Deloitte.

Reimagining the AI-enabled cyber defense organization

As attack surfaces expand and cyber threats continue to increase, it is time to modernize the approach to cyber defense.

Key trends shaping the future of cyber defense:

- Artificial intelligence (AI) agents will be implemented in 60% of all IT operations tools by 2028, which is an increase from less than 5% in 2024[1].
- This shift reflects a broader trend, with 58% of organizations now leveraging AI for threat detection and response[2].
- AI emboldens attackers to launch adversarial attacks, manipulate models, and exploit other risks, such as algorithmic bias and security vulnerabilities in AI models.
- In response, 60% of organizations report fewer security incidents due to AI-driven tools[2].
- Through 2027, 40% of organizations will have adopted formal exposure validation initiatives, most relying on AEV technologies and managed service providers for maturity and consistency[3].
- These trends underscore AI's transformative impact on innovation and operational efficiency, while highlighting the need for broad safeguards and adaptive risk management.

The future of cyber defense: Embracing AI

A blueprint for an AI-powered cyber defense: This model deconstructs and compartmentalizes the processes that run the cyber defense function, provides a foundation to map the current tech capability, and enables a mapping of AI to reimagine cyber defenses. By integrating technology, workforce, operating models, and governance, it offers a practical framework to help organizations modernize their security posture, incorporate AI-powered services, and strengthen defenses against emerging threats. The legend below identifies the new and evolved AI-enhanced services.

Threat monitoring
- Threat intelligence
- Threat hunting
- Threat modelling
- Vulnerability identification
- Security monitoring and analytics
- Insider threat

Data onboarding and enrichment
- Asset discovery
- Data collection/log source onboarding
- Threat actor collection and attribution
- Data normalization and enrichment

Data analytics
- Big data management
- Advanced analytics
- Model training and validation

Governance
- Oversight and decisions
- Policy and standard
- Operating model
- Strategy and roadmap

Reporting
- Key performance indicator (KPI) reporting
- Risk reporting
- Security dashboard/health reporting

Audit and compliance
- Audit support
- Evidence collection
- Regulatory/control requirements

Training
- Workforce training and upskilling
- Wargaming/simulation exercises

Core solutions for cyber defense
- **Attack surface management (ASM)**
- **Threat detection and response**
- **Incident response**
- **Security architecture and engineering**

Response
- Investigation and forensics
- Legal, regulatory coordination
- Recovery coordination
- Crisis communication

Detection engineering
- Data engineering
- Security orchestration, automation, and response (SOAR)
- Content development and tuning
- Detection simulation

Offensive security
- Application and AI security testing
- Continuous penetration testing
- Presence monitoring and reconnaissance
- Red team exercises

Exposure management
- Vulnerability management
- Risk assessment and prioritization
- Configuration management
- Bug bounty

Controls validation
- Breach attack simulation
- Endpoint configuration analysis
- Attack path enumeration
- Human response analysis

Operations
- **Level 1 support**
  - Monitor and alert triage
  - Basic investigation
  - Incident escalation
- **Level 2 support**
  - In-depth investigation and troubleshooting
  - Remediation and reporting
- **Level 3 support**
  - Incident response (IR)/ engineering
  - Tuning and automation
  - Post incident review

Service management
- Standard operating procedure (SOP) management
- Standard operating procedure and performance management
- Knowledge articles and runbooks
- Security incident management

Remediation
- Root cause analysis
- Remediation orchestration
- Incident quarantine and elimination

Platform/solution deployment and administration
- Design and architecture
- Platform/solution build
- Testing
- Solution deployment
- Access control
- Platform administration and maintenance

LEGEND

- **AI-led (significant AI uplift)**
- **AI-assisted (partial AI uplift)**
- **AI-native**

