CYRES CYBERSECURITY | IN COLLABORATION WITH DOX DEFENSE TECH

Seeing the Invisible

Shadow AI and Modern Cyber Risk

Dvir Geva

Consulting CISO, Cyres Cybersecurity

CISSP, CISM, Lead Auditor ISO 27001 and ISO 42001

Johan Van Dam

CEO, Dox Defense Tech

AGENDA

What we will cover

1. **The Shadow AI problem**
   A risk already inside your clients' businesses
2. **Five categories of AI risk**
   What changes, what stays the same
3. **Four myths and the new reality**
   Including a USD 25 million case study
4. **Cyber hygiene with an AI lens**
   A practical framework you can use
5. **Building resilience**
   How we help, and how brokers benefit

A QUICK SCENARIO

It is 10:42 on a Tuesday morning

An employee at one of your commercial clients

...pastes a customer's confidential information into a free public AI tool to "speed up a quote."

Nobody approved it. Nobody knows it happened. The data is now sitting on a server they do not control, may form part of a future training dataset, and is — to put it plainly — outside the business.

This is not a hypothetical. It is happening, today, in organisations across every sector your clients operate in.

DEFINITION

What we mean by "Shadow AI"

Shadow AI

AI tools used inside organisations without the awareness, approval, or governance of IT or leadership.

Think ChatGPT in the browser, Copilot inside a personal Office account, an AI plug-in quietly added to a CRM — none of it on the IT asset register.

78%

of AI users at work bring their own AI tools (BYOAI). 80% at small and medium-sized companies.

Microsoft & LinkedIn Work Trend Index, 2024

+USD 670K

added to the average data breach cost in 2025 when Shadow AI was a contributing factor.

IBM Cost of a Data Breach Report 2025

SECTION 2

Five categories of AI risk your clients now face

1

Data leakage

Sensitive data sent to public AI tools, beyond the company's control.

2

Decision risk

Business decisions made on AI output that may be wrong or hallucinated.

3

Operational dependency

Reliance on an AI tool that can change, fail, or vanish overnight.

4

Supply-chain AI

Clients' own vendors quietly adding AI features to existing tools.

5

Generative content

AI-generated outputs that may breach IP, professional, or regulatory standards.

AI RISK 1 OF 5

Risk 1 — Data Leakage

1

The problem

Proprietary or client data is sent to public AI tools where it may be stored, processed, or used to train future models. Once it leaves the business, it cannot be recalled.

EXAMPLE SCENARIO

A bookkeeper uses ChatGPT to "reconcile this messy CSV." The CSV contains every customer's banking details. Within seconds, that data is on a server in another jurisdiction, potentially feeding a future model.

ASSESSMENT

"Is the business aware of what data its staff are putting into AI tools? Is there any policy in place?"

AI RISK 2 OF 5

Risk 2 — Decision Risk

2

The problem

AI outputs can be wrong, biased, or hallucinated — but they sound confident. Decisions made on that output carry real consequences, and the question of accountability has not been resolved.

EXAMPLE SCENARIO

A junior staff member asks an AI assistant whether a particular indemnity clause is enforceable. The AI says yes, confidently. The client relies on it. Six months later, in court, that answer turns out to be wrong.

ASSESSMENT

"What decisions are being made on AI output — and who is on the hook if those decisions are wrong?"

AI RISK 3 OF 5

Risk 3 — Operational Dependency

3

The problem

Businesses are increasingly building real workflows on top of AI tools. When those tools change pricing, change terms, or disappear, the dependency becomes a liability.

EXAMPLE SCENARIO

A logistics firm builds its dispatch workflow around a third-party AI plug-in. The vendor changes its API. Overnight, the workflow breaks. The business has no fallback because the dependency was never documented.

ASSESSMENT

"If this AI tool went away tomorrow, what would be lost? Is there even a list of where the business is depending on AI?"

AI RISK 4 OF 5

Risk 4 — Supply-Chain AI

4

The problem

Vendors and SaaS providers are quietly adding AI features to existing products. Client data flows into those features — often without explicit notification or consent.

EXAMPLE SCENARIO

A property management firm's CRM rolls out a new "AI assistant" toggle. It is on by default. Tenant data has now started passing through a third-party language model. The MD finds out eight months later.

ASSESSMENT

"Have existing tools added AI features that nobody asked for — and where does that data now flow?"

AI RISK 5 OF 5

Risk 5 — Generative Content Risk

5

The problem

AI-generated content can infringe intellectual property, misrepresent the business, or quietly breach professional or regulatory standards — sometimes all three in the same document.

EXAMPLE SCENARIO

A marketing assistant generates an article using AI. The article includes paragraphs lifted, in substance, from a competitor's blog. It goes out under the company's name. The competitor's lawyers notice within a week.

ASSESSMENT

"What content is being published that was AI-generated — and what may the business be exposed to as a result?"

SECTION 3

Four myths SMBs still hold

And the verified data that contradicts them.

Myth 1
Too small to be a target

Myth 2
Time to patch before exploit

Myth 3
Antivirus protects us

Myth 4
AI helps defenders more

MYTHS 1 & 2

On being a target, and on having time

MYTH 1

"Cybercriminals target big companies."

REALITY

+34% rise in successful exploits, year-over-year.

