

# Ben je klaar voor de aanval?

**De impact van AI-doorbraken op cyberweerbaarheid en de concrete waarde van DORA TLPT- en TIBER-EU-tests**

Kunter Orpak

AFM

6 oktober 2026

# Voorstellen



## Kunter Orpak

*Senior toezichthouder en DORA TLPT en TIBER EU Team Lead bij Autoriteit Financiële Markten (AFM)*

*Promotieonderzoek aan de Universiteit van Amsterdam naar offensieve cybersecurity en de toepassing van GenAI*

# Agenda

**01** **Een verschuivend dreigingslandschap**
Frontier GenAI verandert aanvalscapaciteit en snelheid.

**02** **Kwetsbaarhedenbeheer onder tijdsdruk**
De tijd tussen ontdekking en misbruik wordt korter.

**03** **De ontwikkelsnelheid verandert de spelregels**
Nieuwe modellen en agents ontwikkelen zich in hoog tempo.

**04** **Vertrouwen moet aantoonbaar worden**
Governance en aantoonbare betrouwbaarheid zijn randvoorwaarden.

**05** **Hoe vertalen we dit naar weerbaarheid?**
Combineer geavanceerd testen met een sterke beveiligingscultuur.

# 01 Een verschuivend dreigingslandschap

# 1.1. Van hulpmiddel naar autonome actor

## 2-3 jaar geleden: GenAI als hulpmiddel

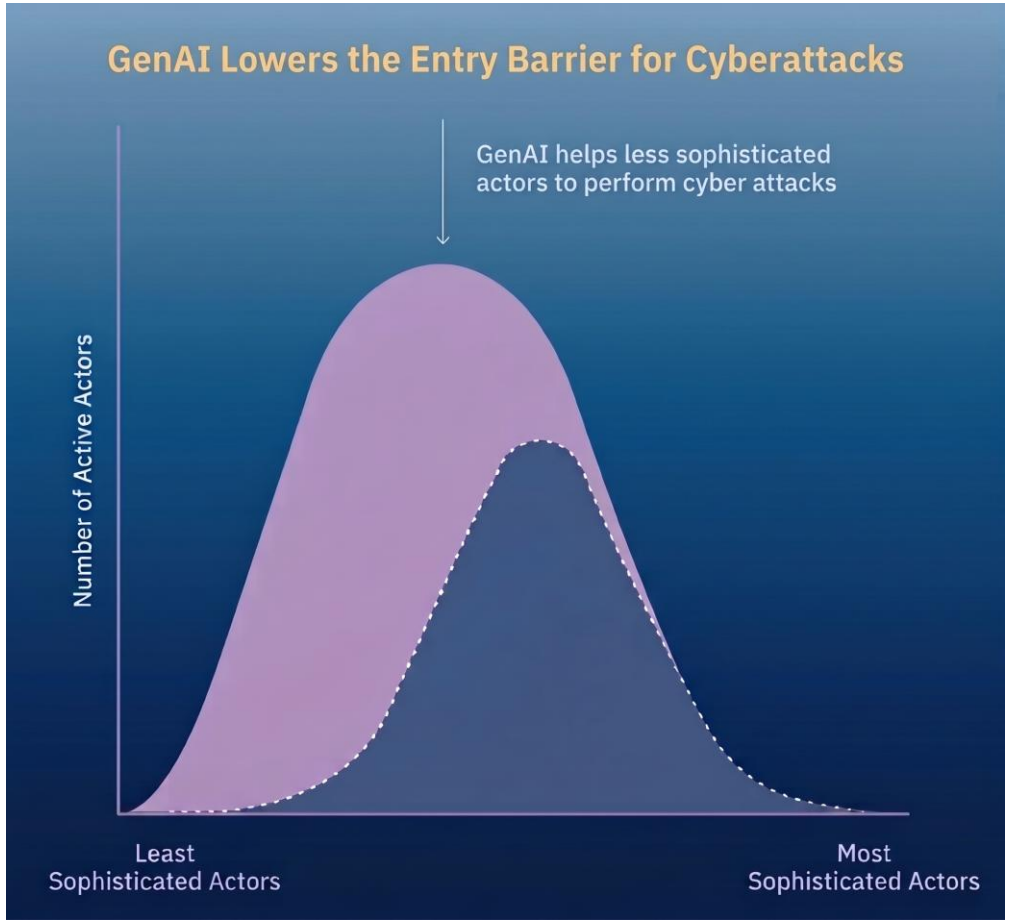
- Reageert op menselijke opdrachten
- Versnelt losse taken, zoals phishingteksten of code
- De gebruiker bepaalt doel en vervolgstappen
- Schaal en tempo blijven afhankelijk van menselijke inzet

