Cyber & AI à l'horizon 2030 : la bataille pour la vitesse

Guy-Philippe GOLDSTEIN

Enseignant à l’Ecole de Guerre Economique

Conseiller PwC Advisory - Cyber & AI impacts on Business Management

A.
“Winter is coming”

GP Goldstein

The Quantum Imperative Timeline

- **Standards finalized:** NIST published three core PQC standards (FIPS 203, 204, 205) in August 2024, with two additional standards (FIPS 206, HQC) expected by 2027
- **US Federal deadline**: **2035** for complete migration, with interim milestones at 2027 (National Security Systems acquisitions) and 2030 (algorithm deprecation)
- **EU timeline:** Three-phase approach with milestones at 2026 (initiate transition), 2030 (high-risk systems), and **2035** (full migration)
- **Industry compliance:** Multiple regulatory frameworks (DORA, NIS2, PCI DSS 4.0) mandate cryptographic agility and PQC planning, with enforcement beginning in 2025

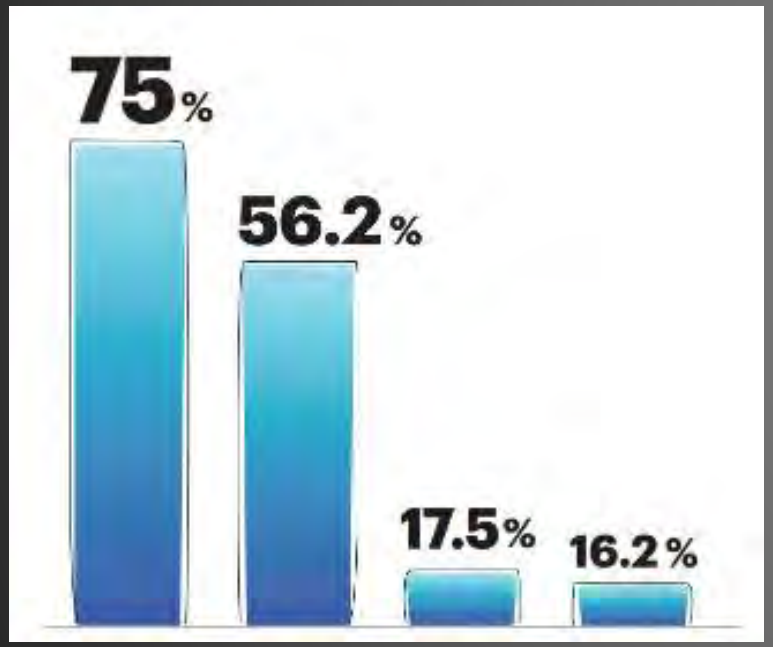
▶ ***Known path (change cryptography) across the industrial ecosystem against an emerging risk***

GP Goldstein

Une ingénierie sociale “amplifiée”

Une explosion du volume et de la qualité des attaques de phishing

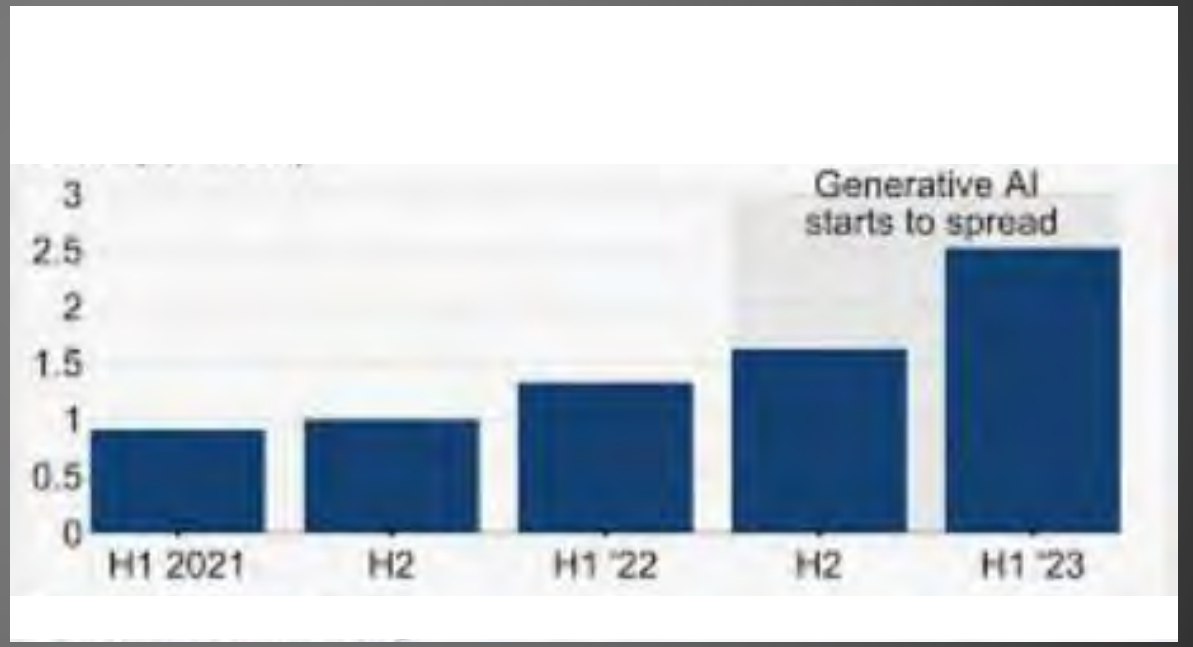
Quelles types d’attaques utilisant l’IA constitue la plus grande menace pour votre organisation?



Source: Team 8

Augmentation des emails frauduleux avec le développement de l’IA generative

(Nbre d’emails pro frauduleux reçu pour 1000 boites emails/ semaine)



Source: Abnormal security

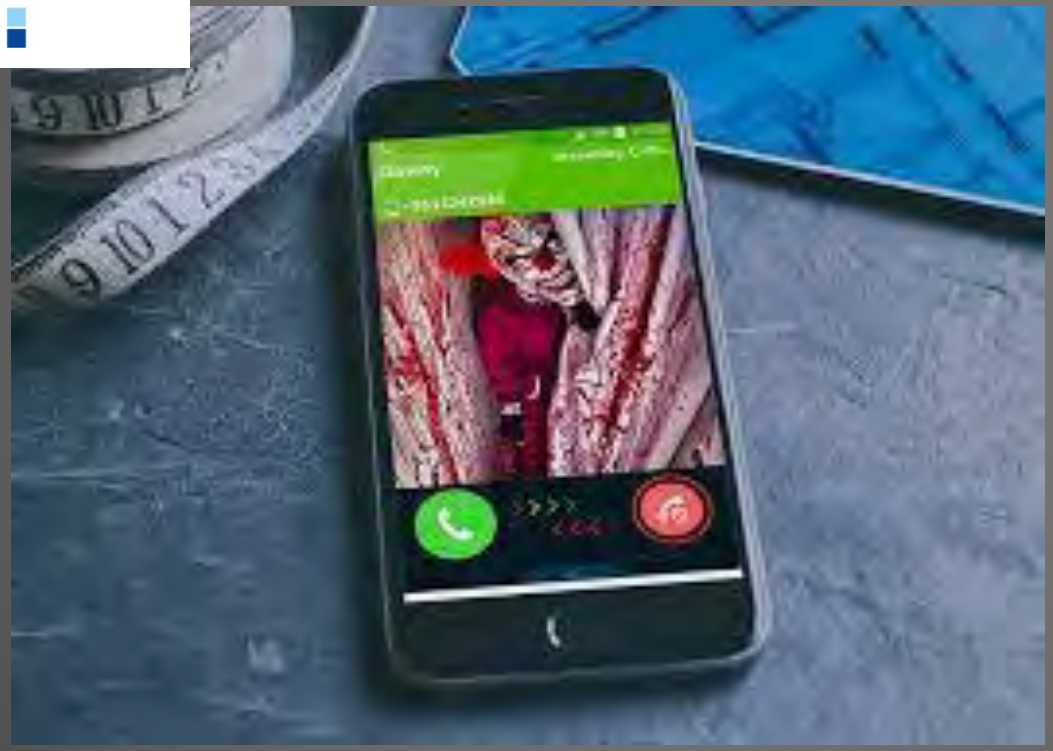
▶***Spearphishing: +4000% depuis novembre 2022 (****Source SlashNext)*

