German-American IT Security Forum

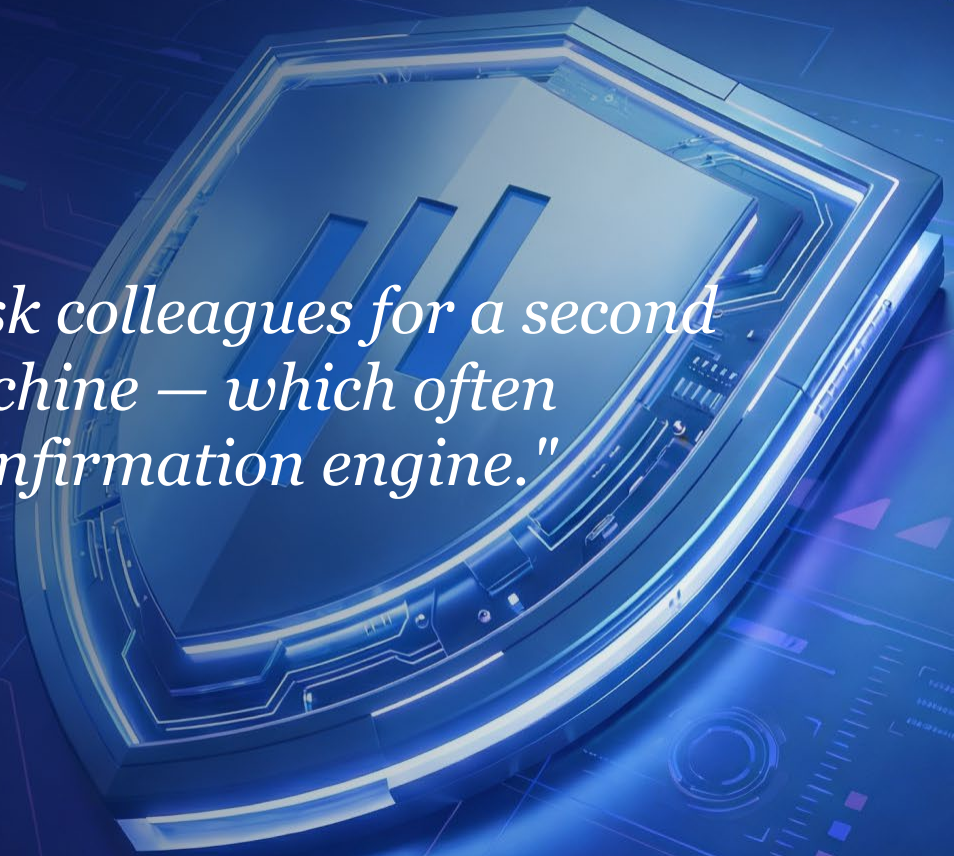
San Francisco, 23.03.2026

AI: Productivity Boost, Skill Atrophy, Security Exposure

Direnc Koyun, KOBIL

Why This Talk?

"People who used to ask colleagues for a second opinion now ask a machine — which often functions as a blind confirmation engine."



AI makes us more productive. That part is real.

- **BCG consultants:** **25% faster,** measurably higher quality (Harvard Business School)
- **Developers:** **55% faster** task completion with Copilot (GitHub)

> *The question is not whether these gains are real. It is what we are trading for them.*

AI helps spectacularly — until it doesn't. And we can't tell the difference.

- **The "jagged technological frontier":** an invisible boundary between where AI helps and where it harms
- **You feel productive on both sides.** Output looks competent on both sides.
- **METR study (July 2025):** developers were 19% slower with AI — but estimated they were 20% faster

> *When feedback becomes unreliable, offloading becomes rational — even when it is objectively harmful.*

We are not losing all skills. We are losing the skills we need to oversee AI.

- The more confident AI's output, the less effort humans invest in evaluating it
- Experienced physicians: measurably worse at detecting cancer after 3 months with AI

> *A tool that makes you better while you use it but worse when you stop is not augmentation. It is dependency.*

When confidence erodes, we stop using AI as a tool. We start asking it for direction.

- “Is this right?” → “How should I approach this?” → “What should I do?”
- A hammer doesn’t tell you what to build. An LLM often does.
- The shift: executor → advisor → decision-maker

> *It starts the moment you stop trusting your own judgment and start asking the machine.*

Shadow AI bypasses visibility and control.

- A large share of enterprise AI use happens outside managed environments
- Personal accounts, browser plugins, copy-paste to public endpoints — invisible to security teams
- It is not negligence. It is rational behavior. The benefit is immediate - the risk is abstract.

> *Security controls require identity, logging, DLP, and policy enforcement. Unmanaged accounts bypass all of it.*

Threat model:

AI raises throughput while lowering review depth, and adds new execution and dependency surfaces.