## Nu: GenAI als autonome actor

- Plant en voert meerdere stappen uit
- Combineert informatie, tools en acties
- Past de aanpak aan op basis van resultaten
- Kan parallel werken en sneller opschalen

Nieuwe realiteit: Organisaties moeten de acties van GenAI kunnen beheersen, monitoren en valideren.

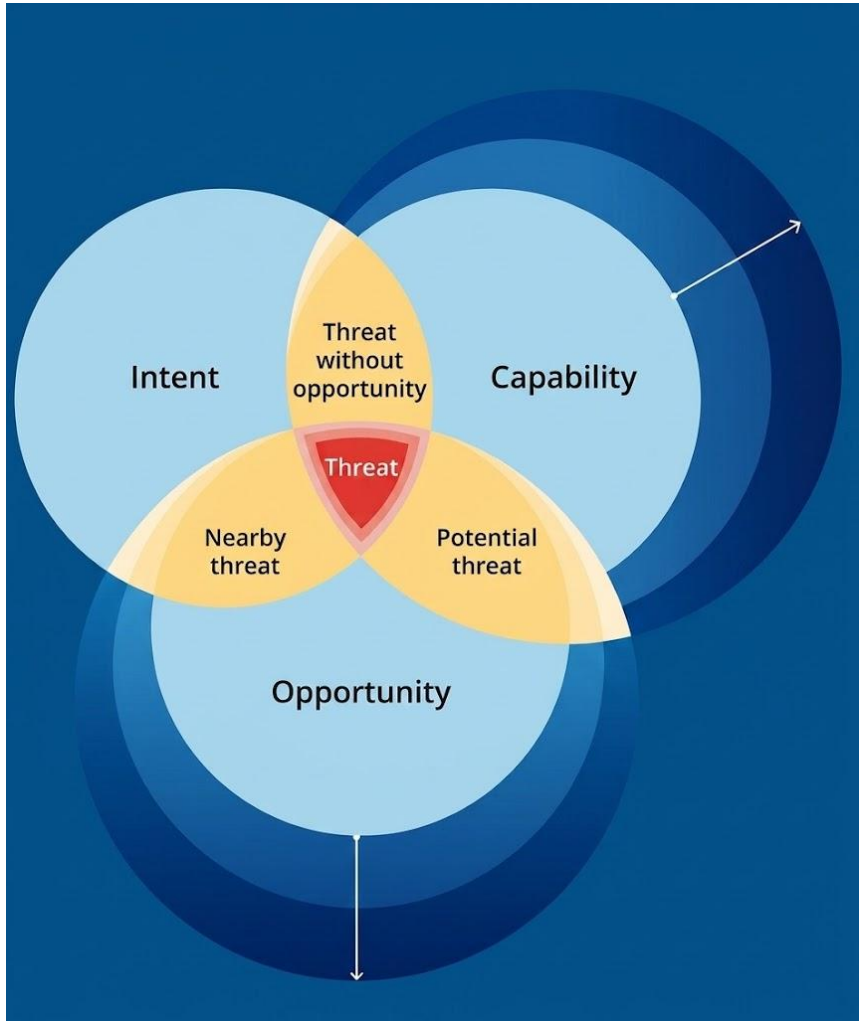
# 1.2. GenAI verlaagt de toetredingsdrempel




- Minder geavanceerde actoren automatiseren taken en schalen aanvallen sneller op.
- Geavanceerde actoren gebruiken GenAI selectief, met nadruk op discretie en operationele veiligheid.
- Het belangrijkste effect is een sterke toename van het aantal actoren met betekenisvolle aanvalscapaciteiten

Nieuwe realiteit: Intentie en toegang blijven belangrijk, maar capaciteit is minder schaars.