GP Goldstein

Les menaces de Demain

Deepfake & voice-cloning: vers la petite criminalité

Charente Libre



- J., habitant d'Angers, reçoit fin septembre 2024 un appel masque sur son portable de sa mère Martine
- "Je suis bloquée à la gare, j'ai perdu ma carte bleue"
- La voix est entrecoupée, l'intonation change – Martine est en fait en Dordogne

▶ Pour CyberMalveillance (Gouv), la voix a été générée par une IA

GP Goldstein

Une ingénierie sociale “amplifiée”
Deepfake & voice-cloning au niveau entreprise

Fraude “au directeur” à l’âge IA

Deepfakes en visio-conférence

Janvier 2024 –Visio conference faussé dans la filiale hong-kongaise de la société d’engineerie Arup ; vol de $25 millions

Source: SCMP, 2024

Voice cloning & visio

Mai 2024 – Voice cloning de Mark Read, ,PDG de WPP, avec faux compte Whatsapp et après connection sur Teams

Source: The Guardian, 2024

GP Goldstein

Une ingénierie sociale “amplifiée”
Les attentes réputationnelles (et financières)

Chantages sur les cadres seniors

Honeypot en Asie

Octobre 2024 – Gang cybercriminel aurait récupéré $46 millions; arrêté par la police de Hong Kong
Cible des professionnels masculins de Taiwan, Singapoure ou de l’Inde

Source: CNN, 2024

Manipulation l’image des dirigeants

Deepfakes avec message video

Avril 2024 – Faux messages des directeurs de la BSE et du NSE (en Inde) afin de manipuler les cours de bourse

Source: FT & Reuters, 2024

GP Goldstein

Le château fort & le réseau

Domaine médiéval

Défenseur >> Assaillant

- **Simple**
- **Le défenseur VOIT l'assailant**
- **Pénalités forte par assaut**

GP Goldstein

Le château fort & le réseau

Domaine médiéval

Défenseur >> Assaillant

- **Simple**
- **Le défenseur VOIT l'assailant**
- **Pénalités fortes par assaut**

VS.

Réseau d'entreprise

Défenseur ~ Assaillant

- **Complexe**
- **Le défenseur NE VOIT PAS l'assailant**
- **Peu de pénalités par assaut**

GP Goldstein

L'entreprise dans son réseau étendu de filiales et de prestataires

Des réseeaux toujours plus complexes et etendues …

- 150 à 400 filiales
- 2.000 à 20.000 fournisseurs

… des assaillants qui ont l'ascendant sur le tempo

- Mean Time To Detection: 10 jours pour notification en interne (2025)

VS

- Vitesse Exfiltration des données: 2 jours de la compromission à l'exploitation

▶ Une course de Vitesse entre Attaquant et Defenseur
▶ Comment accélérer la défense ? Et ralentir l'attaquant?

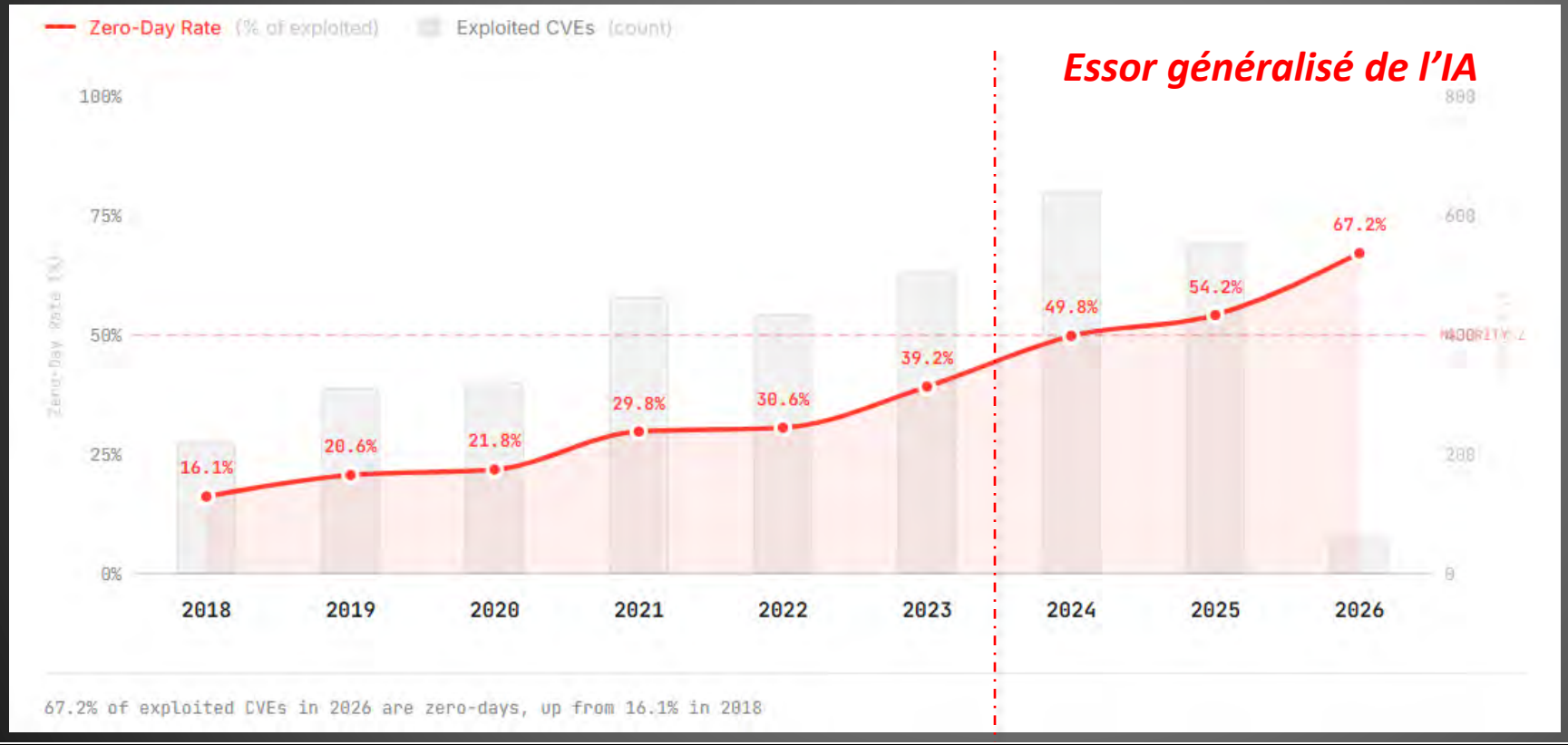
Source: Google, Palo Alto Networks

GP Goldstein

L'IA contribue à l'acceleration du tempo

Vitesse d'utilisation des Zero-Day

% des CVEs où l'exploitation a lieu avant ou le jour même de la divulgation publique



- Montée progressive de la Vitesse d'exploitation depuis 2018
- Entre 2023 et *début* 2026, le taux est passé de 39% à 67%
- Le déploiement des systèmes IA, en particulier agentique, va accélérer le rythme
  - Découvertes: DARPA AICC, Anthropic avec Firefox (3/2026)
  - Exploitation: CVE Genie

▶ ***Début 2026, la durée Moyenne d'exploitation d'un Zero Day est passé à 1,6 jours contre 56 jours en 2024***

