XAI:
ARE WE LOOKING BEFORE WE LEAP?

sosə

CONTENT

This four part report highlights the challenges and advances when it comes to regulations, relevant use-cases, and emerging technologies taking the mystery out of AI.

Explainable AI is getting a lot of attention lately, and for a good reason. Netflix's recent release of its eye-opening and controversial docudrama, "The Social Dilemma" has brought the discussion to the mainstream. Directed by Jeff Orlowski, the film features former top executives that worked at the largest tech companies and highlights their growing concerns regarding the inherently deep, dark cultural effects of social media.

The interviewees sound the alarm on how the public is unknowingly being manipulated. The former employees share revealing insights such as. "If you're not paying for the product, then you're the product", "It's 2.7 Billion Truman shows", and "It's the gradual, slight, imperceptible change in our own behavior and perception that is the product."; creating shockwaves in the media and dinner tables.

While it's ironic that Netflix is the one blowing the whistle, they have a salient point. In a remarkable scene following the dramatization of a content-recommendation algorithm in action, the former executives reveal how misunderstood the algorithms actually are. "There are only a few people who understand how those systems work, and even they don't necessarily fully understand what's going to happen..." a former employee asserts.

Explainable AI (XAI) is an emerging field in machine learning that addresses the lack of transparency in traditional AI algorithms and machine learning such as those depicted in the film. In addition to ensuring transparency, XAI also promotes accountability, traceability, and auditability. It is designed to explicitly contradict the "black box" concept in machine learning, in which even algorithm designers cannot explain why the AI model arrived at a specific decision.

Driven by the need to comply with regulations and industry players' pursuit to understand the rationale behind decisions reached by algorithms, XAI is seeing a rise in its adoption. Corporations are working to create monitored and comprehensible AI models while continuing to benefit from big data and unimpaired computing power.
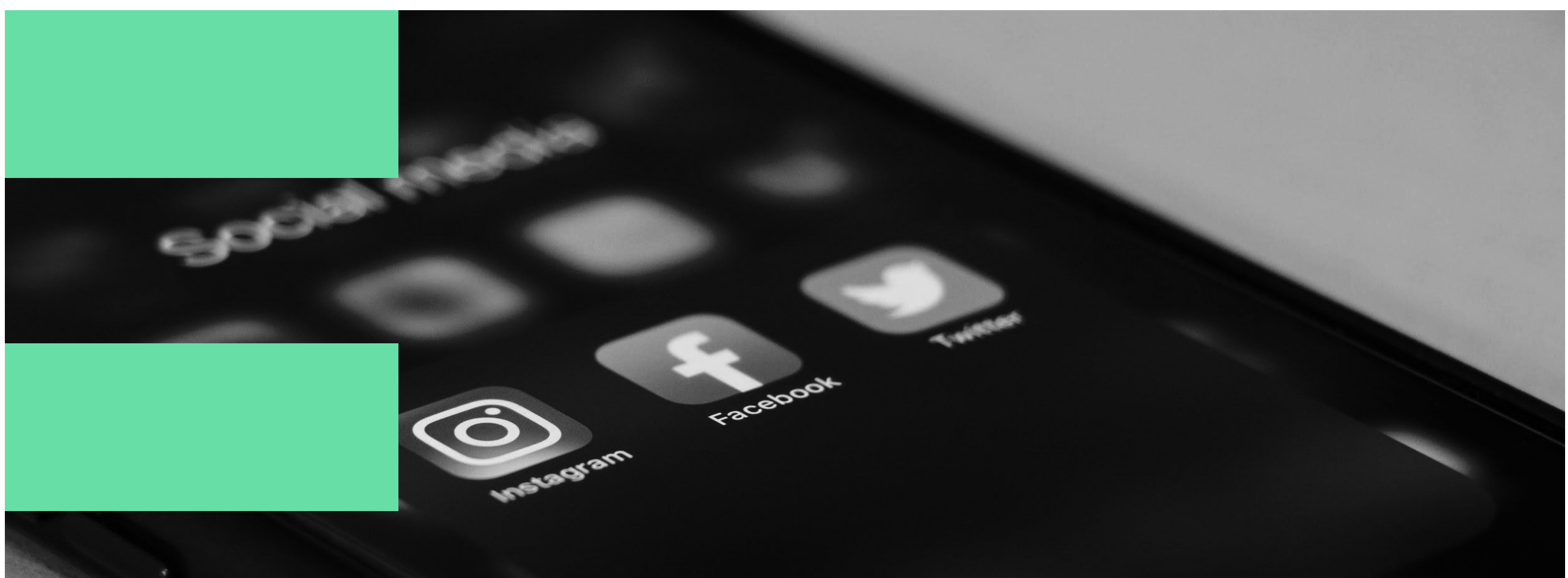
01

XAI
REGULATIONS

In an attempt to implement an ethical and transparent use of AI, nations and trade organizations are establishing guidelines and regulations. These frameworks are provided to boost trust in AI, and by extension, the implementation of XAI. While there are currently several governments leading the way in regulating XAI, we foresee a reality in which this practice will become a global standard in order to continue benefiting from AI, without losing control over it.