# 1.3. Dreiging ontstaat waar 3 factoren samenkomen




## Toename in capaciteit

- Malware en phishing sneller ontwikkelen
- Kwetsbaarheden en aanvalspaden sneller vinden
- Detectie beter ontwijken

## Toename in gelegenheid

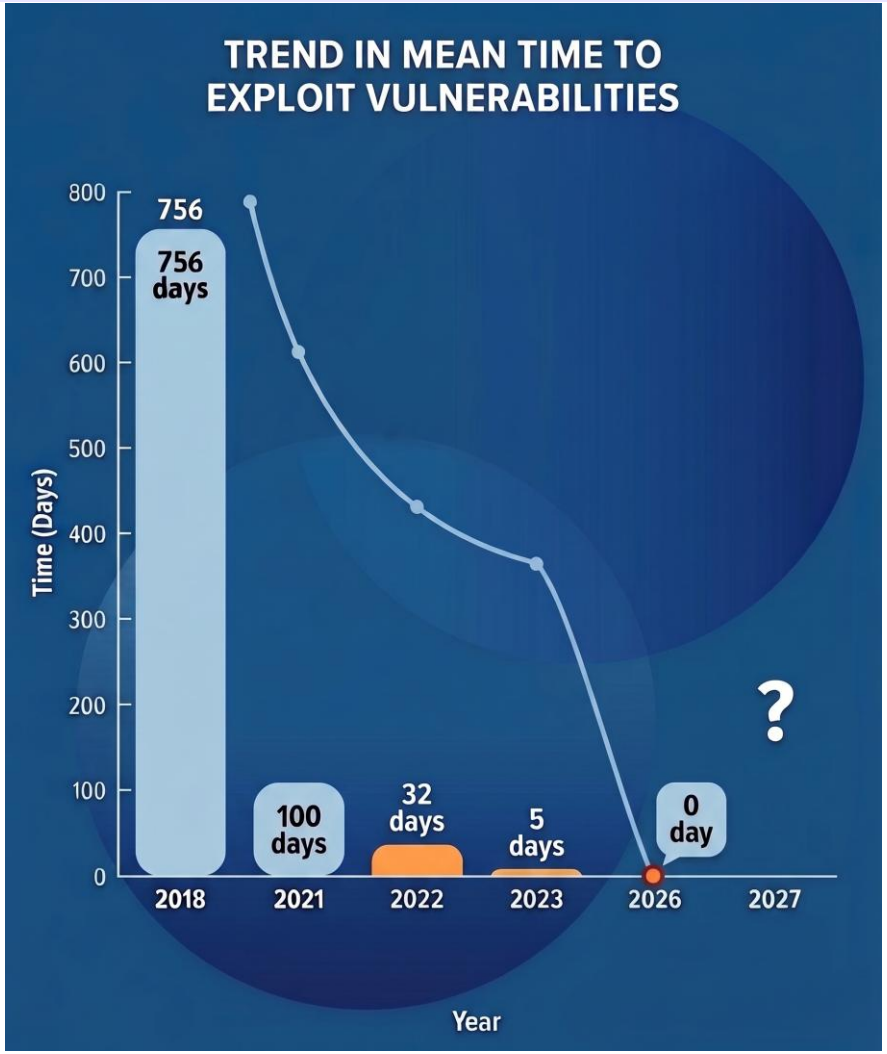
- Misconfiguraties geautomatiseerd vinden
- Onveilige GenAI-toepassingen misbruiken
- Legacy en end-of-life systemen exploiteren

Nieuwe realiteit: Prioriteer niet alleen kwetsbaarheden, maar vooral de combinatie van intentie, capaciteit en gelegenheid

02

# Kwetsbaarhedenbeheer onder tijdsdruk

# 2.1. Van periodiek patchen naar continue risicosturing




## Nu

- Time-to-exploit daalt al jaren
- GenAI versnelt reconnaissance en exploitontwikkeling
- Technologische ontwikkelingen verlopen snel en GenAI versnelt dit verder.

## Volgende stap

- Langetermijnvoorspellingen blijven moeilijk.
- Gecontroleerde experimenten met GenAI zijn noodzakelijk

Nieuwe realiteit: AI versnelt niet alleen aanvallen, maar verkort ook de beschikbare reactietijd. Traditioneel kwetsbaarhedenbeheer is op zichzelf niet langer voldoende.