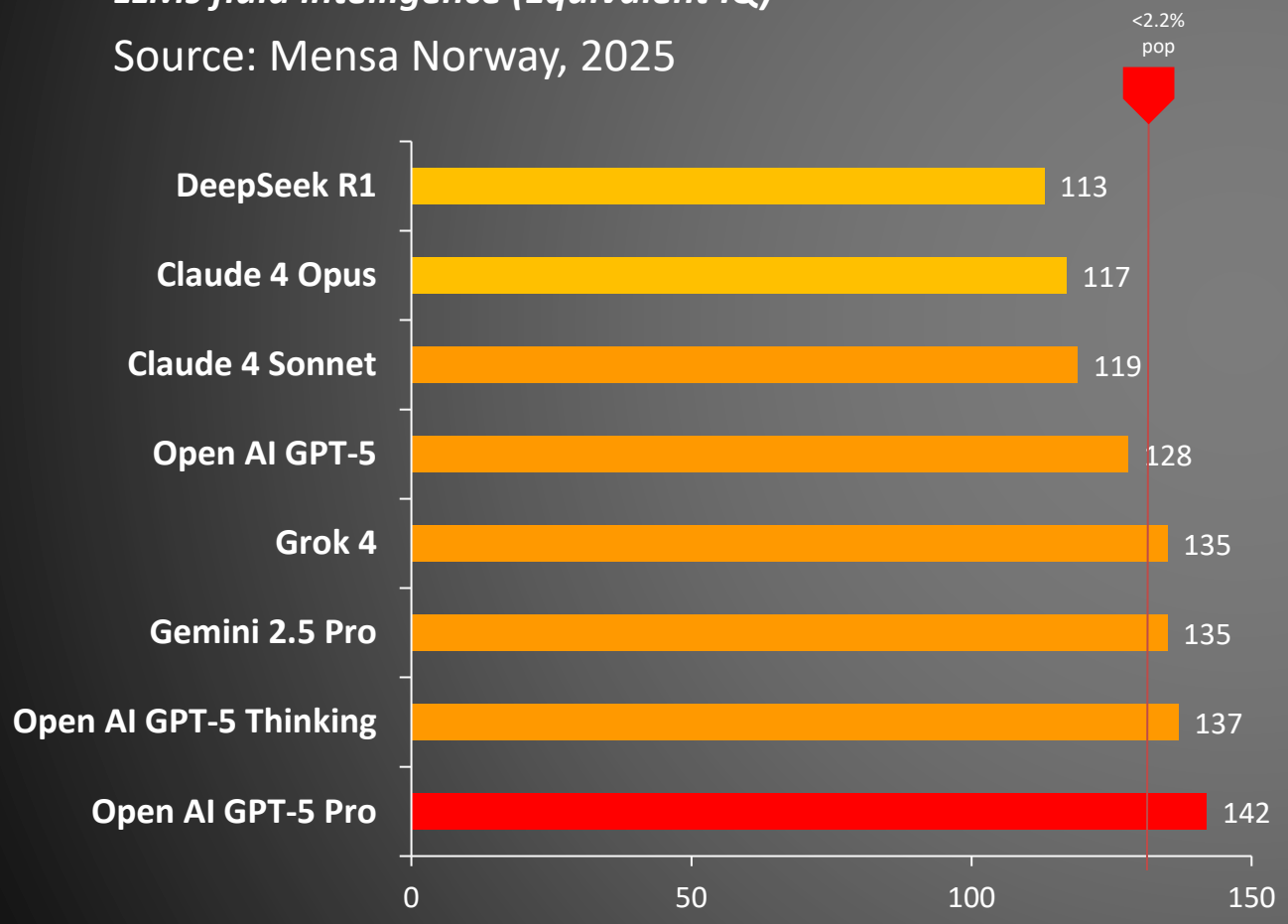
Source: Zero Clock Day

GP Goldstein

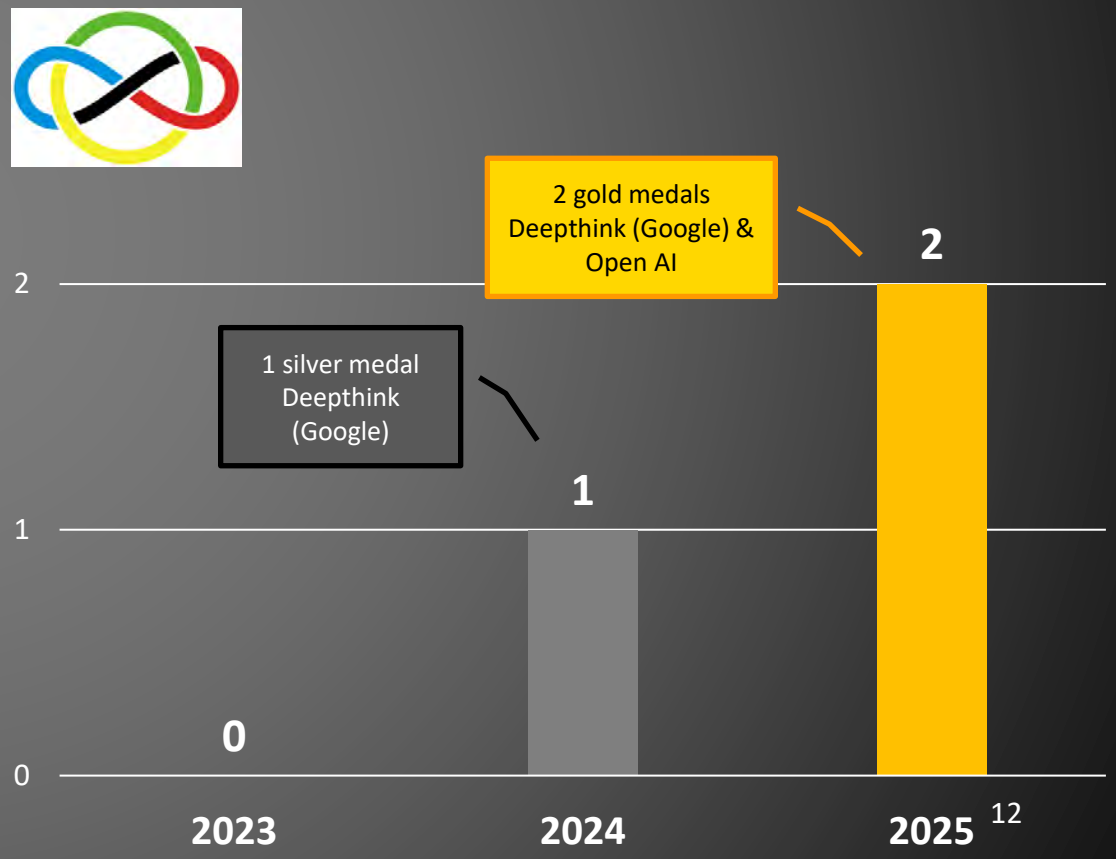
Meanwhile: some AI systems had already surpassed the cognitive abilities of 98% of the human population in many specific tasks

LLMs fluid intelligence (Equivalent IQ)

Source: Mensa Norway, 2025



Medals obtained by LLM at International Mathematics Olympiads



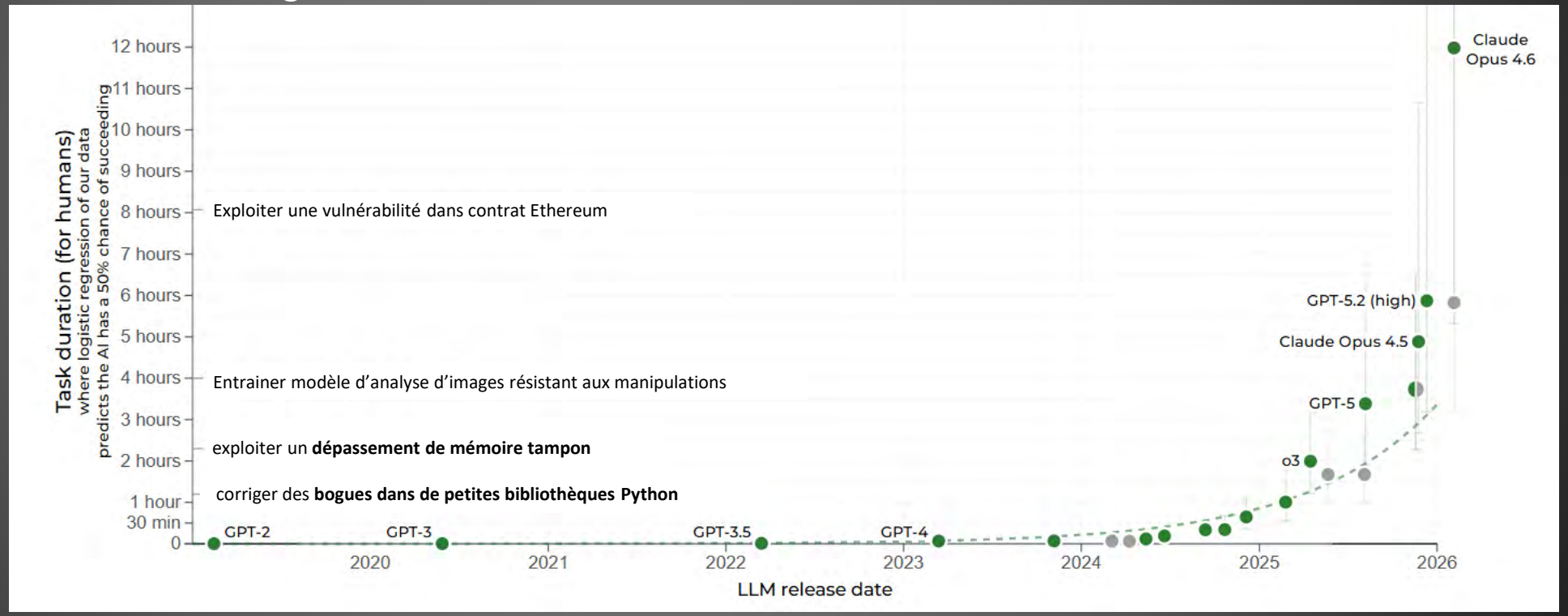
Source: Nature, New York Times

GP Goldstein

Les capacités des LLMs à gérer des tâches complexes de longue durée sont encore limitées mais pourrait doubler tous les 7 mois

