SIA EDUCATION at ISC

The Technology Ethic: Managing Risk in AI Adoption

F. Paul Greene, PhD, AIGP, CIPP/US/E, CIPM, FIP

In partnership with SIA EDUCATION at ISC

Built by RX In the business of building businesses

- Chair, Artificial Intelligence and New Technologies Practice Group
- Chair, Privacy and Data Security Practice Group
- Technology Litigator
- PhD, Germanic Languages and Literatures

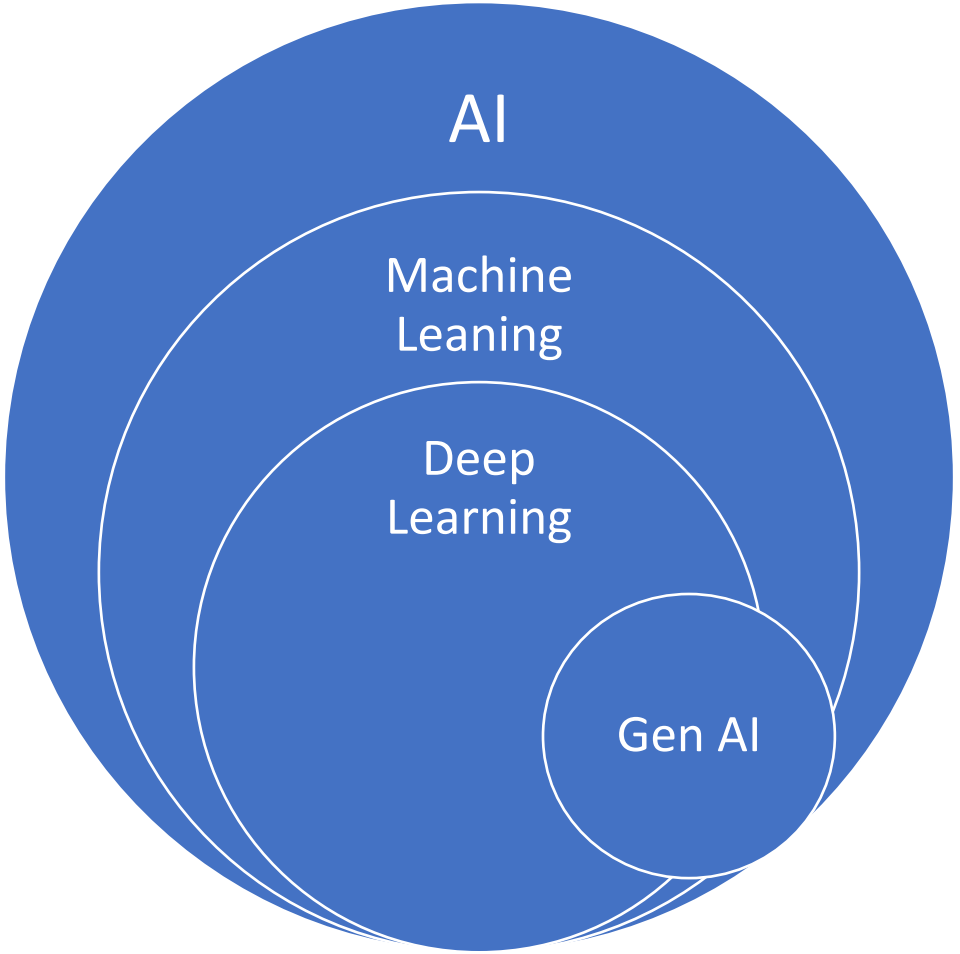
Prompt transformation

- create my avatar
- give a bald head and a goatee
- make it look more like me
- transformed prompt: an avatar of a user interested in technology, ethics, and pop culture, with a bald head and a beard

What is AI?

An engineered or machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments

What is AI?



What is AI?

- **Machine Learning (ML)**
  - Enables machines to improve with experience
  - Uses algorithms to learn from data and make decisions
  - Requires structured data and human intervention
  - Examples: linear regression algorithms, image recognition, propensity scores, behavior predictors and analytics
- **Deep Learning (DL)**
  - Specialized ML with neural networks
  - Learns from unstructured data (images, text)
  - Needs large data sets and computational power
  - Mimics the way a human brain operates – can process and learn from data
  - Examples: deep blue, autonomous vehicles, image and speech recognition

What is AI?