1. Gartner® Research: AI Agents Will Transform Enterprise IT Operations, Cameron Haight, February 2025. 2. Everest Group: Gen AI and the Future of Cybersecurity: Advanced Strategies for Cyber Defense, Yugal Joshi, Kumar Avijit, Arjun Chauhan (2025). 3. Gartner: Market Guide for Adversarial Exposure Validation, Eric Ahlm, Dhivya Poole, Angela Zhao, Mitchell Schneider, March 2025. 4. Gartner® Research: Predicts 2025: Navigating Imminent AI Turbulence for Cybersecurity, Jeremy D'Hoinne, Akif Khan, Manuel Acosta, Avivah Litan, Deepak Seth, Bart Willemsen February

Connect to accelerate

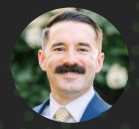
Contact our leaders to dive deeper into the blueprint and reimagine what's possible for your organization.

Mark Nicholson
Principal
Cyber AI GTM Leader
Deloitte & Touche LLP
manicholson@deloitte.com

Naresh Persaud
Principal
AI Transform Leader
Deloitte & Touche LLP
napersaud@deloitte.com

Sharon Chand
Principal
Defense & Resilience Leader
Deloitte & Touche LLP
shchand@deloitte.com

Kamaljeet Singh
Managing Director
Defense Leader
Deloitte & Touche LLP
kamalsingh@deloitte.com



Steve Ruzzini
Senior Manager
Cyber AI GTM
Activation Lead
Deloitte & Touche LLP
sruzzini@deloitte.com

AI for defense: Functional uplifts to achieve greater efficiency

Advanced phishing detection
AI analyzes behavior patterns and email content, which may result in up to a **60-70%** boost in threat detection and remediation

AI-empowered ransomware detection and response
AI agents monitor network activity and can potentially enable **40–60%** improvement in detection and response effectiveness

Digital security operations centers (SOC) analyst
Language models automate incident analysis with the potential to generate **30–35%** cost savings through efficiency, visibility, and risk reduction

AI penetration tester agent (Deloitte's Agent Smith)
Modular agents simulate real-world attacks and offer a potential for up to **30-40%** efficiency gain for AI pen testers

Enhanced anomaly detection for operational technology (OT)
AI refines anomaly detection models with the potential to generate up to **90%** precision in anomaly scores and severity

Automated vulnerability testing and remediation
AI automates scans and tracks remediation with the potential for **40–50%** efficiency gains, enabling exposure solution transformation

**Percentages are estimations based on recent project delivery for 10-15 organizations, ranging from 12-week implementation to multi-year operate engagements, and internal testing in a simulated environment*

Defense for AI: Securing the adoption of AI in the modern enterprise

AI brings new challenges for cyber defenders to address.

Challenges	What is typically done today	How to overcome challenges
Adversarial attacks: Attackers use AI to automate attacks, evade security, and create deepfakes for fraud	Rely on signature-based detection, manual review, and basic anomaly monitoring	Deploy adversarial training for models, use AI-powered threat detection, and invest in deepfake detection tools
Denial of service: Resource-heavy operations from unpredictable inputs can degrade service or drive up costs	Monitor network traffic and set static rate limits	Implement adaptive rate limiting, use AI to detect resource abuse patterns, and isolate critical AI services
Low-and-slow attacks: AI can orchestrate attacks that occur gradually over time, staying below traditional detection thresholds	Focus on high-volume alerts and periodic manual review	Employ behavioral analytics, long-term pattern analysis, and continuous monitoring for subtle anomalies
Polymorphic malware: AI enables malware to change its code or behavior dynamically, defeating signature-based detection tools	Use traditional antivirus and static malware signature	Adopt AI-driven endpoint protection, behavioral-based malware detection, and regular threat intelligence updates
Automated credential stuffing: AI can rapidly test stolen credentials in ways that mimic legitimate user behavior, evading detection	Block known bad IPs, use basic login attempt limits, and static CAPTCHAs	Implement multifactor authentication, credential stuffing detection tools, and user behavior analytics
Model theft and manipulation: Application programming interface (API) exposure and reverse engineering can enable sensitive data theft or misuse	Restrict API access with keys and monitor usage volume	Use API abuse detection, watermark models, encrypt model parameters, and monitor for abnormal query patterns