GIDS

De complete gids voor AI & LLM Security

Shadow AI, prompt injection, EU AI Act, OWASP Top 10 LLM en wat het kost. Met actuele marktdata en bronvermelding.

Maart 2026

IBGIDS NL

INHOUDSOPGAVE

Kerncijfers op een rij

AI-adoptie groeit explosief, maar beveiligingsmaatregelen blijven achter. Deze cijfers laten zien waar de risico's liggen.

77%
van bedrijven meldde een AI-gerelateerd beveiligingsincident in 2024
IBM Cost of Data Breach 2025 [1]

68%
van medewerkers gebruikt minstens 1 niet-goedgekeurde AI-tool op het werk
Netskope Cloud & Threat Report 2025 [2]

USD 670K
extra kosten per datalek door shadow AI (niet-goedgekeurde AI-tools)
IBM Cost of Data Breach 2025 [1]

60%
van organisaties heeft geen formeel AI-beveiligingsbeleid
Practical DevSecOps / Lakera 2026 [3]

5%
van organisaties voelt zich vertrouwd met hun AI-beveiliging
Practical DevSecOps 2026 [3]

80%+
van phishing-campagnes gebruikt nu AI-gegenereerde of AI-verbeterde content
ENISA Threat Landscape 2025 [4]

USD 1,9M
besparing per datalek bij extensief gebruik van AI security tools
IBM Cost of Data Breach 2025 [1]

Aug 2026
deadline voor EU AI Act high-risk AI-systeemverplichtingen
EU AI Act [5]

1. Wat is AI & LLM Security?

AI & LLM Security richt zich op het beschermen van kunstmatige intelligentie-systemen en large language models tegen misbruik, datalekken en ongeautoriseerd gebruik.

Als je organisatie generatieve AI inzet -- of dat nu ChatGPT, Copilot, een eigen chatbot of een AI-agent is -- dan verwerk je data via systemen die fundamenteel anders werken dan traditionele software. LLMs genereren antwoorden op basis van patronen in hun training, en dat maakt ze kwetsbaar voor manipulatie (prompt injection), datalekken (sensitive information disclosure) en onvoorspelbaar gedrag (hallucinations) [6].

AI Security omvat daarom meerdere disciplines: het beveiligen van de AI-modellen zelf, het controleren van input en output, het bewaken van wie welke AI-tools gebruikt (shadow AI discovery), en het naleven van regelgeving zoals de EU AI Act en NIS2 [5].

OWASP TOP 10 VOOR LLM APPLICATIONS

De OWASP Top 10 voor LLM Applications (2025) is de industriestandaard voor het identificeren van AI-specifieke risico's [6]:

NR	RISICO	OMSCHRIJVING
LLM01	**Prompt Injection**	Manipulatie van modelgedrag via directe of indirecte input
LLM02	**Sensitive Info Disclosure**	Onbedoelde blootstelling van PII of bedrijfsgeheimen
LLM03	**Supply Chain**	Risico's in pre-trained models en third-party componenten
LLM04	**Data Poisoning**	Manipulatie van trainingsdata
LLM05	**Improper Output Handling**	Onvoldoende validatie van modeloutput
LLM06	**Excessive Agency**	Te veel autonomie voor AI-agents
LLM07	**System Prompt Leakage**	Blootstelling van beveiligingsinstructies
LLM08	**Vector & Embedding Weaknesses**	Kwetsbaarheden in RAG-systemen
LLM09	**Misinformation**	Generatie van feitelijk onjuiste content
LLM10	**Unbounded Consumption**	Ongecontroleerd gebruik van AI-capaciteit

2. Waarom is het belangrijk?

De kloof tussen AI-adoptie en AI-beveiliging is alarmerend. 90% van organisaties zet GenAI in, maar slechts 5% voelt zich vertrouwd met de beveiliging ervan.

De financiele impact is concreet: een datalek met een AI-component kostte in 2024 gemiddeld USD 4,88 miljoen, het hoogste bedrag ooit gemeten [1]. Shadow AI -- het gebruik van niet-goedgekeurde AI-tools door medewerkers -- voegt daar gemiddeld USD 670.000 aan toe. En 77% van medewerkers plakt bedrijfsdata in AI-chatbots, waarvan 22% vertrouwelijke persoons- of financiele gegevens bevat [2].

HET SHADOW AI-PROBLEEM

Shadow AI is het grootste directe risico voor de meeste organisaties. 68% van medewerkers gebruikt minstens een niet-goedgekeurde AI-tool [2], en 86% van organisaties heeft nul zichtbaarheid in welke data naar externe AI-providers stroomt [7]. De oplossing is niet verbieden, maar faciliteren: wanneer organisaties goedgekeurde AI-tools aanbieden, daalt ongeautoriseerd gebruik met 89% [7].

