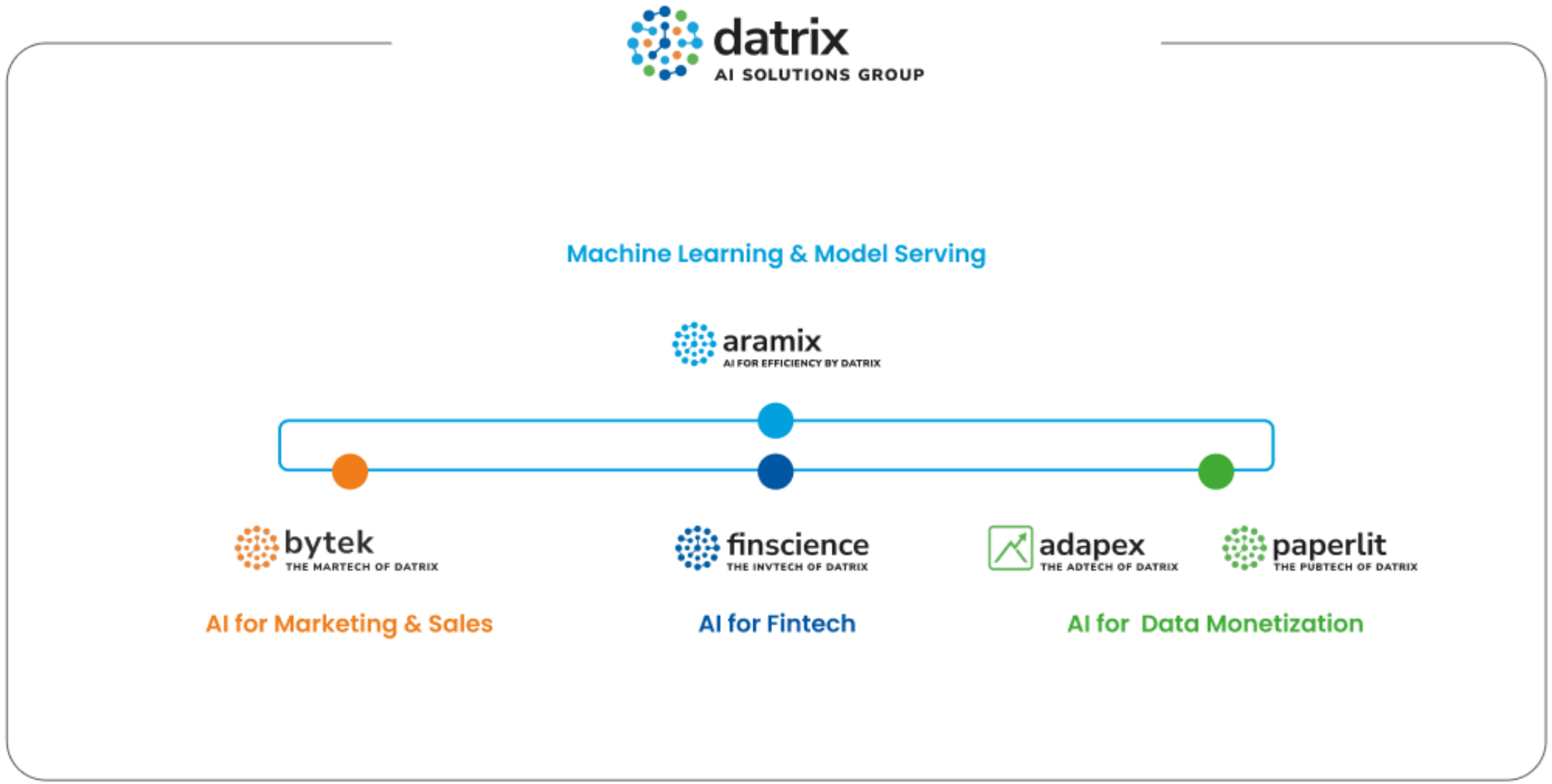
https://datrixgroup.com

Datrix is an Italian SME with headquarters in Milan and offices in Rome, Viterbo, Cagliari and New York