Durée de la tâche (mesurée en temps necessaire à des professionnels humains), réalisée avec un taux de succès de 50% ou plus

Source: METR.org



Date de publication du modèle

▶En 2026, certains modèles d'IA Générative pourront effectuer l'équivalent de tâches humaines de plusieurs heures

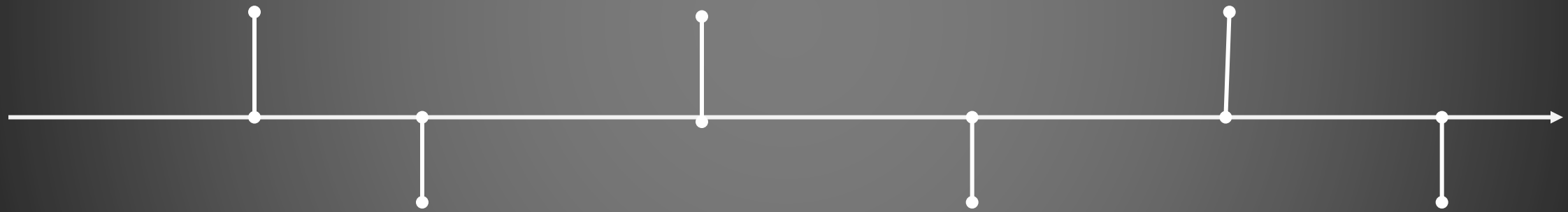
GP Goldstein

Les signes se multiplient d'une montée du hacking augmenté / Hacking IA Autonome

MITRE: Résultats OCCULT
(février 2025)
- DeepSeek R1 score de 92% sur benchmark (GPT4o proche)
- Llama 3 excelle à mouvements latéraux & approaches "living off the land"

HackerOne Leaderboard
(juin 2025)
- XBOW (syteme AI autonome) arrive Numéro 1 du classement

CVE-Genie
(Sep2025)
- Equipe de chercheurs de Boston U. & UC Santa Barbara crée un LLM qui sur la base de CVE crée un exploit verifiable avec taux de success de 51%



Palisades Research
(Mai 2025)
- Systèmes AI autonomes arrivent dans le Top 10% - Top 5% de competitions CTF

DARPA AI Cyber Challenge
(aout 2025)
- Decouvertes par l'ensemble des équipes de 54 vulnérabilités en 4 heures

Sean Heelan (Ind. Res.)
(Jan 2026)
- **"Industrialisation de la génération d'exploits" :** des agents créant 40 exploits distincts sur 6 scénarios, la plupart de ces scénarios étant résolus par GPT5.2 et Opus 4.5.

GP Goldstein

Experts declarations evolving every other month

Janvier 2025

"growing but currently limited risks."

Octobre 2025

"The UK National Cyber Security Centre predicts that by 2027, general-purpose AI systems will almost certainly (95–100% confidence) make cyber offence more effective and efficient"

Octobre 2025

- Heather Adkins, VP, Security Eng., Google
- Bruce Schneier, Harvard Kennedy School
- Gadi Evron, Fmr Chairman IL-CERT

In "Autonomous AI Hacking and the Future of Cybersecurity", CSO, 10/2025

« Tipping point on the horizon
AI agents now rival and sometimes ***surpass even elite human hackers*** *in sophistication. . […] In this future, attack capabilities could accelerate* ***beyond our individual and collective capability to handle.»***

▶ ***Unknown path (what to do?) across the industrial ecosystem against a new emerging risk***

B.
Horizon 2030:
Nouveau champ de bataille

GP Goldstein

Une ingénierie sociale amplifiée

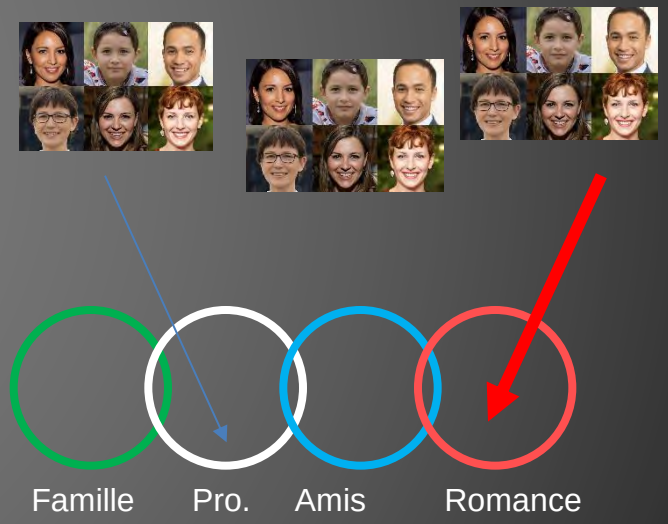
Pas de confiance implicites sur les reseaux en ligne / AR / VR…

Environment

Cf. www. thisPersonDoesNotExist.Com

Menaces

Manipulation cognitive avancées



Macro or Micro Targeting
Multiple effets psycho

GP Goldstein

…Dans un future très proche?

Avatars coordonnées à travers différents types de réseaux

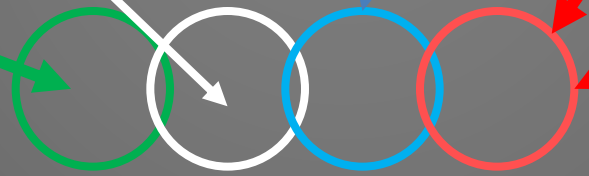
Faux collègue sur
<u>Réseaux corporate</u>
- Infostealer
- Spearphishing
- Recrutement

Reconnaissance
Recrutement

Black-mail

Cyber-harassment
sur enfants de cible

Seduction



Modèle psycho social avancé

Jumeau numérique de l'orga

Multi-Micro Targeting – multiples cibles « à l'échelle » à travers toute l'organisation