> **Business case:** Organisaties die AI security tools extensief inzetten besparen gemiddeld USD 1,9 miljoen per datalek [1]. Bij een gemiddeld MKB-incident van EUR 75.000--150.000 is een AI security assessment van EUR 5.000--15.000 een fractie van de potentiele schade.

3. Hoe werkt het? Het beveiligingsproces

AI Security is geen eenmalige actie, maar een continu proces van inventarisatie, beleid, technische maatregelen en monitoring.

1 AI-inventarisatie en risicoscan

WEEK 1--2

Breng alle AI-systemen en -gebruik in kaart: welke tools worden ingezet, door wie, welke data wordt verwerkt? Detecteer shadow AI. Classificeer systemen op risico (EU AI Act categorisering).

2 AI-beleid en governance

WEEK 2--4

Stel een AI-gebruiksbeleid op: welke tools zijn goedgekeurd, welke data mag worden ingevoerd, wie is verantwoordelijk? Richt een AI governance-structuur in met rollen en verantwoordelijkheden.

3 Technische beveiligingsmaatregelen

WEEK 3--6

Implementeer guardrails: input/output filtering, prompt injection detectie, data loss prevention voor AI-kanalen. Overweeg een LLM firewall voor productie-AI-systemen. Configureer logging en monitoring.

4 Testing en validatie

WEEK 5--8

Voer AI red teaming uit: test je AI-systemen op prompt injection, data leakage, jailbreaks en de OWASP Top 10 LLM. Documenteer bevindingen en stel een remediatie-roadmap op.

5 Continue monitoring en optimalisatie

DOORLOPEND

Monitor AI-gebruik, detecteer anomalieen, update beleid bij nieuwe AI-tools of regelgeving. Train medewerkers periodiek in veilig AI-gebruik. Evalueer en verbeter guardrails op basis van incidenten.

4. Wat kost het?

De kosten voor AI Security variëren sterk, afhankelijk van het aantal AI-systemen, de complexiteit en het gewenste beschermingsniveau.

TIER	OMSCHRIJVING	PRIJSINDICATIE	EENHEID
Basis	AI-inventarisatie, shadow AI scan, quickscan OWASP Top 10 LLM, AI-gebruiksbeleid review	EUR 3.000 -- 8.000	per assessment
Standaard	Volledige AI security assessment: prompt injection testing, data leakage analyse, governance gap analyse, EU AI Act readiness check	EUR 8.000 -- 25.000	per assessment
Premium	Continue AI security monitoring: LLM firewall, real-time prompt scanning, shadow AI detectieplatform, compliance dashboard, managed service	EUR 2.000 -- 8.000	per maand

PRIJSBEPALENDE FACTOREN

- Aantal AI-systemen en LLM-implementaties
- Complexiteit van integraties (standalone chatbot vs. embedded agents)
- Volume data dat door AI-systemen wordt verwerkt
- EU AI Act classificatie (high-risk vs. limited risk)
- Eenmalige assessment vs. continue monitoring
- Aantal medewerkers dat AI-tools gebruikt
- Sector en gevoeligheid van verwerkte data

MKB-TIP

Begin met een basis AI-inventarisatie en shadow AI scan (EUR 3.000--8.000). De meeste MKB-bedrijven ontdekken hiermee al onbekend AI-gebruik dat direct risico vormt. Pas daarna besluit je over continue monitoring.

5. Waar moet je op letten bij een aanbieder?

AI Security is een relatief nieuw vakgebied. Niet elke cybersecurity-aanbieder heeft de expertise in huis.

SELECTIECRITERIA

- **Kennis van OWASP Top 10 LLM** -- De aanbieder moet deze standaard kennen en toepassen
- **Ervaring met EU AI Act compliance** -- Kan de aanbieder je helpen met AI Act readiness?
- **Shadow AI discovery-capaciteit** -- Kan de aanbieder ongeautoriseerd AI-gebruik detecteren?
- **Technische diepgang** -- Ervaring met prompt injection testing, RAG security, LLM firewalls
- **Sector-ervaring** -- AI-risico's verschillen per sector (finance, zorg, overheid)
- **Pragmatische aanpak** -- Focus op werkbare oplossingen, niet alleen verbieden

10 VRAGEN VOOR JE AANBIEDER

1. Hoeveel AI security assessments hebben jullie uitgevoerd?
2. Hoe detecteren jullie shadow AI-gebruik in onze organisatie?
3. Welke OWASP Top 10 LLM-risico's testen jullie concreet?
4. Kunnen jullie ons helpen met EU AI Act readiness?
5. Hoe gaan jullie om met prompt injection testing op productiesystemen?
6. Welke LLM firewall of guardrail-oplossingen adviseren jullie?
7. Hoe ziet jullie rapportage eruit en welke metrics meten jullie?
8. Bieden jullie ook training voor onze medewerkers aan?
9. Wat is de doorlooptijd van een assessment en de vervolgstappen?
10. Kunnen jullie referenties geven van vergelijkbare organisaties?