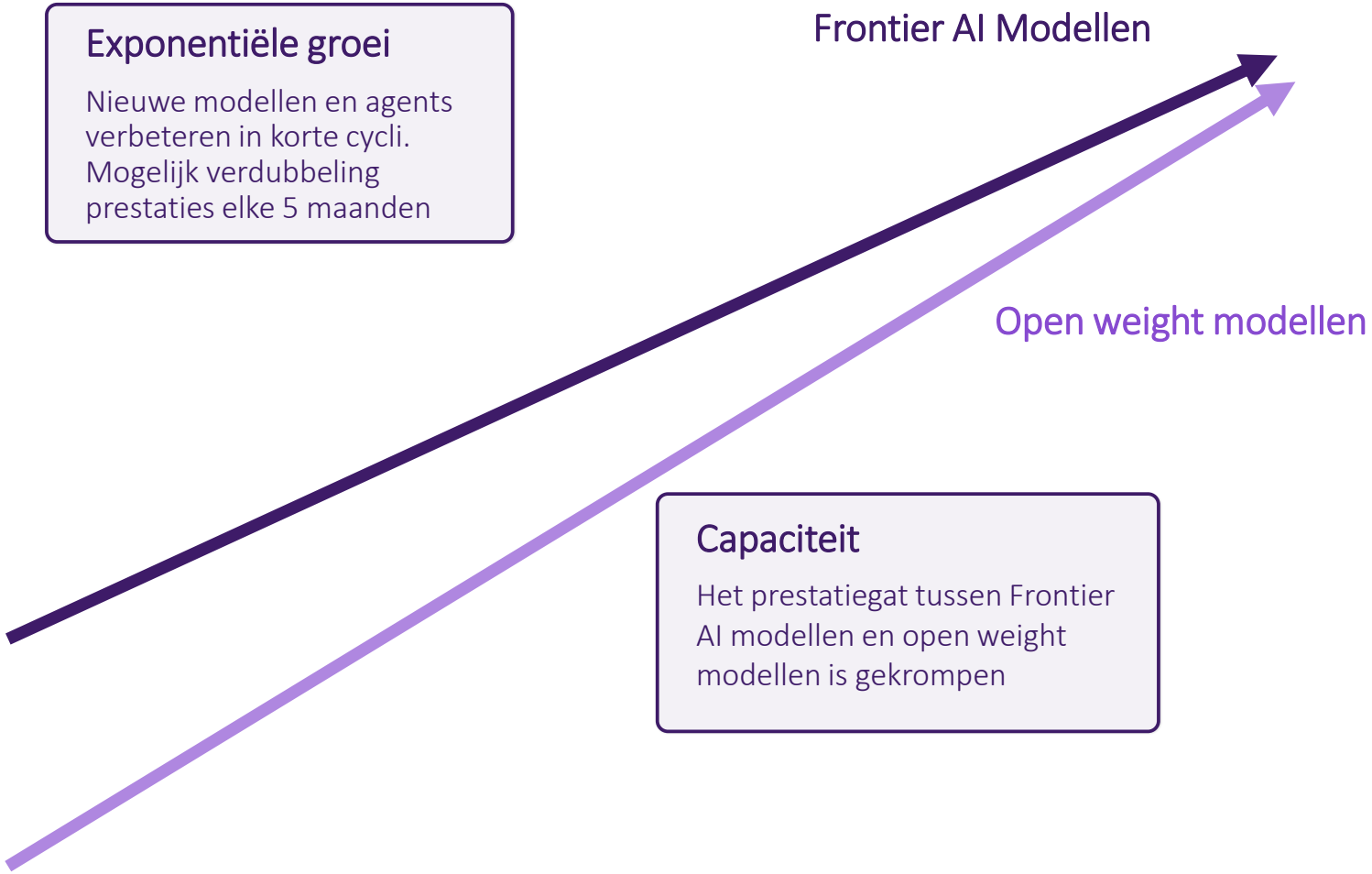
03

# De ontwikkelsnelheid verandert de spelregels

# 3.1. GenAI-capaciteiten ontwikkelen snel

### Exponentiële groei

Nieuwe modellen en agents verbeteren in korte cycli. Mogelijk verdubbeling prestaties elke 5 maanden