EU

The High-Level Expert Group on Artificial Intelligence released the Ethics Guidelines for trustworthy AI. As the UN aims to regulate the requirements, these guidelines provide a clear roadmap for future legislation. The EU Data Protection Regulation (GDPR) also requires that profiling and decisioning systems, typically equipped with AI, are comprehensible.

UNITED STATES

The White House has released 10 principles for government agencies to adhere to when proposing new AI regulations for the private sector; the move follows the American AI Initiative. Although we are used to seeing a hands-off approach, this initiative may force corporations to accelerate a wide adoption of explainable AI.

CHINA

The 'New Generation Artificial Intelligence Development Plan' (AIDP) outlines China's AI policy objectives. By 2030, China seeks to become the world's center for AI. While pursuing material goals, the AIDP released eight principles that emphasize how AI development should begin by enhancing the well-being of humanity. The plan highlights the importance of transparency, responsibility, collaboration, and agility to address new and emerging risks. Although still vague, we are set to see new regulation being developed around explainability.

02

CORPORATIONS MAKING STRIDES: Notable Use Cases

In heavily regulated industries such as finance - transparency, accountability, and trustworthiness of data-driven AI decisions is essential. For these reasons, leading consultancies such as Deloitte and PWC are first in line to roll out XAI methodologies to support their clients and mitigate risk.

Deloitte.

Deloitte

Deloitte aspires to support clients in highly regulated industries through "safe" innovation that leverages AI and simultaneously complies with ever-evolving regulation. Deloitte aiStudio developed an XAI tool, Lucid ML, to provide customers with a better understanding of the ML models' inner workings and the drivers behind specific risk resolutions.

pwc

PwC

PwC partnered with an auto insurer to automate and accelerate the claim estimation process. During the project, PwC developed several AI and analytical tools that successfully detected and processed details the estimators had missed. To build further trust and transparency in the platform, PwC implemented XAI, creating simplified visualizations illustrating the logic behind the predictions and determinations.
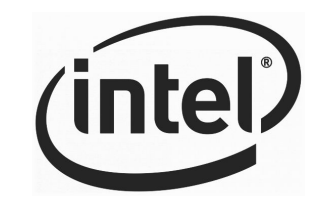
Intel

The world's largest semiconductor chip manufacturer partnered with DarwinAI to research the algorithm's transparency and explainability. Jointly, they have increased AI performance on a variety of hardware offerings. The developed solutions are more compact, accurate, and tuned for the device where they are deployed.

Lockheed Martin

The aerospace and defense giant sought to improve customers' understanding and trust in their AI solutions by providing them with complete visibility into their production AI systems while ensuring high performance. LMT announced a strategic collaboration with DarwinAI to improve neural network efficiencies and create more accurate and robust AI models.

The company also announced a strategic collaboration and investment in Fiddler, an XAI startup, to build trust in AI models by stationing explainability as an added value..