> **RED FLAGS**
>
> Wees alert als een aanbieder: geen ervaring heeft met LLM-specifieke risico's (alleen traditionele pentesting), de OWASP Top 10 LLM niet kent, geen concreet plan heeft voor shadow AI detectie, of claimt dat een enkele tool alle AI-risico's oplost.

6. Veelgemaakte fouten

AI Security is nieuw terrein. Deze fouten zien we regelmatig bij organisaties die hun AI-gebruik willen beveiligen.

1. AI-gebruik verbieden in plaats van faciliteren

Veel organisaties reageren op AI-risico's door het gebruik te verbieden. Het resultaat: medewerkers gebruiken AI-tools alsnog, maar nu volledig buiten het zicht van IT. Wanneer je goedgekeurde alternatieven aanbiedt, daalt ongeautoriseerd gebruik met 89% [7].

2. AI behandelen als een IT-project

AI Security raakt de hele organisatie: HR (medewerkergebruik), juridisch (EU AI Act), compliance, IT en management. Behandel het als een organisatiebreed governance-onderwerp, niet als een IT-taak.

3. Alleen focussen op externe dreigingen

Het grootste AI-risico voor de meeste organisaties is intern: medewerkers die bedrijfsdata in publieke AI-tools plakken. 77% doet dit, en 22% deelt daarbij vertrouwelijke informatie [2].

4. Geen AI-inventaris bijhouden

Je kunt niet beveiligen wat je niet kent. 60% van organisaties heeft geen formeel AI-beveiligingsbeleid [3], en de meeste weten niet eens welke AI-tools in gebruik zijn.

5. EU AI Act uitstellen

De high-risk AI-verplichtingen gelden vanaf augustus 2026. Conformiteitsbeoordelingen, quality management systemen en logging-vereisten kosten maanden om te implementeren. Begin nu, niet wanneer de deadline nadert.

6. Vertrouwen op een enkele technische maatregel

Een LLM firewall alleen is niet genoeg. Effectieve AI Security combineert technische maatregelen (guardrails, monitoring), governance (beleid, rollen), en cultuur (training, bewustwording).

7. Compliance: EU AI Act, NIS2 en regelgeving

De regelgeving rond AI verscherpt snel. Twee wetten zijn bijzonder relevant: de EU AI Act en de Cyberbeveiligingswet (NIS2).

EU AI ACT

De EU AI Act is de eerste alomvattende AI-wetgeving ter wereld. De wet classificeert AI-systemen in risicocategorien [5]:

CATEGORIE	VOORBEELDEN	VERPLICHTINGEN
Verboden	Social scoring, manipulatieve AI	Volledig verboden (sinds feb 2025)
High-risk	AI in HR, kredietbeoordeling, kritieke infra	Conformiteitsbeoordeling, logging, menselijk toezicht (aug 2026)
Limited risk	Chatbots, deepfakes	Transparantieverplichtingen
Minimaal risico	Spamfilters, AI in games	Geen specifieke verplichtingen

Boetes: Tot EUR 35 miljoen of 7% van jaaromzet (verboden praktijken), tot EUR 15 miljoen of 3% (high-risk), tot EUR 7,5 miljoen of 1% (overige) [5].

NIS2 / CYBERBEVEILIGINGSWET

De Cyberbeveiligingswet (Nederlandse implementatie van NIS2) treedt naar verwachting Q2 2026 in werking [8]. AI-systemen die onderdeel zijn van netwerk- en informatiesystemen vallen onder de scope. Circa 10.000 bedrijven worden direct geraakt, met boetes tot EUR 10 miljoen of 2% van jaaromzet.

AVG / GDPR

AI-systemen die persoonsgegevens verwerken moeten voldoen aan de AVG. De Autoriteit Persoonsgegevens focust in 2025 op toezicht op algoritmen en AI [9]. Boetes tot EUR 20 miljoen of 4% van jaaromzet.

8. Verschil met verwante oplossingen

AI Security overlapt met andere cybersecurity-disciplines, maar heeft een eigen focus.