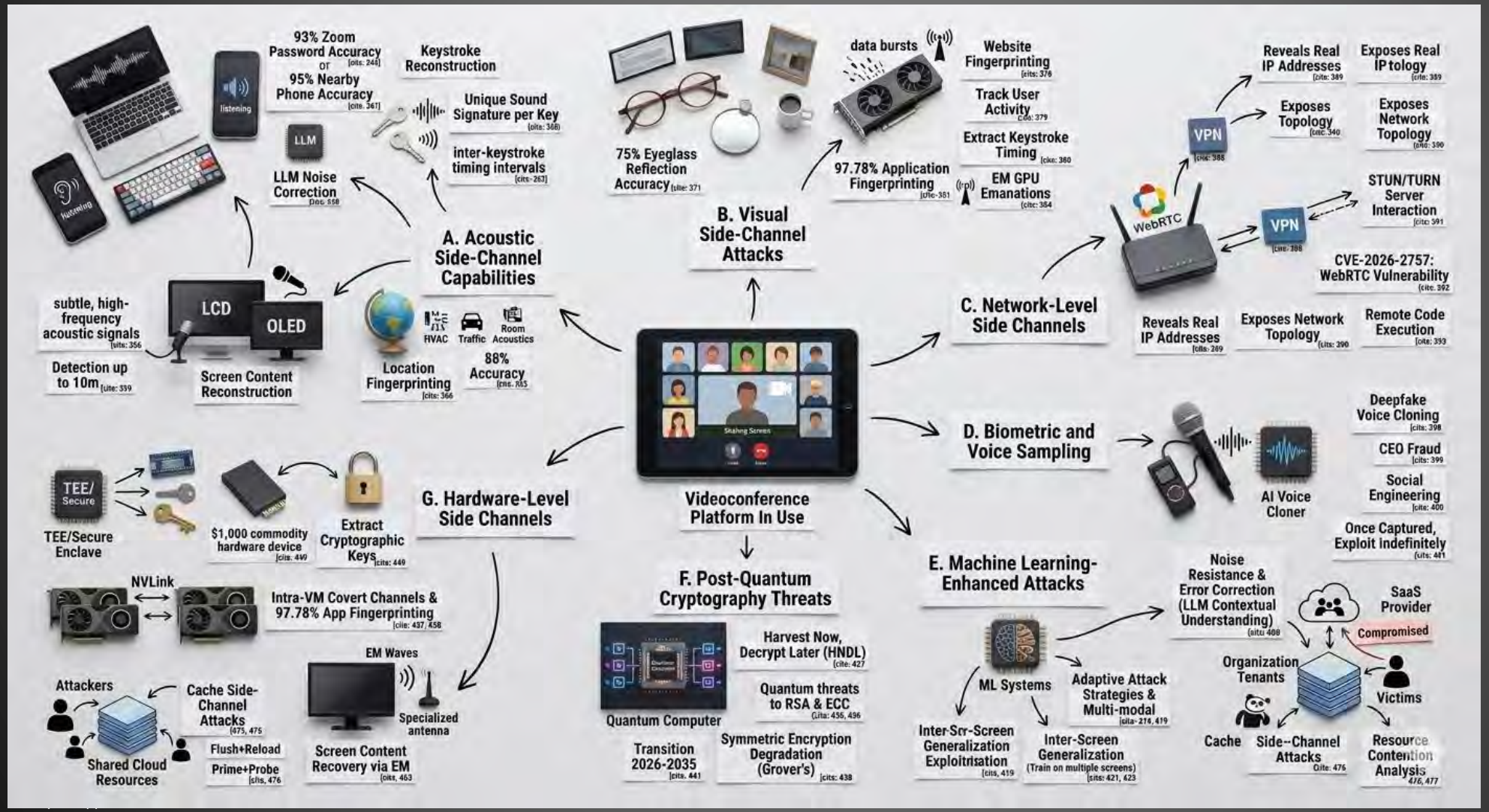
Capacité à comprendre l'esprit humain
- Traits de personalités Big 5
- "Theory of Mind" adulte des GPT 4.0 (Michal Kosinski / Stanford - 2023)

Capacité à persuader:
- Supérieur aux humains: 82% plus de chances de remporter un débat qu'un humain (EPFL, Fondazione Bruno Kessler - 2024)

➢ ***Risque emergent, y compris retournement pour insider threat***

GP Goldstein

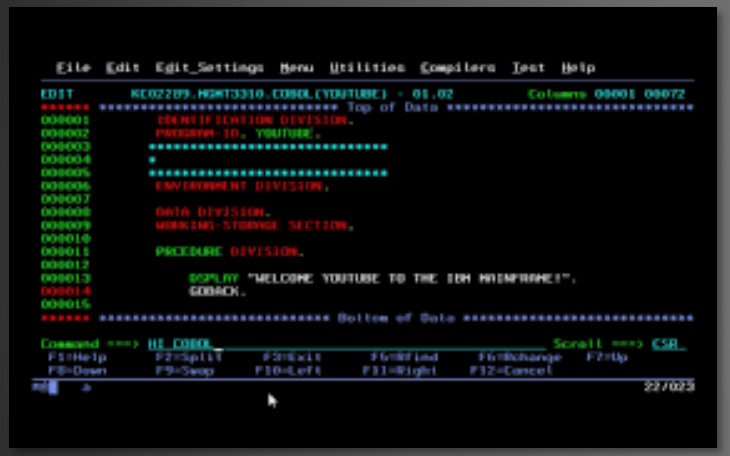
Explosion of side-channels attacks



GP Goldstein

Nouvelles menaces
Zero protection implicite – ni air gap, ni vieux code

“Vieux” COBOL protégé ?

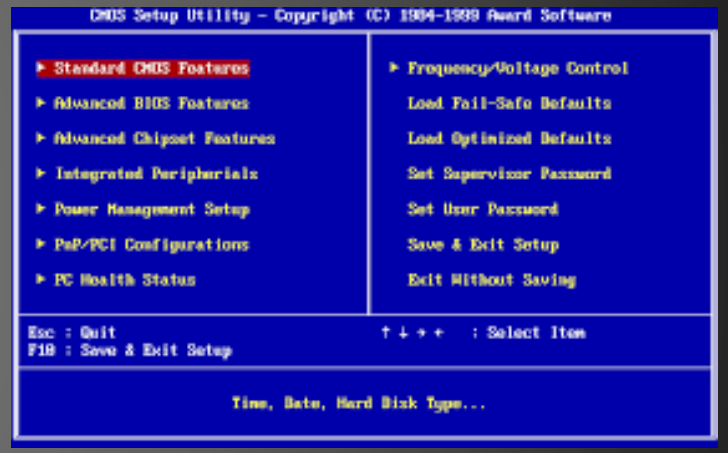


Ex: Cobol analysé par Claude (02/2026)

OT trop particuliers?

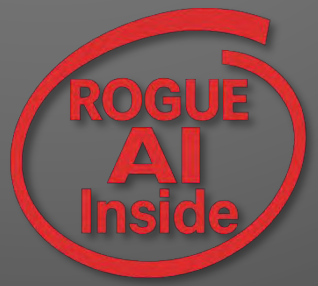
Ex: plusieurs projets de recherche (ICCPulse, Norvège; NVIDIA Morpheus..)

BIOS Firmware?



Ex: plusieurs projets dont la startup BootLoop (Y Combinator)

Vulnerables à



GP Goldstein

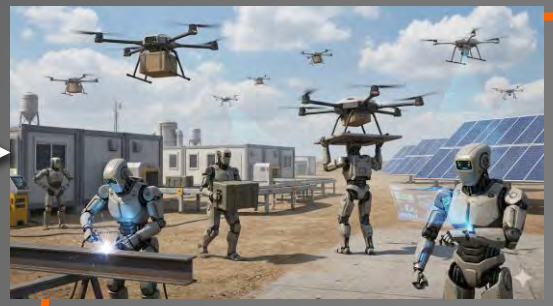
The Great Convergence: "Tsunatech"