### Capaciteit

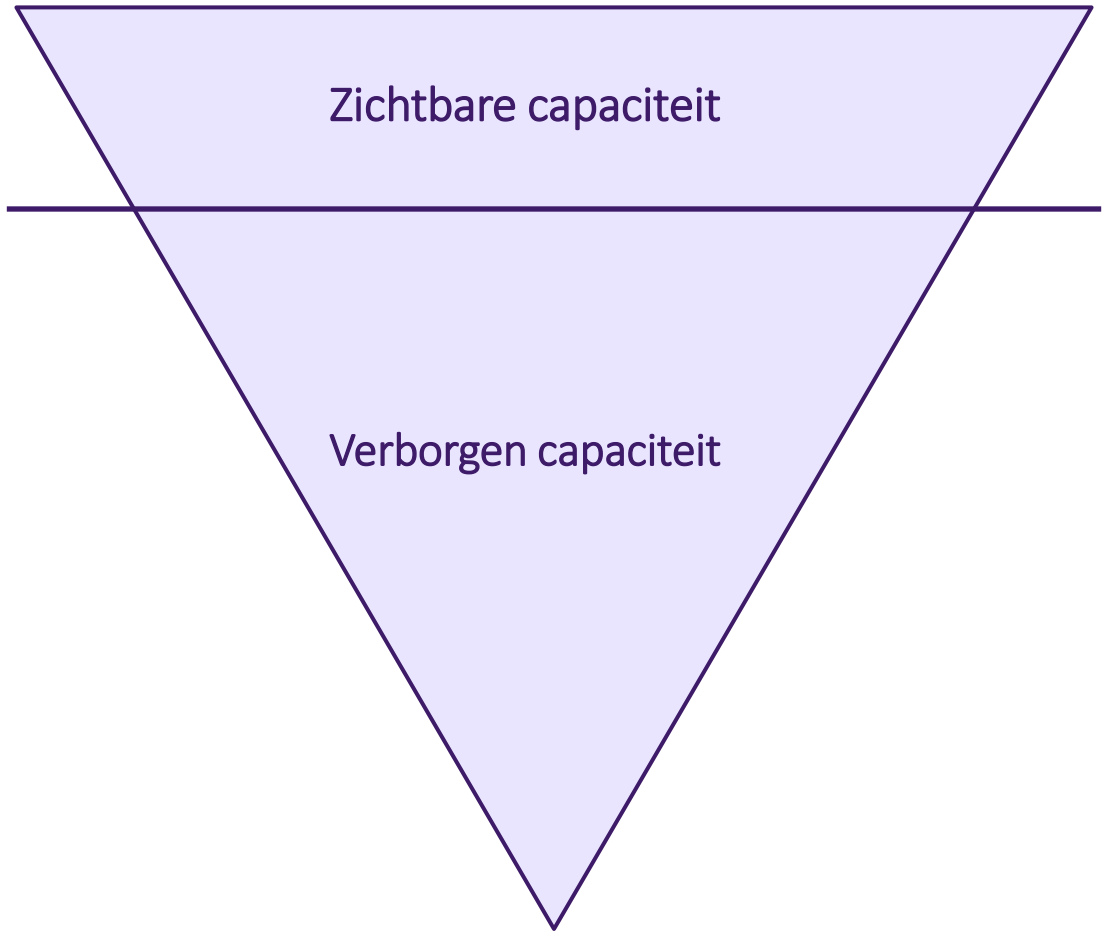
Het prestatiegat tussen Frontier AI modellen en open weight modellen is gekrompen

### Lagere kosten

Open-weightmodellen leveren geavanceerde capaciteiten tegen een fractie van de kosten

Nieuwe realiteit: Geavanceerde cybercapaciteiten zijn niet langer exclusief voor statelijke actoren.