Datrix SpA is listed on Euronext Growth Milan

~100 employees
~€16M turnover in 2022

DATRIX: Sustainable AI solutions for data-driven Business Growth

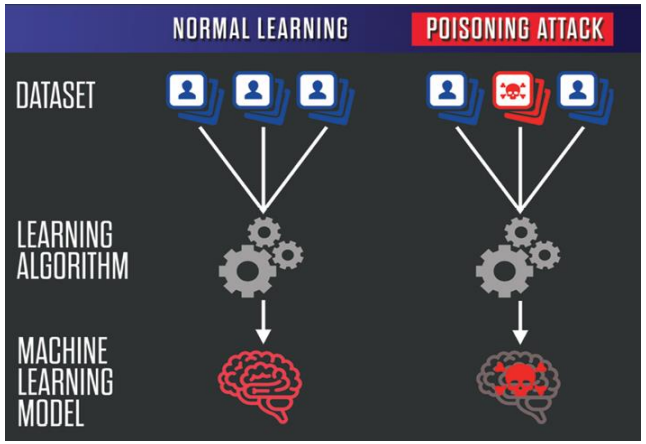


Artificial intelligence is now part of our everyday lives

Artificial intelligence can be both a blessing and a curse for cybersecurity

Artificial Intelligence Attacks Taxonomy

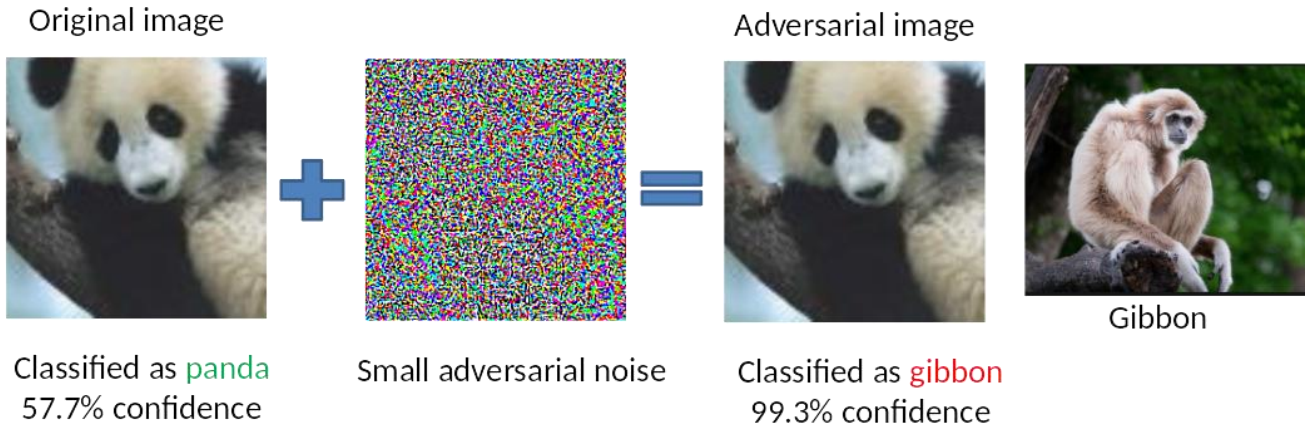
Data poisoning → *Training time*



Most of the production ML models retrain their models periodically with new data.