Side channel attacks

Phys. Interception

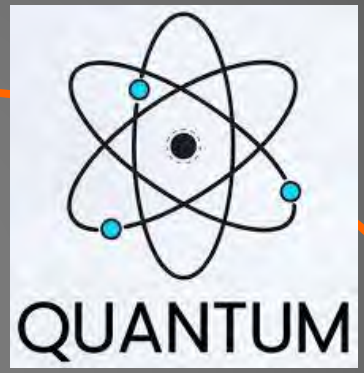
Cyber-Phys effects

Robotics

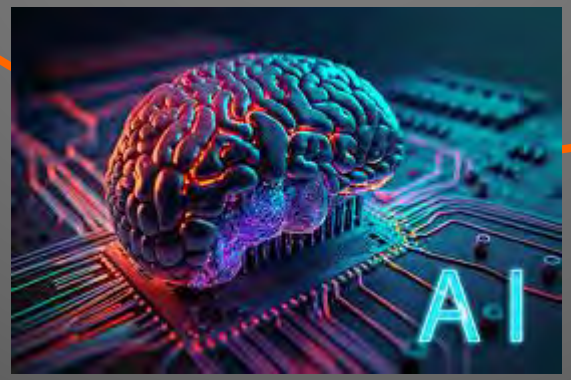


New cyber-physical effects

Quantum simulation



New correlation & causality models



Orchestration

GP Goldstein

The Great Convergence: “Tsunatech” with new people

Democratization of expertise from AI:
New recruits

EXTERNAL

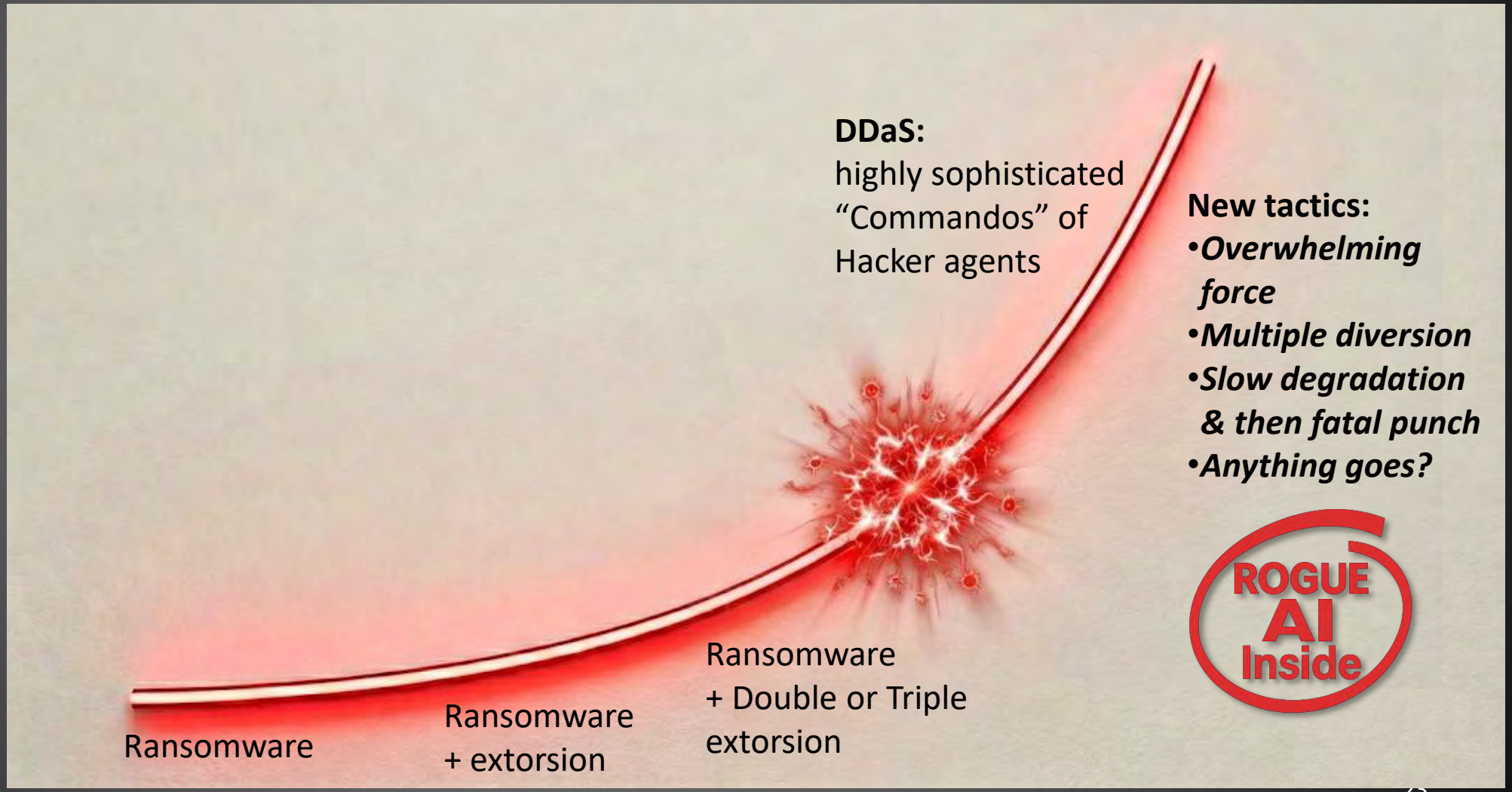
- Old crime transitioning
- Democratization of hacking – Vibe Hacking

INTERNAL

- Insider threat
- Insiders recruited by coercion

GP Goldstein

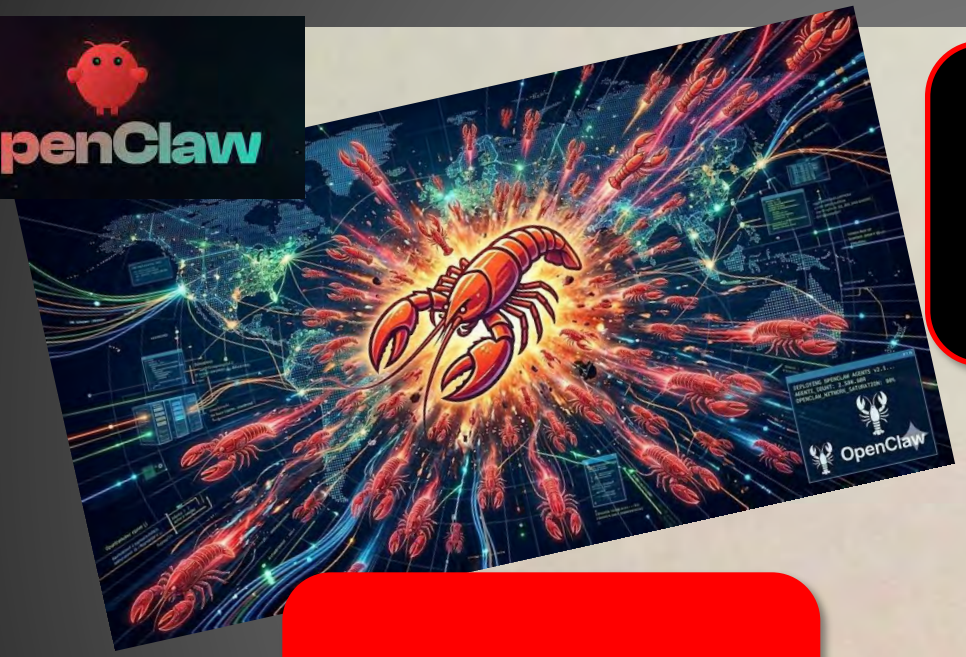
DDaS - Distributed Deployment of Agent Swarms with "precambian explosion" of tactics & endgoals



GP Goldstein

DDaS - Distributed Deployment of Agent Swarms with “precambian explosion” of tactics & endgoals

OpenClaw



Claude-Flow's BatchTool: jusqu’à 100 agents / 60 types d’agents spécialisés

3-5 agents / Jusqu’à 10 agents (02/26)

DDaS:
highly sophisticated “Commandos” of Hacker agents

New tactics:
- ***Overwhelming force***
- ***Multiple diversion***
- ***Slow degradation & then fatal punch***
- ***Anything goes?***