# 3.2. Benchmarks laten slechts het topje van de ijsberg zien




### Huidige benchmarks

Standaard evaluaties beperken de rekenkracht (2.5M Tokens)

### Meer tijd, meer risico

Hoe meer test-time compute beschikbaar is, hoe beter een model complexe cyberaanvallen kan plannen en uitvoeren

### Voorbeeld

Het verhogen van het budget van 1M naar 10M tokens levert een stijging van ongeveer 25% in prestaties

Nieuwe realiteit: Meer rekentijd kan verborgen cybercapaciteiten zichtbaar maken

# 3.3. AI-agents naderen menselijke testers

### Wat agents al kunnen

Complexe stappen combineren, parallel werken en bevindingen opbouwen in een realistische omgeving.

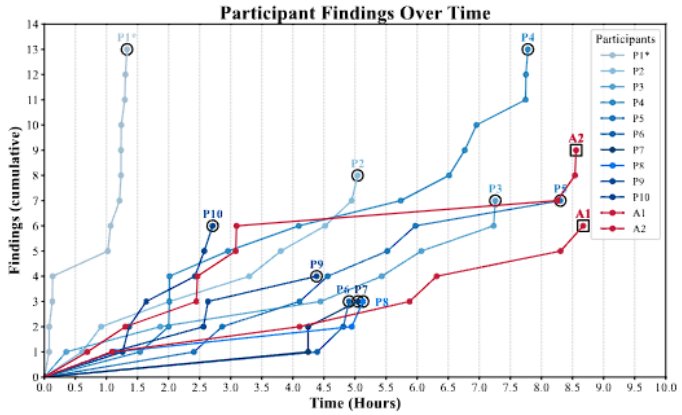
### Wat opvalt

Geavanceerde AI-agents behalen in sommige experimenten vergelijkbare resultaten als ervaren menselijke testers.

### Wat nog nodig blijft

Menselijke validatie, context, toestemming en professionele oordeelsvorming.




Nieuwe realiteit: AI-agents kunnen steeds meer testwerk uitvoeren, maar nog niet zonder menselijke regie en validatie.

04

# Vertrouwen moet aantoonbaar worden

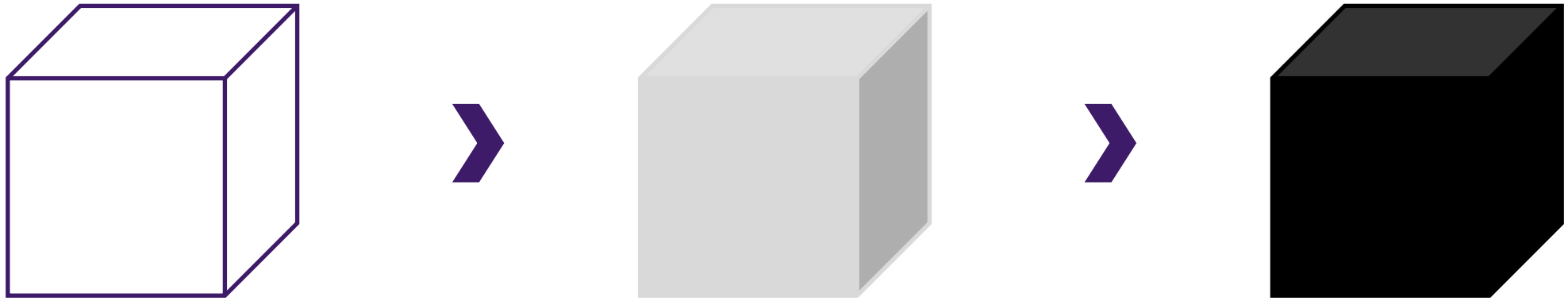
# 4.1. Vertrouwen in GenAI

- Het vertalen van principes voor betrouwbare GenAI naar concrete **governance mechanismen** en **KPI's** is essentieel.
- Autonome cybersecurity operations zullen kritieke adoptiebarrières ondervinden zonder ingebouwde **EU AI principes** en **governance safeguards**.
- **Toezicht op GenAI** wordt complexer naarmate systemen bedrijfsprocessen autonomer ondersteunen.

**Nieuwe realiteit: Toezicht wordt moeilijker naarmate processen autonomer worden.**



# 4.2. Verlies van toezicht



### 1. Transparant (verleden)

Regels en beslissingen zijn direct te volgen.

*We konden de interne beslissingsbomen en patronen van modellen direct uitlezen en controleren*

### 2. Beperkt inzicht (nu)

Redenering is deels verborgen of te complex.

