A GenAI Reality Check: Security, Privacy, and Governance

Jim Richberg, Head of Cyber Policy and Global Field CISO

March 2026

Agenda

- Definitions and level setting
- The big picture and key paradoxes
- Recommendations

Level Setting

2 main types of Artificial Intelligence: **Predictive/Discriminative** and **Generative**

I will focus on ***transformer***-based Large Language Model (***LLM***) Natural Language Processing (***NLP***) models that leverage ***Deep Learning***

There are multiple types of GenAI, but all share the following attributes:

- **Non-deterministic behavior:** models are probability-based, which means their answers can vary even when the same question is asked multiple times
- **Lack of memory:** Models do not retain memory of previous interactions and treat each query independently. Developers are trying to overcome this, which can bring major benefits but which poses significant security and privacy risks
- **Context window limitations:** There is a limit to how much data a model can process in a single query, which restricts the depth and detail of its responses.
- **Hallucinations:** state of the art commercial models currently hallucinate 0.7-10% on average

Rapidly expanding GenAI use

- **Gen AI usage in organizations is doubling annually**
- 86% (?!) of workers use AI tools at least once a week
- 60% of employees are willing to use “shadow AI” (unapproved tools or accounts) if it helps them get the the job done/meet work deadlines
- 60% of leaders acknowledge AI-based cyber threats are outpacing their security team’s ability to deal with them.
- 70% of C-level executives are willing to let faster production outweigh concerns over security in GenAI use
- Non-technical leaders and the work force are currently twice as likely as IT and security professionals to embrace the widespread use of GenAI
- ~66% of business executives and IT professionals said their organization’s use of GenAI outpaces their understanding of the technology

Key Paradoxes as we embrace the use of GenAI

Fundamental Gaps in our understanding of GenAI

- Generalization
- “Grokking”

Fundamental/structural security vulnerabilities

- Prompt injection
- Data poisoning

Security/Privacy Recommendations: *Structure and Process*

- Governance matters! ***Establish and iterate on an AI Governance process***

  We don't yet have a standardized model, so "when in doubt, look to NIST", which has a Risk Management Framework

  - Understand AI workloads by defining their purpose, data sources, and intended outcomes.
  - Apply Responsible AI principles (privacy, security, fairness, inclusiveness, transparency, accountability) to identify vulnerabilities.
  - Evaluate external dependencies like third-party datasets, APIs, and models for security, bias, and compliance risks.
  - Assess integration risks with existing systems to prevent cascading failures, performance bottlenecks, or security gaps.
  - The Framework includes policy documentation focused on governance

- **Appoint a Chief AI Officer**
- **Produce and publish a good faith AI research Safe Harbor policy**

Security/Privacy Recommendations: *Practical 'technical' advice*

- Don't rely on guardrails to adequately protect you!
- Don't start from scratch by ignoring security basics
- Leverage GenAI for cybersecurity
- **Focus on AI identity & access control and account privileges**
  - Consider a "virtual HR"-like process for onboarding and terminating AI agents
- Secure your AI access points

Security/Privacy Recommendations: for programmers

- Inspect AI-generated code for security as well as functionality!
- Use controlled/vetted open-source libraries
- For ‘vibe coding’, run the code through the LLM again with a prompt to *find and fix known vulnerabilities* before it is used

  “the 5 R’s of AI-application development”

  - Reason—have a clearly defined goal
  - Restrict (make small incremental changes)
  - Review (assume subtle failures are present)
  - Revert (have a rollback capability)
  - Retest after any changes
    (content from a ‘lessons learned’ presentation at Stanford University)

Closing Thoughts

- Leverage the full range of potential partners
- Shape/participate in AI governance within your organization
- Ask questions!

Thanks for your attention, and feel free to contact me at:

Jrichberg@Fortinet.com

Jim Richberg | LinkedIn

For a deeper dive, scan the QR code to register for the upcoming webinar:

Securing AI Usage in Practice: Identity, Visibility, and Guardrails in 2026


FORTINET