ROGUE AI Inside

Ransomware

Ransomware + extorsion

Ransomware + Double or Triple extorsion

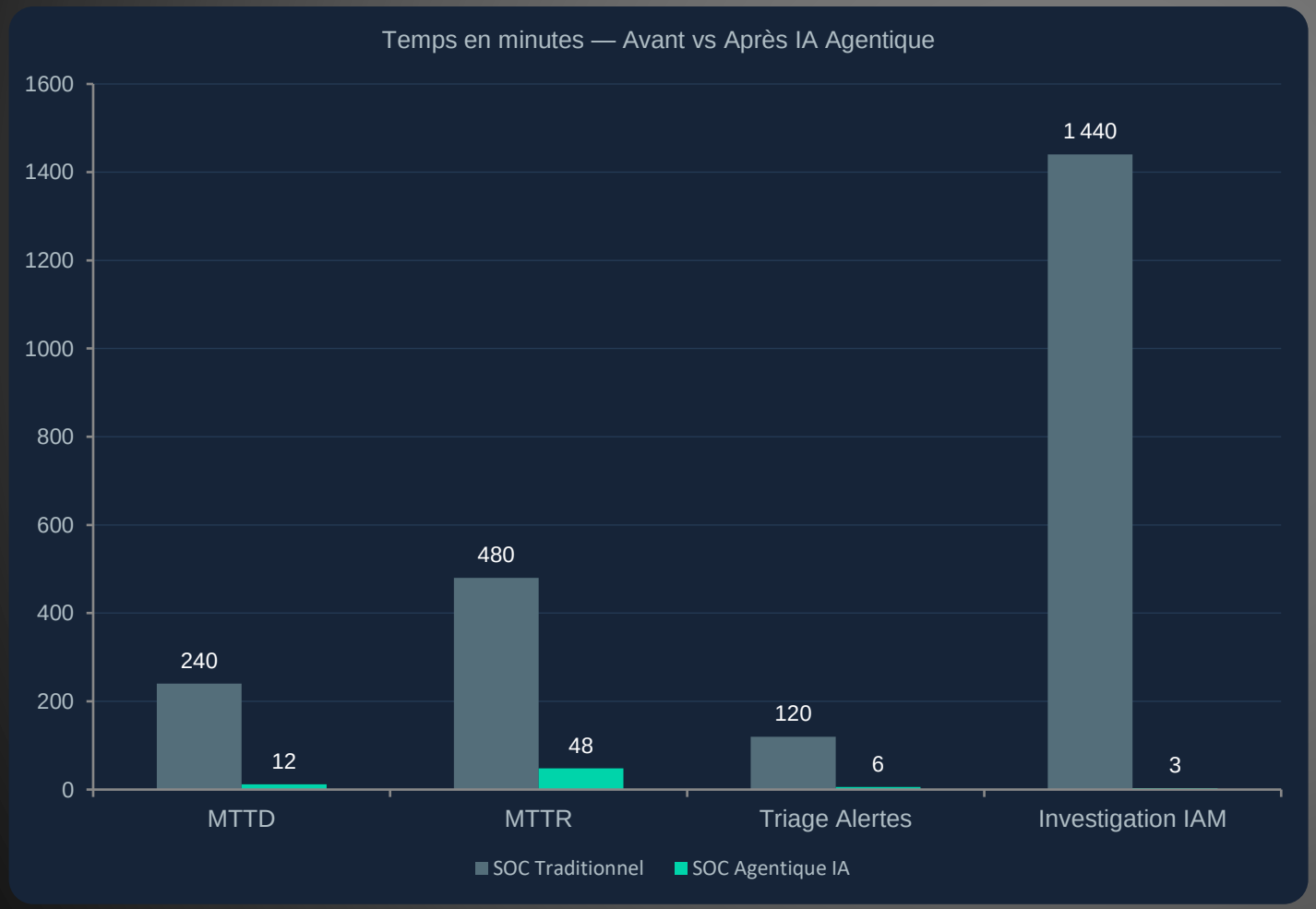
C.
Horizon 2030:
La nouvelle Défense

GP Goldstein

A. Develop AI Agentic cybersecurity

IMPACT OPÉRATIONNEL : DÉTECTION & RÉPONSE

Les agents IA réduisent drastiquement les temps de détection et de remédiation



20× — **MTTD plus rapide**
Stellar Cyber — clients en production

−60% — **Réduction MTTR**
Benchmark SOC agentiques 2025-26

3 min — **vs 1 jour (IAM)**
Services financiers — ReliaQuest

Sources : Stellar Cyber, Torq, ReliaQuest, Prophet Security, Omdia 2025-2026

B. Recode / Reinforce

Critical elements of IT corporate stack

Formal proof programming

GP Goldstein

C. Nettoyer le code automatiquement

Ex: GOOGLE BIG SLEEP : CHASSER LES ZERO-DAYS PAR IA

LE CONTEXTE : « VULNPOCALYPSE »

30 000 — **Bugs NVD non résolus**
Backlog croissant — [un]prompted 2026

$1 Md — **VC dans l'AI pen-testing**
Course à l'armement offensive

0 — **Faux positifs tolérés**
Contrainte de design non-négociable

ARCHITECTURE TRI-SYSTÈME

01 — BIG SLEEP
Découverte autonome de vulnérabilités zero-day. Étudie le codebase, analyse les bugs passés, variant analysis, confirme l'exploitabilité via debugger + Python.

02 — CODEMENDER
Remédiation et hardening automatique. 72 fixes upstream déjà livrés. Aurait pu rendre CVE-2023-4863 (libwebp zero-click iOS) inexploitable.

03 — SECURE-BY-DEFAULT LLM
Entraîner l'IA à écrire du code sécurisé dès la génération. Éliminer les vulnérabilités à la source, avant même le commit.

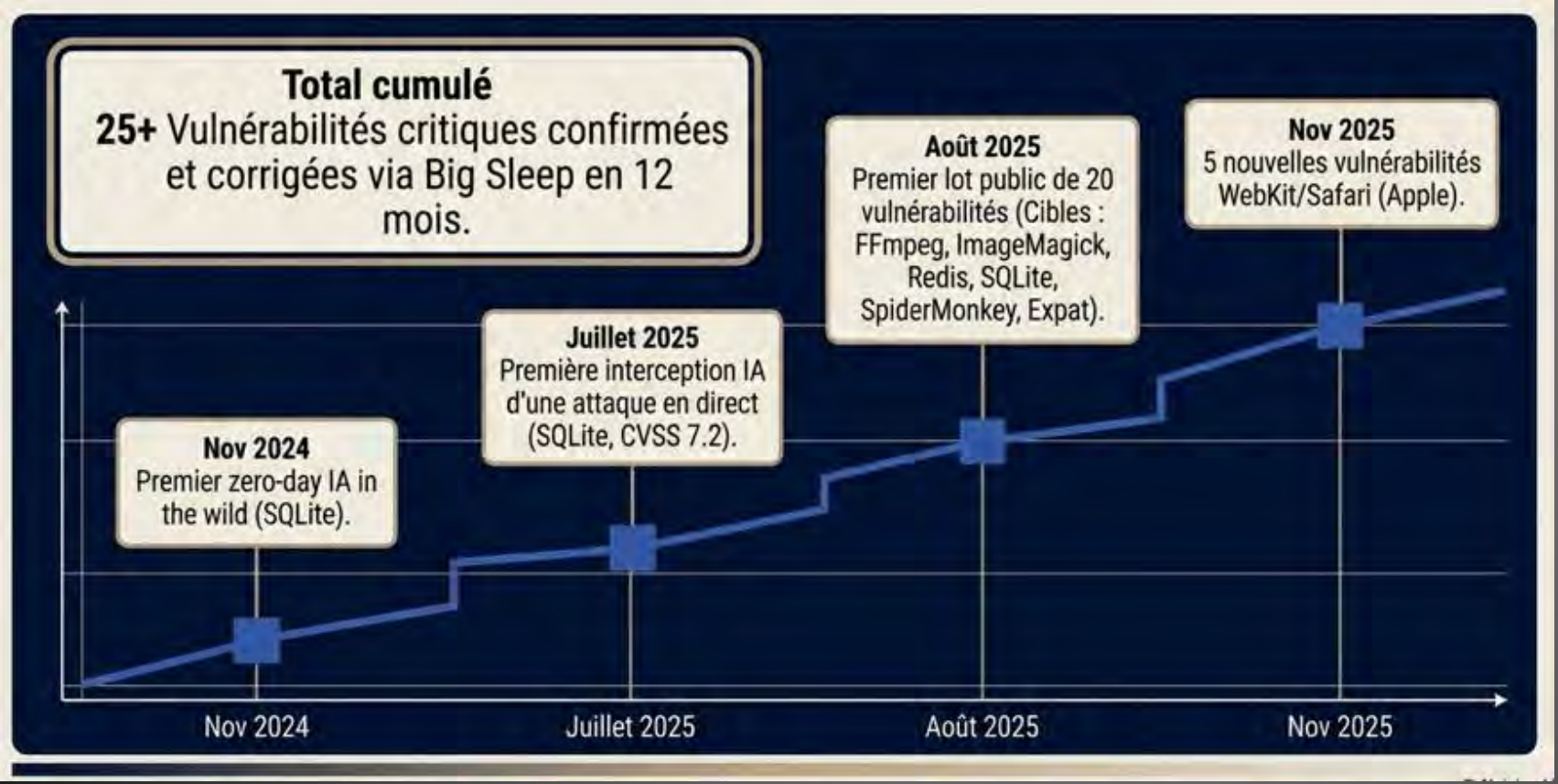
➢ **DeepMind × Project Zero — « How do we eliminate every software vulnerability on Earth? »**

Source : [un]prompted AI Security Conf., SF, 3-4 mars 2026 — H. Adkins (VP Security Google) & J. Flynn (VP Security DeepMind)

