Cybersécurité & Antifragilité

Beyond Resilience

Séminaire Aristote — Janvier 2026

Franck Rouxel — klaerenn.fr

L'impasse — La nemesis

L'IA rogue auto-réplicante

Une IA offensive qui explore, mute, s'adapte
plus vite que toute défense.

Elle ne cherche pas à reproduire le connu.
Elle crée l'inédit.

→ *Elle explore en permanence la terra incognita des possibles*

L'asymétrie mathématique

Côté défensif

Amélioration linéaire

L'IA augmente l'efficacité
des experts existants

→ *Amélioration proportionnelle*

Côté offensif

Croissance exponentielle

Capability uplift :
accès aux techniques sophistiquées

Throughput uplift :
échelle et vitesse

Google DeepMind — Framework for Evaluating Cyberattack Capabilities (2025)
12 000+ incidents réels, 20+ pays, 7 archétypes d'attaque
Phases les plus disruptées : reconnaissance, weaponization, evasion
→ *L'IA n'apporte pas de capacités « breakthrough » — vitesse et échelle suffisent*

Distinguishability Collapse — Résultats formels

Theorem 4.6 — Distinguishability Collapse

La détection devient computationnellement impossible
quand le coût du mimétisme est polynomial (condition réalisée par l'IA)

Theorem 4.13 — Irreversibility

L'effondrement est permanent — pas de technologie défensive symétrique

Theorem 4.16 — Kill Chain Composition

Amélioration IA : exponentielle en attaque, additive en défense

« Denning (1987) → Forrest (1996) → Wagner & Soto (2002) → Collapse (2026) »

Contraintes structurelles irréductibles

La défense

- Doit protéger chaque point d'entrée
- Contrainte de couverture totale
- Dépend du « patient 0 » — découvre toujours après

L'attaque

- N'a besoin que d'une faille
- Loi du maillon faible
- 0,6% des vulnérabilités causent 80% des dégâts

→ *L'équation devient mathématiquement déséquilibrée*

L'échec documenté

4,88 M$ — coût moyen d'une breach

82,6% — phishing assisté par IA

+500% — rançons en un an (400k$ → 2M$)

Conformité respectée. Compromission effective.

→ *29 000 CVE/an, 32% exploitées le jour même*

→ *L'approche risque s'effondre face à l'inédit*

Les fausses sorties

Niveau 1 — Mieux détecter

Semantic Indistinguishability Theorem (6.3)

À la couche sémantique, l'intention ne peut être inférée
des observables — information-théoriquement impossible.

Promise-Theoretic Impossibility (5.6)

La détection est une imposition qui échoue
quand l'adversaire ne promet pas de révéler son intention.

→ *Pas computationnellement difficile*
→ *Information-théoriquement impossible*

Niveau 2 — Apprendre les invariants

Surfaces, couplages, dépendances, contraintes

L'intuition

Apprendre la topologie plutôt que les patterns d'attaque.

Le piège

Une architecture qui s'auto-adapte au passé
optimise pour un futur qui ne se reproduira pas.

> → *On reste dans le paradigme défensif, un cran au-dessus*

Niveau 3 — La machine de Gödel

Schmidhuber (2003)

Un système qui réécrit son propre code
quand il peut prouver que la réécriture l'améliore.

Auto-amélioration récursive et optimale.

Le sommet de l'optimisation

→ *La machine améliore comment elle résout*
→ *Pas comment sa finalité survit aux perturbations*
→ *Sommet de l'optimisation = toujours l'impasse*

La théorie des systèmes non-linéaires suggère que la guerre moderne repose moins sur l'application linéaire de la force que sur la compréhension et l'influence des interactions dynamiques.

— Dominique Luzeaux, Revue Défense Nationale, 2025

Penser un système vivant

Edgar Morin plutôt que Taleb

Un système vivant n'est pas une machine à optimiser.

Il s'auto-organise.
Il fait émerger des propriétés que ses composants n'ont pas.

→ *La complexité n'est pas la complication*
→ *Le tout est plus que la somme des parties*

« La partie est dans le tout et le tout est dans la partie »

— Edgar Morin

CrowdStrike — La bonne question

Juillet 2024

8,5 millions de machines paralysées en 78 minutes.
Hôpitaux, aéroports, banques.

La mauvaise question

Comment éviter le bug CrowdStrike ?

La bonne question

Comment l'hôpital continue de soigner
malgré la perte de ses systèmes ?

L'antifragilité existe déjà — chez les attaquants

Ransomware-as-a-Service

L'arrestation d'un affilié renforce le réseau.

Ils ne protègent pas leurs composants.
Ils maintiennent des structures qui survivent
à n'importe quelle perturbation.

→ *Les marges d'inefficience financent la substituabilité*
→ *L'« inefficacité » apparente est une prime d'assurance*

Le pivot nécessaire

Detect and respond

↓

Constrain and verify

La détection a atteint ses limites fondamentales.
Les métadonnées, les signaux contextuels,
les contrôles architecturaux — hors du scope du collapse.

→ *Seule voie viable*

Le renversement

L'antifragilité s'applique à la finalité,

pas à la défense.

C'est une sortie du paradigme mécaniste
de la cybersécurité.

Pas son optimisation.

« Concevoir des systèmes qui se renforcent par l'attaque elle-même, qui utilisent l'agression comme signal d'amélioration »

Ce que je ne sais pas

Comment formaliser cette approche au-delà de la détection

Si les organisations peuvent sortir du paradigme
alors que la réglementation les y maintient

Comment accompagner ce changement de niveau conceptuel

Territoire largement inexploré

Merci

klaerenn.fr

DOI: 10.5281/zenodo.18151116