Automation has made small and mid-sized businesses the most efficient targets. Lower defences, faster payouts, less negotiation. The 2025 Verizon DBIR documented a 34% year-over-year increase in successful vulnerability exploitation across the dataset.

Source: Verizon DBIR 2025

MYTH 2

"If a vulnerability is disclosed, we have time to patch."

REALITY

Median time from disclosure to exploit: zero days.

For new critical vulnerabilities on edge devices in 2025, the median window between public disclosure and mass exploitation was zero days. Average time-to-exploit overall: 5 days. Average time to patch edge devices: 209 days.

Source: Verizon DBIR 2025

MYTH 3 — THE CASE STUDY

When walls and antivirus are not the perimeter

MYTH 3 *"Our antivirus and firewall protect us."*

USD 25.6M

Lost in a single fraud

Arup engineering, Hong Kong office

Early 2024

How it happened

A finance staff member at Arup's Hong Kong office joined a video call. **Every person on that call was a deepfake.**

The "CFO" and several "colleagues" — all AI-generated — instructed the employee to authorise fifteen transfers to five bank accounts. Total losses: USD 25.6 million.

No firewall, no antivirus, and no patching programme stops this. The attack target was a human being making a judgement call on a video.

Sources: Fortune, CNN, CFO Dive — May 2024

MYTH 4 & THE SHIFT

From walls to resilience

MYTH 4 *"AI helps defenders more than attackers."*

Both sides benefit. But attackers iterate without compliance, ethics, or governance overhead — defenders carry that weight. The advantage flows to whoever can move faster, and right now, that is the attackers.

THE SHIFT — FROM WALLS TO RESILIENCE

Visibility

Seeing what is actually happening in your environment.

Detection Speed

Knowing fast when something is wrong.

Response Capability

Acting effectively when it matters.

Meet Mythos — and why it changes the math

WHAT IS MYTHOS?

Anthropic's Mythos demonstration showed AI systems autonomously discovering new software vulnerabilities, building working exploits, and conducting complex automated attacks — at machine speed.

It is no longer a question of whether AI can do this work for attackers. It already does.

WHAT IT MEANS FOR DEFENDERS

1. **The vulnerability supply is rising — fast.**
   AI can surface flaws in code at a pace no human team can match.

2. **Time-to-exploit keeps collapsing.**
   What used to be days between disclosure and active exploitation is now hours, sometimes less.

3. **The asymmetric advantage favours attackers.**
   Patching has structural limits that AI-driven offensive tooling does not share.

Source: "The AI Vulnerability Storm: Building a Mythos-ready Security Program." Cloud Security Alliance CISO Community, SANS Institute, OWASP Gen AI Security Project. Released April 2026; updated May 2026.

SECTION 4

Cyber hygiene, now with an AI lens

Traditional Cyber Hygiene	AI-Era Equivalent
Know what hardware and software you have	Know what AI tools your people are using
Control who has access to what	Control who can use AI with what data
Patch known vulnerabilities	Monitor AI vendor and model changes
Train employees to spot phishing	Train employees on safe AI use and AI-generated social engineering
Back up your data	Govern what data leaves the business via AI tools
Have an incident response plan	Include AI-specific incidents (deepfake fraud, AI data leak)
Vet your vendors	Vet your vendors' AI use, and AI vendors specifically

If your client cannot answer the basic visibility questions in 30 seconds —

that is a gap worth a conversation.

The 30-second test

HOW WE HELP

The Cyres approach — Map. Assess. Resolve. Monitor.

1 Map

Discover the full environment, including shadow IT and shadow AI. You cannot protect what you cannot see.

2 Assess

Identify the risks that actually matter for that specific business. Not a generic checklist — a contextual assessment.

3 Resolve

Practical, prioritised steps to improve resilience, with senior CISO-level guidance throughout.

4 Monitor

Ongoing oversight as the environment evolves. Resilience is not a one-off project — it is a posture.

DELIVERY

Three ways Cyres engages with clients

Advisory

Strategic guidance, without a full-time hire.

Periodic or on-demand CISO leadership, assessments, and board-level advice. We bring senior expertise to where it is needed, when it is needed.

Typically: retainer or scoped project

Build

Closing the gap, with measurable outcomes.

Project-based deployment of controls, processes, and frameworks. Defined scope, defined timeline, structured handover.

Typically: fixed scope, fixed term

Run

Resilience, maintained over time.

Ongoing services and tooling that keep the cybersecurity posture aligned as the business and threat landscape evolve.

Typically: ongoing, subscription-style

FOR THE ROOM

Why this matters to you as brokers

1

Better-protected clients carry better risk profiles.

And better risk profiles tend to make for easier renewal conversations.

2

Brokers who can refer credible cybersecurity advice become more valuable advisors.

Especially as clients increasingly expect their broker to be the first call when something has gone wrong.

3

Partnering with Dox Defense Tech for a cross-discipline local offering.

Bringing comprehensive cybersecurity solutions to East London and the entire Eastern Cape through meaningful collaboration with local security tech leaders, providing holistic risk reduction to your doorstep.

THE NEXT STEP

If a 30-minute conversation would help.

For follow-up conversations with any of your clients on cyber visibility, Shadow AI discovery, or resilience-building, please reach out to our partner:

PARTNER CONTACT

Johan Van Dam

Dox Defense Tech

info@doxdefense.tech

Thank you.

Questions?

Ensuring technology is used to build rather than to destroy.