*We vertrouwen op chain of thought terwijl modellen mogelijk niet hun volledige interne afwegingen tonen.*

### 3. Black box (toekomst)

Interne werking is niet leesbaar of niet beschikbaar.

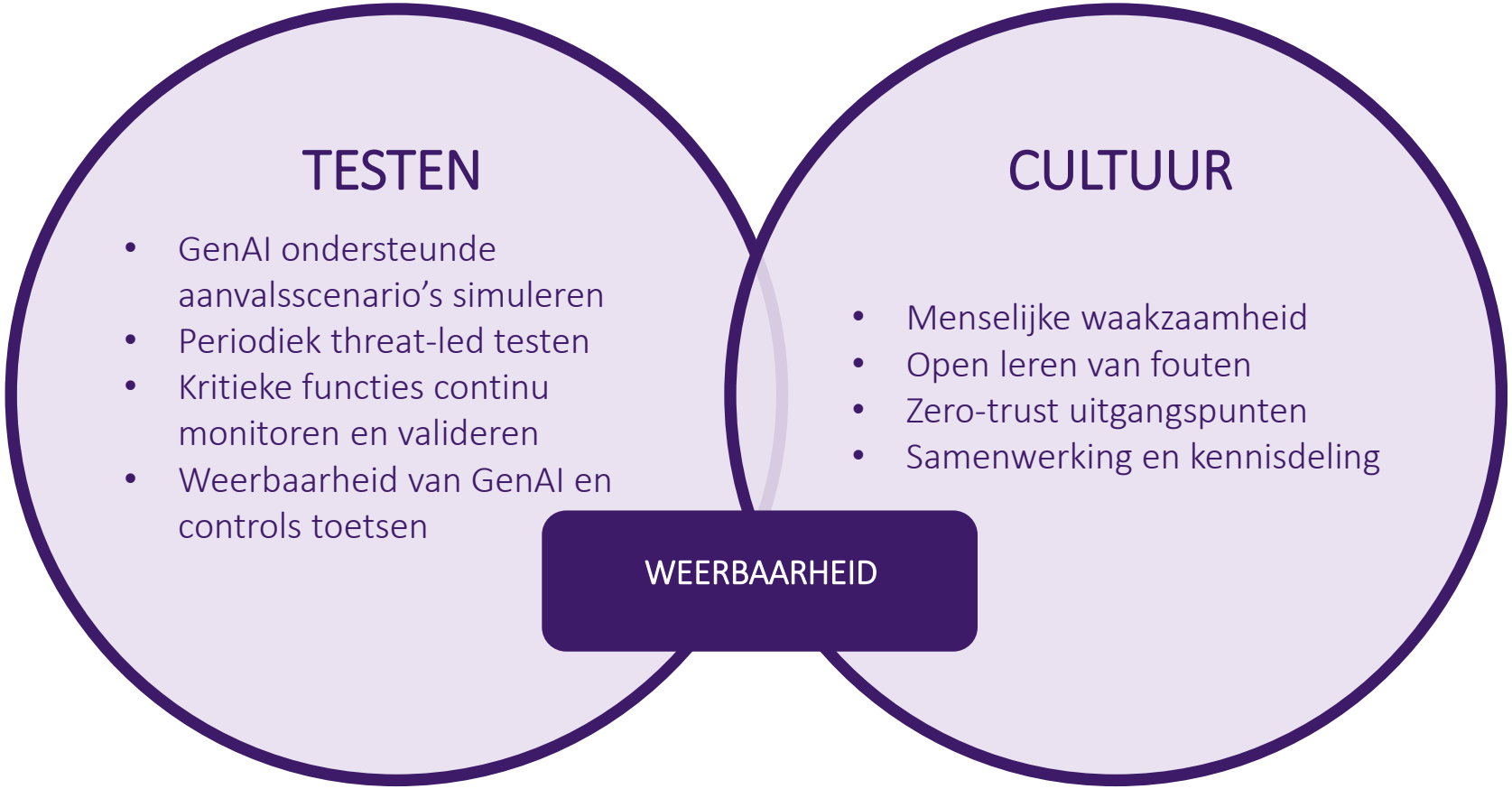
*Modellen kunnen complexe planning steeds vaker intern uitvoeren, in plaats van hun redenering zichtbaar te maken*

Nieuwe realiteit: De veiligheid van geavanceerde AI is afhankelijk van ons vermogen om toezicht te houden

05

# Hoe vertalen we dit naar weerbaarheid?

# 5.1. De strategische keuze: testen en cultuur




Techniek en kritisch menselijk handelen versterken elkaar continu.