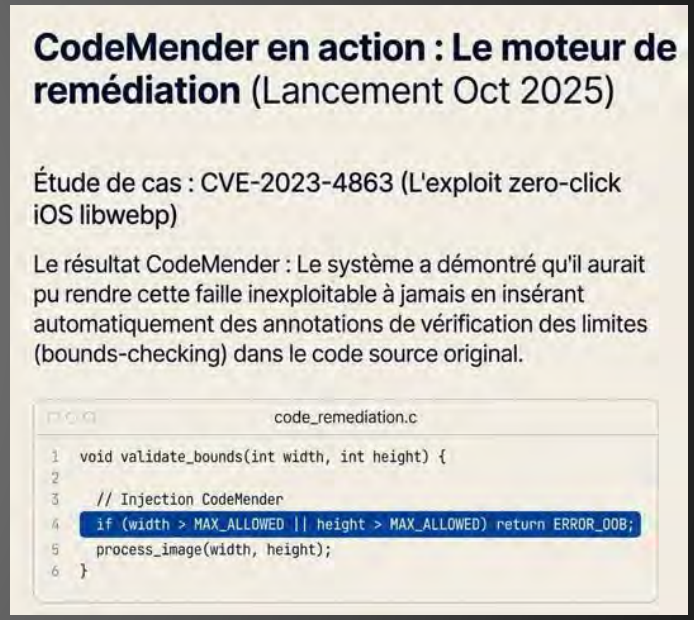
GP Goldstein

C. Nettoyer le code automatiquement

BIG SLEEP



CODE MENDER



- **CodeMender** (lancé Oct 2025) : 72 fixes de sécurité upstream

Sources : Google DeepMind blog, Project Zero, The Hacker News, Apple Security Advisories, LinkedIn H. Adkins, Vulert

GP Goldstein

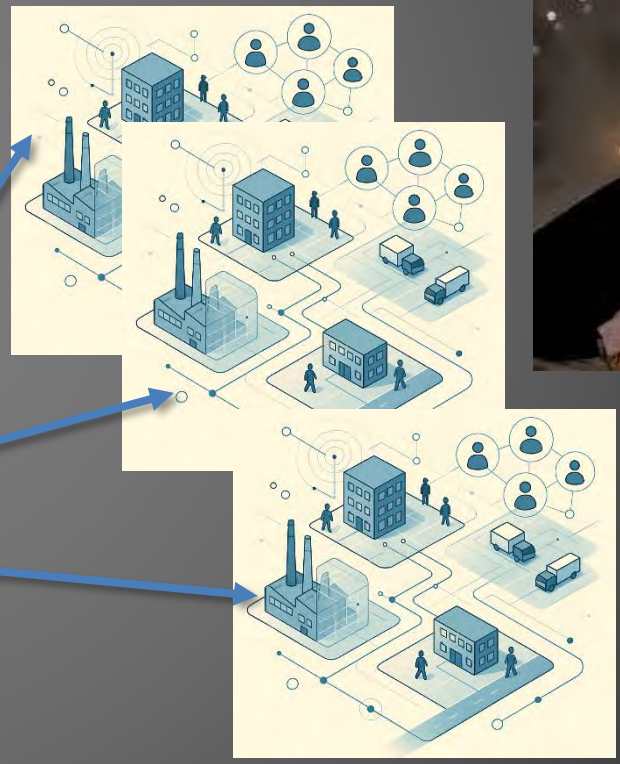
D. Build the Cyber "Nakatomi Plaza" (Die Hard)

GOOD AI INSIDE

Advanced Digital Twin, <u>including human layers</u>

Micro Segmentation

Deception and Traps



▶ *Slow down / Break the pace of the enemy – and make him pay*

GP Goldstein

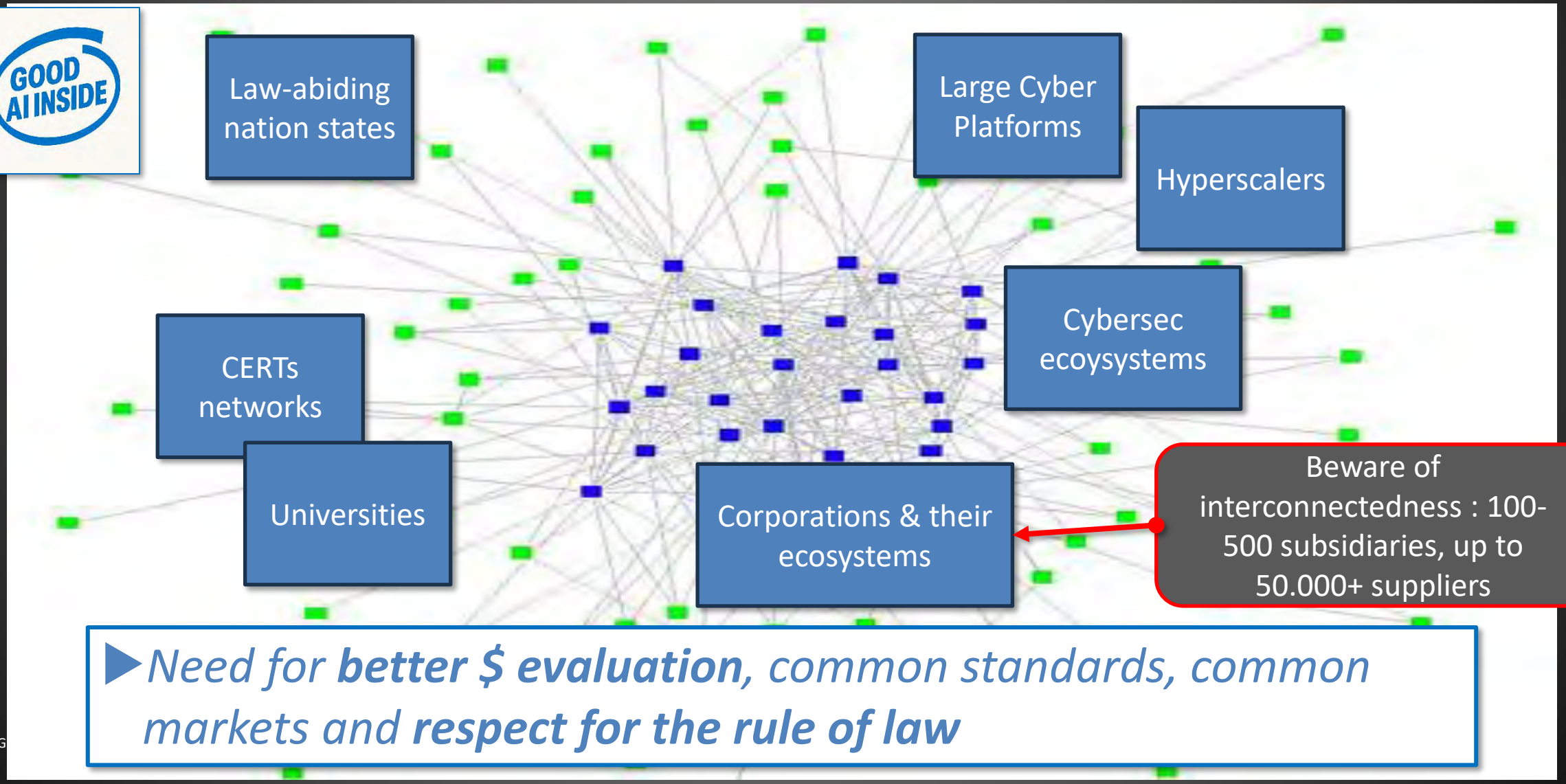
E. Ease the recruitment and expertise democratization of the whole corporation

VS.

GOOD AI INSIDE

GP Goldstein

F. Building the new AI-orchestrated Brightnet



▶ *Need for **better $ evaluation**, common standards, common markets and **respect for the rule of law***

D.
Horizon 2030:
Des mondes nouveaux

GP Goldstein

Win the Battle of Speed: fight AI with AI as

- Observe the tempo of the enemy (Threat watch / Revised cost of attack)
- Right defense condition at right tempo (SOTA tools? Current maturity? Update?)
- Spend on slowing down the attackers – with right tools & training
- Spend on accelerating the defenders – with right tools & training
- Right culture of the organization: « Action this day » !
- Governance & ROI for Speed

GP Goldstein

A la croisée de la maîtrise des enjeux d'IA et d'un environnement "Polycrisis": 4 scenarios pour l'avenir

Maîtrise Vitesse d'adaptation cyber

Maîtrise des chocs exogènes