AI Security Is Not a Brand

Governance, Risk, and the Reality Behind “Safe AI”

Doug Jacobson

IOWA STATE UNIVERSITY

The Claim: “We can only use Copilot… everything else is unsafe.”

- Many organizations hear this.
- Is it really a security argument?
- **The problem with this is:**
  - Shuts down conversation
  - Limits innovation
  - Creates false confidence
  - Replaces thinking with tool selection

What This Talk Is

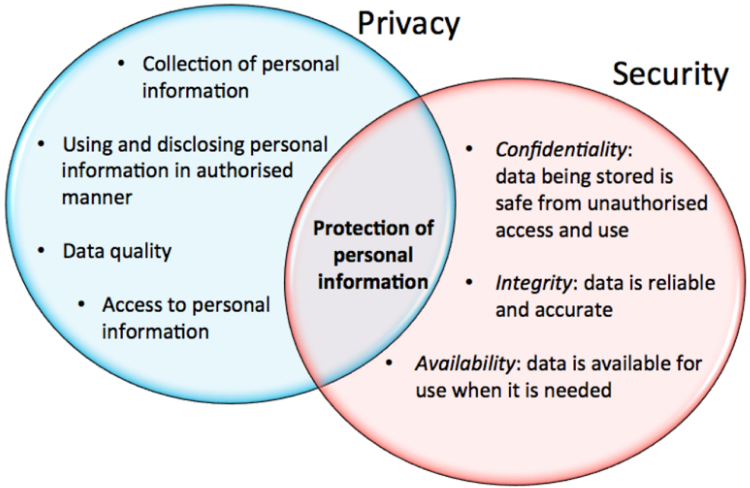
- NOT vendor comparison
- NOT anti-Microsoft

- IS about risk, decisions, governance

Core Thesis

- Security is not determined by brand

- Security is determined by:
  - Governance
  - Identity & access
  - Data controls
  - Architecture
  - Agreements



How many organizations deploy AI

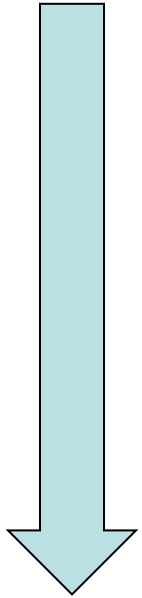
AI TOOL: “We use Copilot → We’re safe”

IMPLEMENTATION: (Undefined / inconsistent)

DATA CONTROLS: “Don’t put sensitive data…” no enforcement)

IDENTITY & ACCESS: “Everyone has access”

GOVERNANCE: Policies? Agreements? Later…



How we should deploy AI

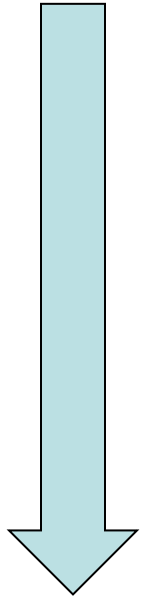
GOVERNANCE: Policies • Legal • Compliance • Agreements

IDENTITY & ACCESS: Who • Authentication • Roles

DATA CONTROLS: Classification • Retention

IMPLEMENTATION: APIs • Integrations • Guardrails • Monitoring

AI TOOL: (Copilot, ChatGPT, etc.)

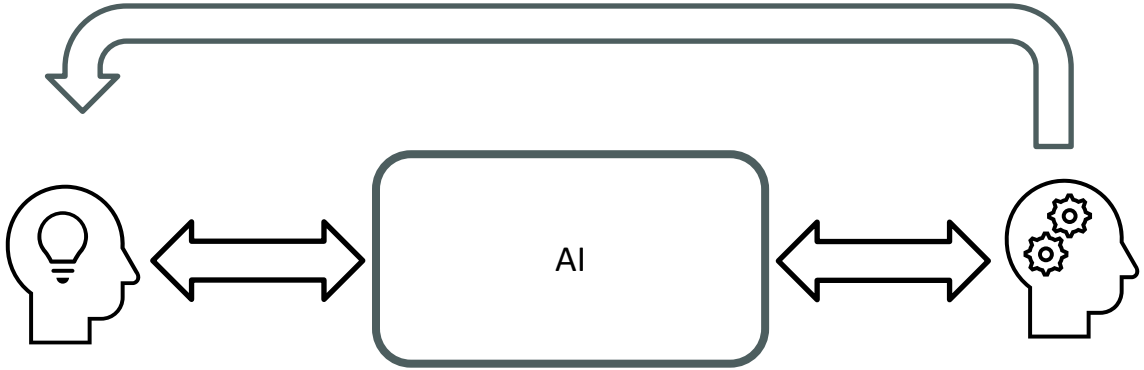


How AI Works

- Predicts patterns
- Generates responses
- Does not think

- Risk depends on use

AI / Human interaction model



Human Assisted AI
- Clarity of requests
- Ideate
- Reenforce
- Refine

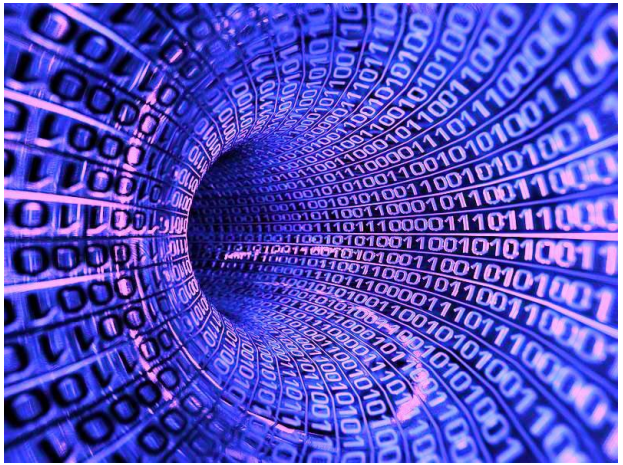
Guardrails
- Ethics
- Privacy
- Bias

AI Assisted Human
- Verify
- Question
- Challenge
- Ideate

Source: Doug Jacobson, Stephanie Hansen

Where Does Data Go?