Data poisoning attacks manipulate training data to induce misclassification to a specific test sample or a subset of the test sample

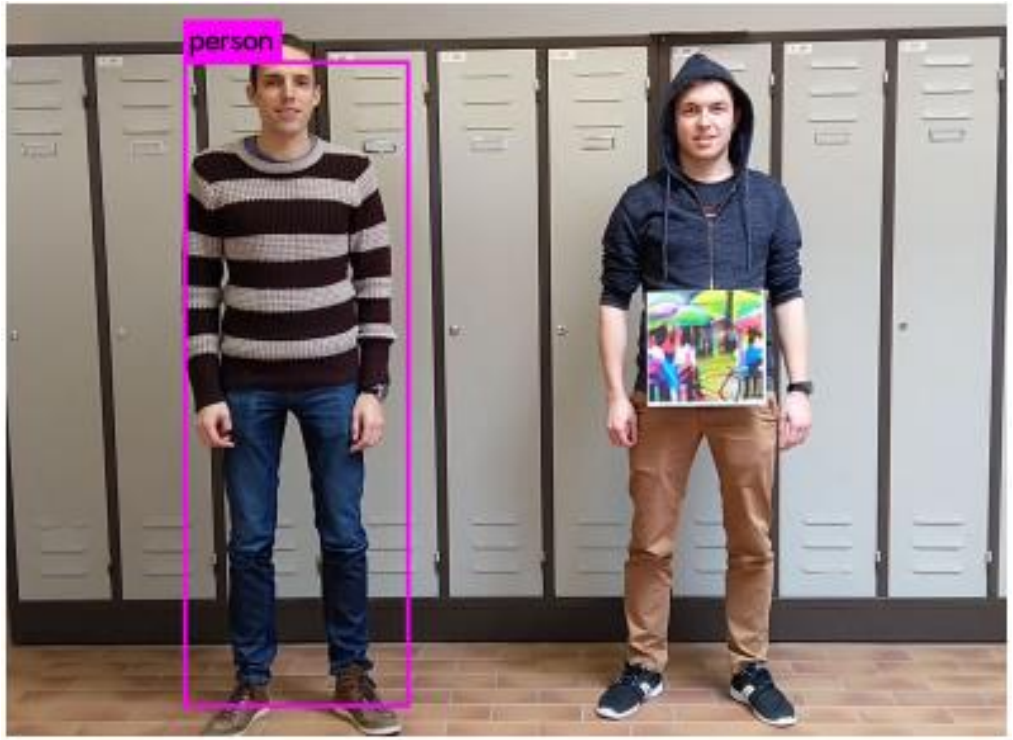
Adversarial attack → *Inference time* (Evasion)



An evasion attack happens when the network is fed an “adversarial example” — a carefully perturbed input that looks and feels exactly the same as its untampered copy to a human — but that completely throws off the classifier