- **Generative AI**
  - Creates new content (text, images, music)
  - Employs ML and DL techniques
  - Generates innovative outputs not explicitly programmed: chaos built in
  - Examples: ChatGPT, Stable Diffusion, modern chatbots, DeepSeek
- **Key Differences**
  - Data Requirements: DL > ML; Gen AI is only as good as its source data
  - Learning Process: ML is guided; DL discovers patterns; Gen AI can repeat, accentuate, create bias
  - Output: ML/DL interpret; Gen AI creates; hallucinates*

What is AI?

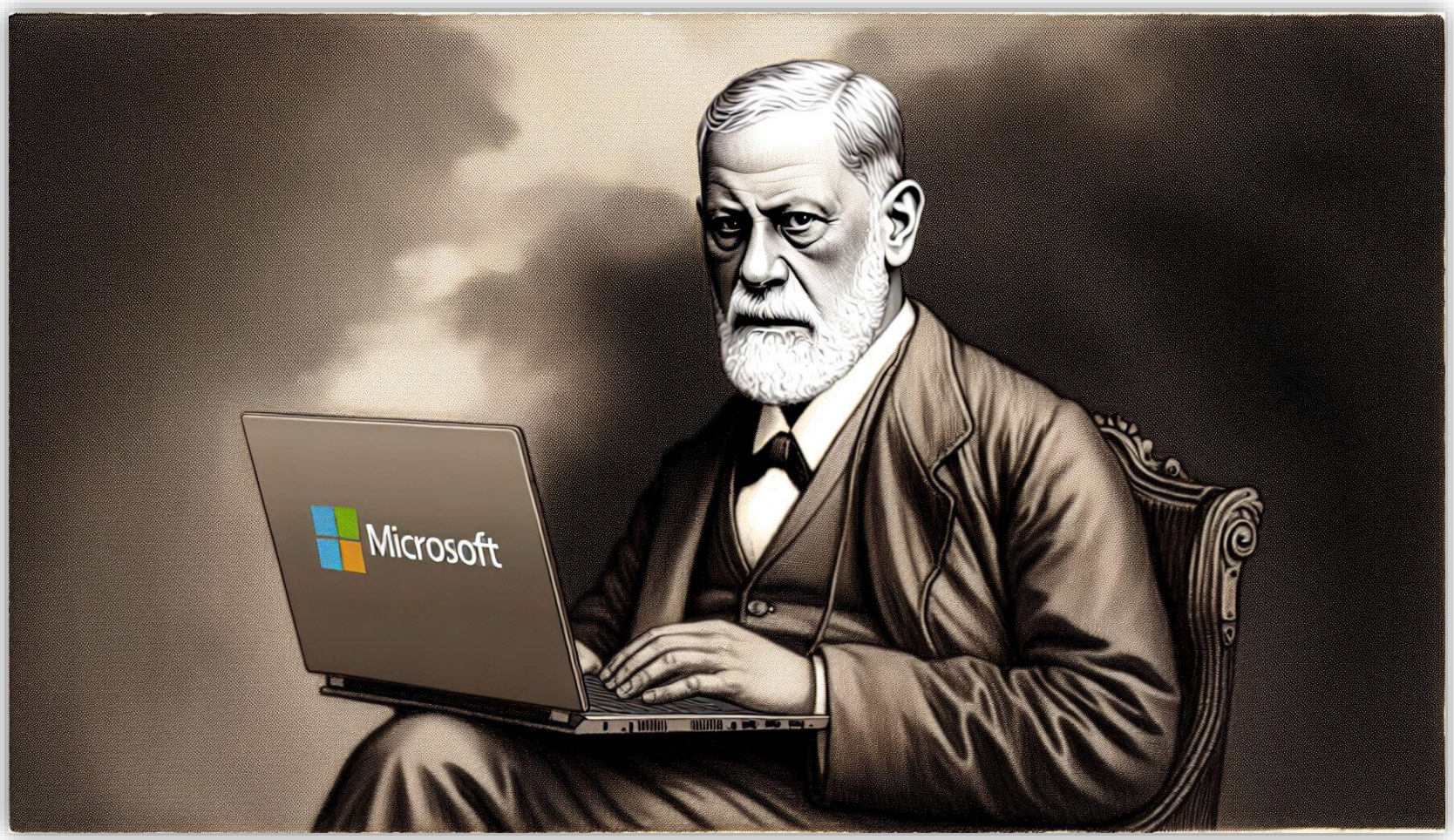
- **Agentic AI**
  - Built on generative platform, hence exhibits same benefits and risks
  - Can automate routine functions
  - Legal and ethical grey zone: what can your AI agent agree to and who's responsible for its actions?