“Works” ≠ “Secure”

- AI-generated code optimizes for function, not security
- AI hallucinates software packages consistently — attackers register them as malware (**Slopsquatting**)
- AI agents: **Prompt injection** = breach at machine speed with legitimate credentials

The Vicious Circle

Boost → Offload → Lower verification → More incidents & leaks → More reliance → Weaker self-assessment → ↻

> *The loop is self-concealing. The degradation in capability reduces the capacity to recognize the degradation.*

AI raises the floor for individuals. It lowers the ceiling for the group.

- **Writers with AI:** individually more creative, collectively more similar
- **The “Creative Scar”:** creativity drops below baseline after AI is removed — and stays there
- LLMs optimize for the most probable next token. Originality is, by definition, improbable.

> *When everyone uses the same tools, trained on the same data, the output converges.*

Knowledge is not burning. It is being overwritten.

1. **Trust Advantage** — AI delivers, trust is earned
2. **Contaminated Sources** — AI-generated content is a significant and growing share of the web
3. **Closed Loop** — AI trains on AI output, models collapse toward the mean
4. **Epistemic Conformity** — AI becomes the default interface to knowledge

> *The library does not burn. It simply becomes smaller, one probable token at a time — while the catalog insists it contains everything.*

This is not Terminator. This is Wall-E.

- We optimize for comfort and offload what is hard
- The next generation may never acquire an AI-independent baseline
- The costs are deferred, diffuse, and self-concealing

Three principles. Not awareness — controls.

Visibility: Managed accounts, logging, DLP, policy enforcement.

Verification gates: Security scanning, dependency checks, human explainability.

Agent controls: Default-deny, least privilege, sandboxed execution, full action audit.

The Digital Library of Alexandria does not burn. It simply becomes the only library anyone remembers how to read.

The Power House in security and digital identity

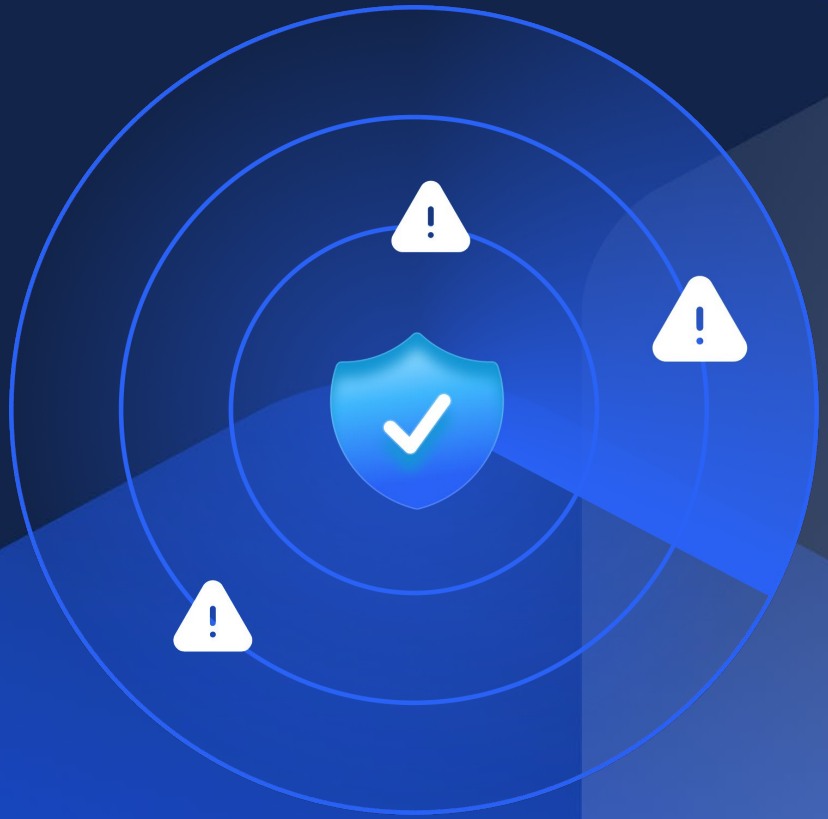
Trusted by more than **5000+ companies** **150M+ user** since more than **38+ Years**

Deutscher Bundestag
Bundesnetzagentur
BASF We create chemistry
UBS
ABB
IBM
COMMERZBANK
DATEV
VakıfBank
MIGROSBANK
KB
Consorsbank
SIEMENS
Deutsche Telekom
AIRBUS
ING
SOCIETE GENERALE
Dropbox
DHL
Sparkasse
UniCredit
AXA
EBA EUROPEAN BANKING AUTHORITY
SAP

Engineered in Germany

KOBIL

Thank You!



Engineered in Germany