Wells Fargo

The financial services company implemented H2O's XAI solutions to gain actionable insights and manage potential risks in AI models. The company wanted to understand the challenges and opportunities explainable AI face in order to comply with regulations whilst maintaining and enhancing customer trust.
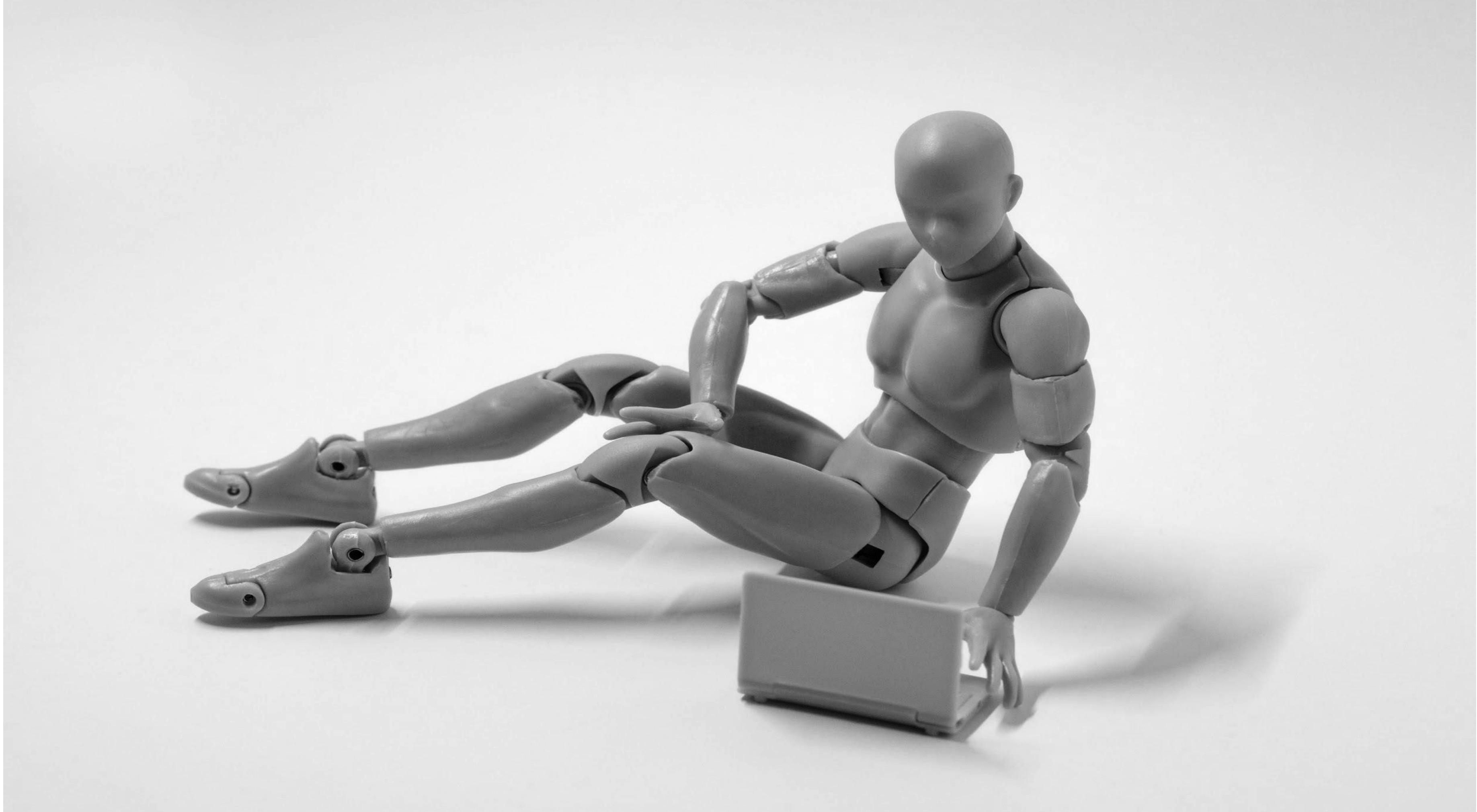
03

INTERVIEW SPOTLIGHT

The future of XAI with Dr. Luca Marighetti
Group Head Tech Transformation at Swiss Re

🎗 **Swiss Re**

We had the pleasure of speaking with Dr. Luca Marighetti, Group Head Tech Transformation at Swiss Re, SOSA's valued partner. A prominent leader in digital risk perception, Dr. Marighetti shared his insights, concerns, predictions, and solutions on the future of XAI.

What are the dangers of AI that cannot be explained? How does it affect the world of insurance?

"While AI is a great tool, it also leads to mistakes and hence damage, things can go wrong along the way. We call this Algorithmic Risk. There are two fundamental challenges we face while dealing with algorithmic risk – attribution, liability and retribution. Can you trace back what went wrong? Who is responsible when things go wrong? Who pays in the end, given that you cannot put an algorithm into jail?

So long as AI remains untransparent and unexplainable, it becomes a rather complex matter. Nowadays, more and more people are coming to understand elementary digital risk, seeking to offload that liability. Yet, you can only quantify the risk if AI is explainable. That is where the insurer comes into place - understanding the algorithm, holding the liability, and carrying the risk involved. Our recent work with Hitachi was initiated as we understood that although AI brings them

> As more engineers rely on open code, infected codes can create backdoors that are very hard to track down. That's why interpretability and transparency are so important.

What do you consider digital risk?

"Let me take a step back. It all started when we started to engage with cyber risk resulting from malevolent attacks. At the time, there was a consensus that cyber risk meant malevolent attacks. But then, I had dinner with the philosopher Daniel Dennett, where we discussed the problems of liability in algorithms. When I got back home, I connected the dots: The world is delegating more and more processes and decisions to machines, machines are vulnerable, and the damage potential is substantial. I realized that cyber risk from malevolent attacks is only a fraction of the risks involved, and it is built of several components.