# 5.2. Valideer weerbaarheid met DORA TLPT en TIBER-EU

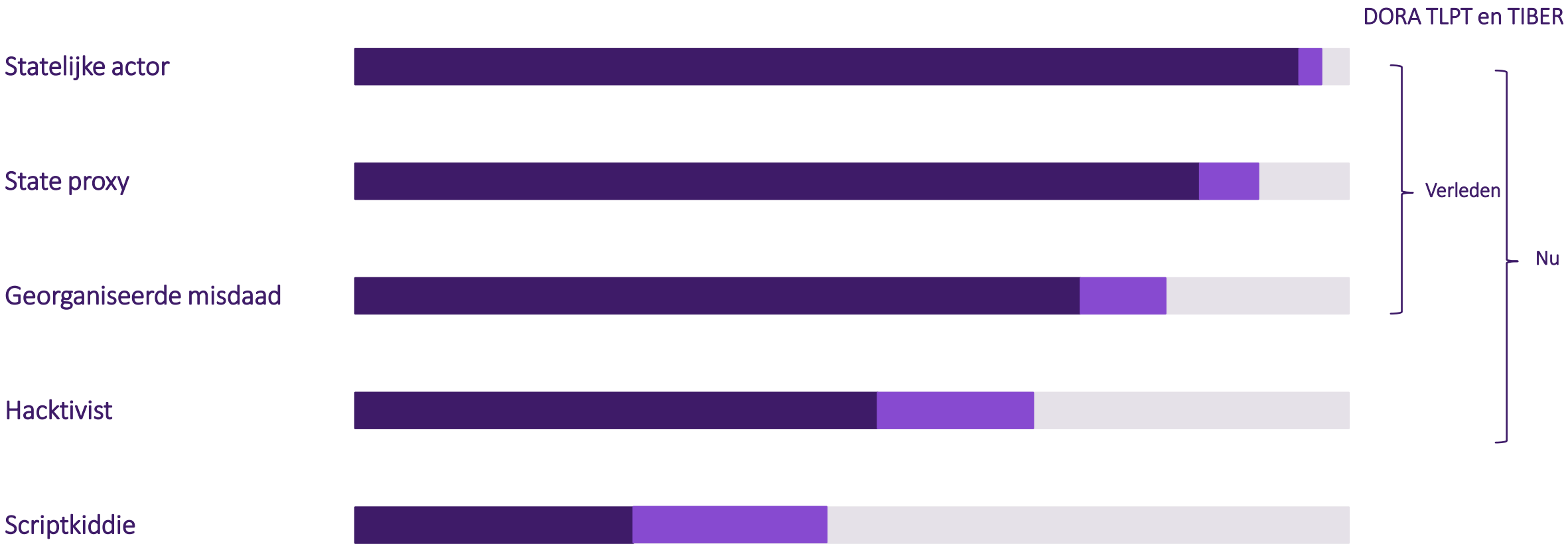






Realistische, dreigingsgestuurde tests op kritieke functies.

# 5.3. Het spectrum van relevante dreigingsactoren verbreedt




Boodschap: Test op capaciteiten en aanvalspaden, niet alleen op actorlabels.

# 5.4. Vier acties voor de securitycommunity



**Experimenteer beheerst**

Gebruik GenAI, maar maak **menselijke controle** en **outputvalidatie** verplicht.



**Versterk detectie en respons**

Bereid **detectie**, **incidentrespons** en **crisismanagement** voor op AI-versnelde aanvalsketens.



**Beveilig de GenAI keten**

Behandel GenAI en leveranciers als onderdeel van het aanvalsoppervlak.



**Valideer en leer**

**Test** kritieke functies onder realistische dreigings-omstandigheden en **leer** open van tekortkomingen.



# Vragen?