DISCIPLINE	FOCUS	VERSCHIL MET AI SECURITY
AI Security	Beveiliging van AI-systemen en LLMs	Specifiek gericht op AI-risico's: prompt injection, shadow AI, model security
Data Security	Bescherming van data (encryptie, DLP)	Bredere scope, AI Security focust op data in AI-context
Application Security	Beveiliging van applicaties (SAST, DAST)	AI-applicaties hebben unieke risico's die traditionele AppSec niet dekt
Cloud Security	Beveiliging van cloud-infrastructuur	AI draait vaak in de cloud, maar AI Security gaat over het model, niet de infra
Security Awareness	Training van medewerkers	AI Security omvat ook awareness, maar voegt technische AI-beveiligingslagen toe

9. Trends 2025--2026

LLM Firewalls

Een nieuw marktsegment (geschat op USD 30--260 miljoen in 2025) dat snel groeit [10]. LLM firewalls filteren prompts en responses in real-time op injection-aanvallen, data leakage en policy violations. Vergelijkbaar met hoe een web application firewall (WAF) webverkeer filtert, maar dan specifiek voor AI-communicatie.

Agentic AI Security

Met de opkomst van autonome AI-agents -- systemen die zelfstandig taken uitvoeren, tools aanroepen en beslissingen nemen -- ontstaan nieuwe risico's rond "excessive agency" (OWASP LLM06). Het beveiligen van AI-agents vereist nieuwe aanpakken voor autorisatie, scope-beperking en menselijk toezicht [6].

AI Security Governance Gap

AI-tools zijn bij 73% van organisaties ingezet, maar real-time governance is bij slechts 7% geimplementeerd [11]. Dit gat van 66 procentpunt is het grootste risico voor 2026. Organisaties die nu investeren in governance lopen voorop.

10. Aan de slag

Begin vandaag met het in kaart brengen van je AI-gebruik. De eerste stap is altijd: weten wat je hebt.

DRIE DIRECTE ACTIES

1. Maak een inventaris van alle AI-tools die in je organisatie worden gebruikt (inclusief shadow AI)
2. Stel een basis AI-gebruiksbeleid op: welke tools zijn goedgekeurd, welke data mag worden gedeeld?
3. Plan een AI security assessment om je specifieke risico's te identificeren

> **Hulp nodig?** Op ibgids.nl/word-gematcht word je vrijblijvend gematcht met AI security specialisten die passen bij jouw sector, bedrijfsgrootte en budget. Binnen 48 uur ontvang je een vergelijking.

Of neem contact op via info@ibgids.nl voor persoonlijk advies.

Bronnenlijst

[1] **IBM Cost of Data Breach Report 2025** -- ibm.com/reports/data-breach

[2] **Netskope Cloud & Threat Report: Shadow AI 2025** -- netskope.com/resources/cloud-and-threat-reports/cloud-and-threat-report-shadow-ai-and-agentic-ai-2025

[3] **Practical DevSecOps AI Security Statistics 2026** -- practical-devsecops.com/ai-security-statistics-2026-research-report/

[4] **ENISA Threat Landscape 2025** -- enisa.europa.eu/publications/enisa-threat-landscape-2025

[5] **EU AI Act - High-level Summary** -- artificialintelligenceact.eu/high-level-summary/

[6] **OWASP Top 10 for LLM Applications 2025** -- owasp.org/www-project-top-10-for-large-language-model-applications/

[7] **Lasso Security - What is Shadow AI?** -- lasso.security/blog/what-is-shadow-ai

[8] **NCSC - Cyberbeveiligingswet (NIS2)** -- ncsc.nl/cyberbeveiligingswet-nis2/bereid-je-voor

[9] **GDPR Enforcement Netherlands** -- cms.law/en/deu/publication/gdpr-enforcement-tracker-report/netherlands

[10] **TechTarget - LLM Firewalls** -- techtarget.com/searchsecurity/feature/LLM-firewalls-emerge-as-a-new-AI-security-layer

[11] **Exabeam AI Accountability Report** -- teleinfotoday.com/press-releases/exabeam-research-ai-accountability-becomes-the-new-mandate-as-cybersecurity-economics-shift

[12] **ISACA - Shadow AI Auditing 2025** -- isaca.org/resources/news-and-trends/industry-news/2025/the-rise-of-shadow-ai-auditing-unauthorized-ai-tools-in-the-enterprise

[13] **Lakera AI Security Trends 2025** -- lakera.ai/blog/ai-security-trends

[14] **CBS Cybersecuritymonitor 2024** -- cbs.nl/nl-nl/longread/aanvullende-statistische-diensten/2025/cybersecuritymonitor-2024?onepage=true

[15] **Verizon Data Breach Investigations Report 2025** -- verizon.com/business/resources/reports/dbir/

[16] **Legal Nodes - EU AI Act 2026 Updates** -- legalnodes.com/article/eu-ai-act-2026-updates-compliance-requirements-and-business-risks

[17] **Thales Data Threat Report 2025** -- thalesgroup.com/en/press_release/2025-thales-data-threat-report-reveals-nearly-70-organizations-identify-ais-fast

[18] **OWASP Gen AI Security Project** -- genai.owasp.org/