Artificial Intelligence Attacks - Evasion examples

Artificial Intelligence Attacks - Evasion examples



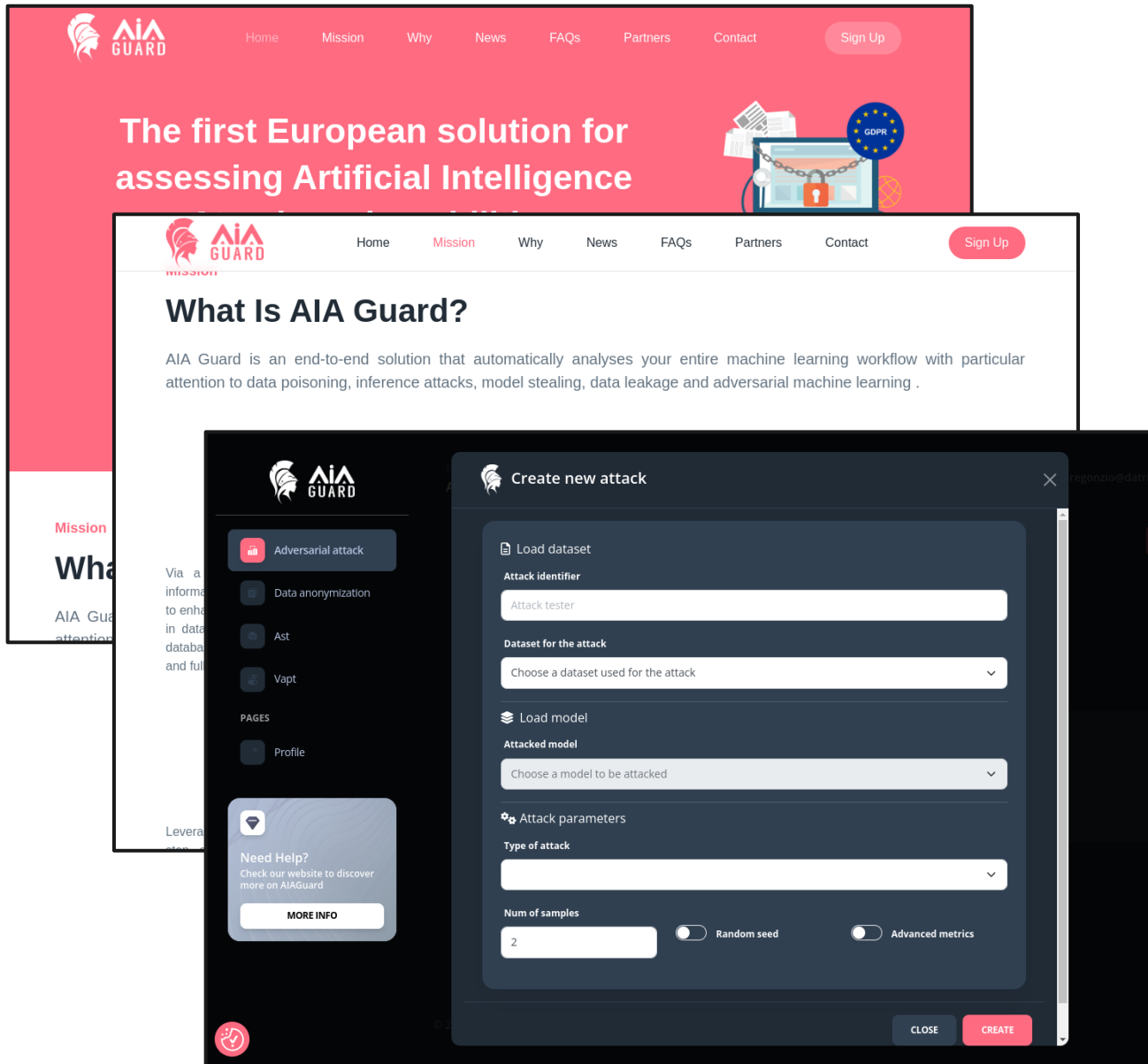
https://www.amazon.com/Adversarial-Anti-Facial-Recognition-Camouflage-Invisibility/dp/B085NFJQM4

Our platform

AIA Guard performs a comprehensive evaluation of the security and privacy risks during the entire AI lifecycle: from dataset creation to model training and the production release.

AIA Guard leverages **Machine Learning** algorithms to analyse dataset, model hardening, source code analysis, and risk evaluation.

The very first European solution for assessing Artificial Intelligence Attacks vulnerabilities



GDPR COMPLIANT

www.aiaguard.com

AIA Guard modules

AST

Application Security testing

- Identifies security weaknesses and vulnerabilities in the source code of the machine learning application.
- Scans software dependencies for known vulnerabilities.

DATASET ANALYSIS

Detect sensitive information

- Identifies opportunities for data leaks on sensitive data and personal information (phone numbers, emails, zip code, etc).

ADVERSARIAL ATTACK

Evasion analysis

- Focused on textual ML models
- Hijack the model toward a misleading behavior generating adversarial examples that can fool the target classifier

VAPT

Vulnerability assessment penetration test

- Detects and try to exploit vulnerabilities on the exposed services and APIs which can make them subject to malicious activities.

REPORT GENERATION

Actionable insights

- Generates clear reports of all the executed procedures, suggesting corrective actions when possible.

datrix
AI SOLUTIONS GROUP

Thanks !