First, malevolent attacks indeed; The intentional exploitation of computer systems, networks, and technology.

Second, legacy system failure; IT systems can fail for various reasons, including hardware or software glitches, power surges, physical perils, and botched upgrades. The Delta case was a remarkable and public incident, where the company simply couldn't figure out what went wrong for days.

Third; algorithmic risk; results from algorithmic complexity and algorithmic malpractice. Concerning as it may be, AI applications may lead to severe consequences. The notorious two fatal Boeing 737 Max 8 crashes in Oct. 2018 and March 2019 allegedly resulted at least partly from an algorithm related matter.

Fourth, human errors; a janitor may accidentally shut down a computer system, and employee might not execute the necessary patches or inadvertently expose critical data to the outer world. Also, open-source is becoming an increasing risk. As more engineers rely on open code, infected codes can create backdoors that are very hard to track down. That's why interpretability and transparency are so important."

How do you foresee the future of AI? Will explainable AI ever catch up with rapid and disruptive innovation?

"I think the algorithmic risk problem is quickly catching a widespread attention, from MIT and Stanford to Tsinghua. At the same time, regulatory pressure is clearly leading to a point where either you make AI interpretable or be out of business. On the other hand, the benefits from AI are significant enough to mobilize the resources needed to solve the problem. So, I am optimistic."

04

TECH COMPANIES TAKING THE MYSTERY OUT OF AI

Listed below are selected companies innovating groundbreaking explainable AI solutions.
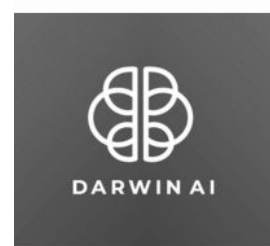
Darwin AI

Developer of GenSynth, a platform that processes AI systems such as computer vision, natural language processing, or speech recognition to yields highly optimized, and compact versions.

HQ: Ontario, Canada
Year founded: 2017
Founding stage: Undisclosed
Disclosed amount raised: $3.4M

Website | Linkedin

Fiddler Labs

An AI engine that enables companies to better understand the rationale behind their machine learning and AI models by analyzing, validating, and monitoring.

HQ: California, United States
Year founded: 2018
Founding stage: 2018
Total raised: $13.2M

Website | Linkedin

Kyndi

An auditable AI software used by critical government and commercial institutions. It analyzes long-form text and acts on opportunities and threats. The platform applies XAI to provide full transparency over the output.

HQ: California, United States
Year founded: 2014
Total raised: $32M
Website | Linkedin

Diveplane

Solutions that provide full understanding and decision-transparency in support of ethical AI policies and data privacy strategies. The patented AI-powered business solutions can be applied across multiple industries such as finance, real estate, and defense.

HQ: North Carolina, United States
Year founded: 2017
Total raised: $6.5M
Website | Linkedin

H2O.Ai

An open-source machine learning automation platform. The platform deploys machine learning and predictive analytics for use in predictive maintenance and operational intelligence, enabling enterprises to access transparent and accountable data products.

HQ: California, United States
Year founded: 2011
Total raised: $146.5M
Website | Linkedin

Superwise.ai

An AI production monitoring platform that decreases data science teams' workload and labor costs by monitoring and analyzing critical data points affecting the models' performance.

HQ: Tel Aviv, Israel
Year founded: 2019
Total raised: $4.6M
Website | Linkedin

Mona labs

An AI-based algorithm quality check system. The platform ensures that inconsistencies or errors do not negatively impact the model's quality while providing users with complete transparency into how the data and models behave in the real world.

HQ: Tel Aviv, Israel
Year founded: 2018
Total raised: $3.9M
Website | Linkedin

Akur-8

An AI insurance pricing platform designed to provide automation and optimization for insurance carriers. The company's platform integrates algorithms dedicated to insurance pricing that can immediately spot anomalies and discover new patterns, enabling insurance carriers to improve their pricing models.

HQ: Paris, France
Year founded: 2018
Total raised: $11.7M
Website | Linkedin

Adopt or Die: XAI on the fast-track for extensive adoption.

COVID-19 has rapidly increased the demand for AI/ML offerings across all sectors. According to a recent study, more than half of enterprises have accelerated their automation initiatives, encouraging the evaluation of AI for new use-cases. The current pandemic has also disrupted many industries with extensive supply and demand shock, coupled with a great amount of uncertainty.

While enterprises, regulators, and customers demand a deeper understanding of these complex models', XAI is on the fast-track to extensive adoption.

CONTACT US.
WITH ANY REQUESTS
OR INQUIRIES.

MATAN BEN-GIGI
Corporate Innovation at SOSA
matan.b@sosa.co