- Depends on platform, config, API, agreements
- Not all tools behave the same



The Real Question

- NOT: Which AI is safe?
- IS: What risks exist in how we use

Four Types of Risk

- Cybersecurity
- Perceived
- Compliance
- Vendor comfort

Cybersecurity Risk

- Data leakage
- Identity compromise
- Misconfiguration
- API misuse
- Bias
- Validation

Data Leakage

- Goes beyond feeding the model
- Sensitive prompts
- Internal exposure

- Mitigation: classify data, control access

Bias

- Bias in data and prompts
- AI has inherent bias based on the data and the question
  - Impacts decisions
- AI might not have the most recent answer
- Review results

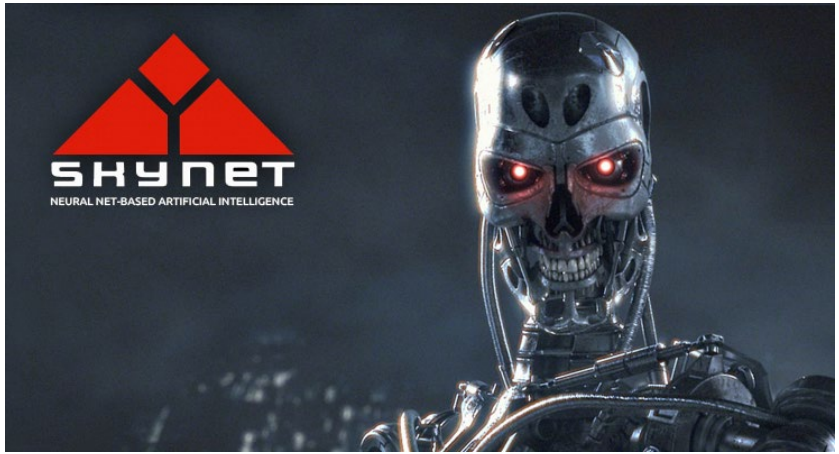
Validation

- AI hallucinations
- Bad data

- Always verify

Perceived Risk

- Fear driven by headlines
- Misunderstanding
- Over-restriction



Compliance Risk

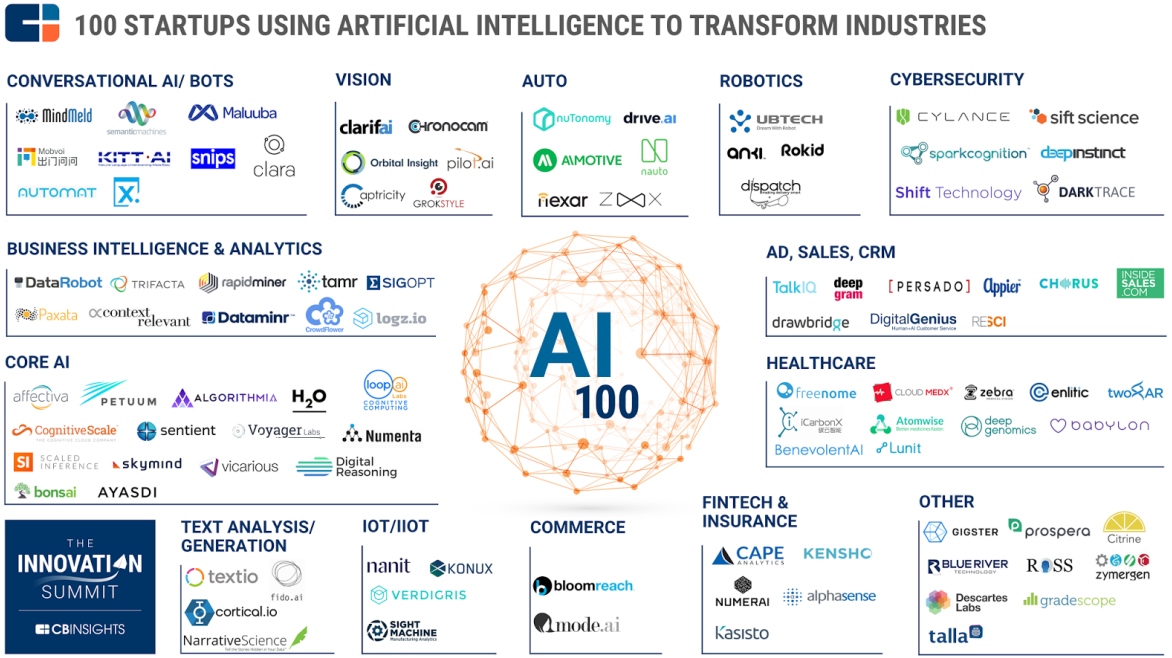
- Regulations
- Data handling
- Retention
- Auditability
- Disclosure
- Ownership
- Legal exposure

Governance Framework

- Objectives
- Stakeholders
- Risk assessment
- Policies
  - Training policies
  - Acceptable use
  - Vendor agreements

Vendor Comfort

- Familiar vendor
- Easy approvals
- Comfort ≠ security



What Makes AI Safe

- Identity
- Data classification
- Monitoring
- API governance

Questions to Ask

- Where does data go?
- Who has access?
- Is it stored?
- Training use?
- Contract terms?

Final Takeaways

- Security ≠ vendor
- Governance > tools
- AI is a tool

Q&A

- You don’t outsource responsibility when you outsource tools.

- dougj@iastate.edu