All that glitters . . .

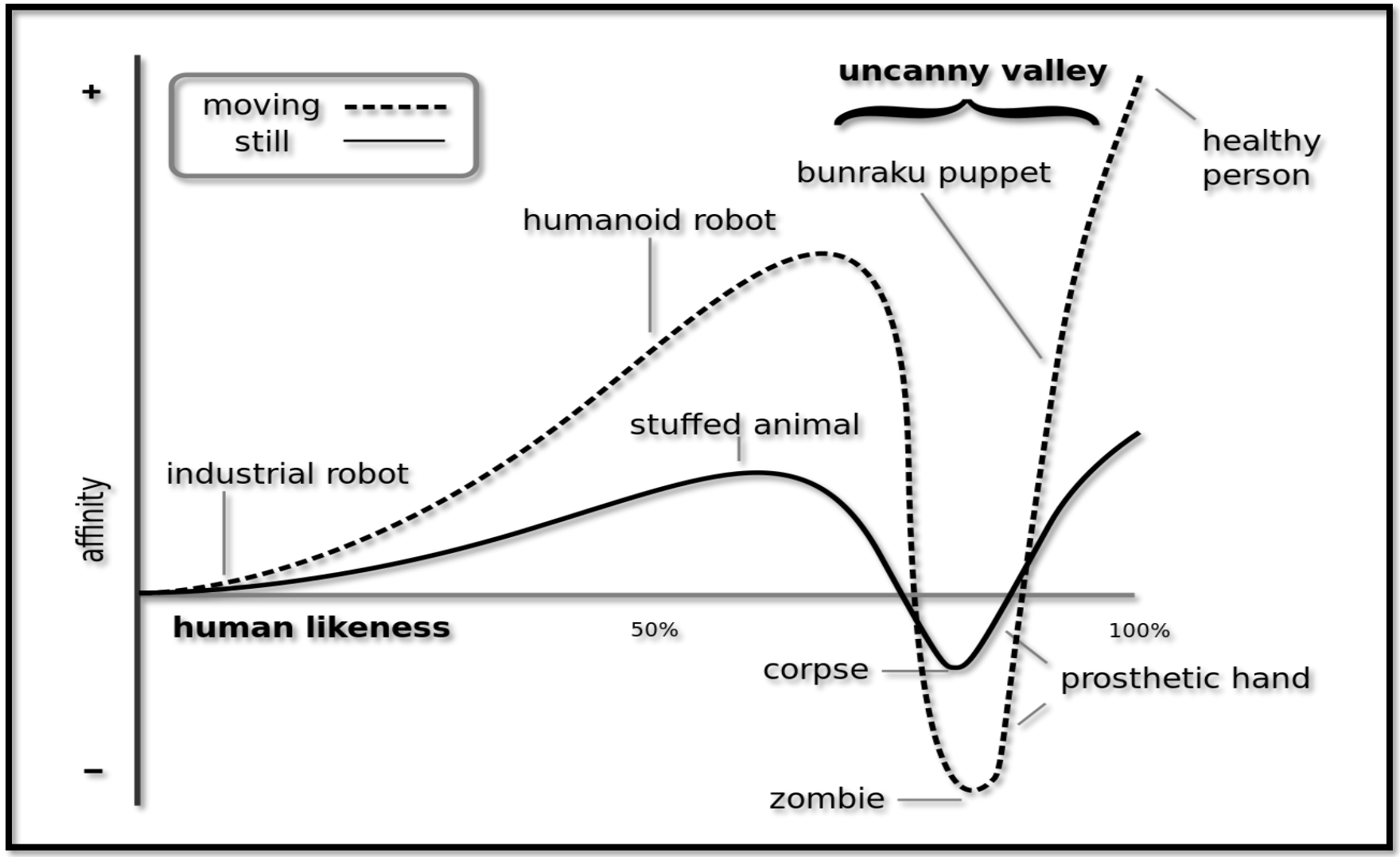
Identifying the Problem: How Do We Feel About AI?

For informational purposes only.
Does not constitute legal advice.

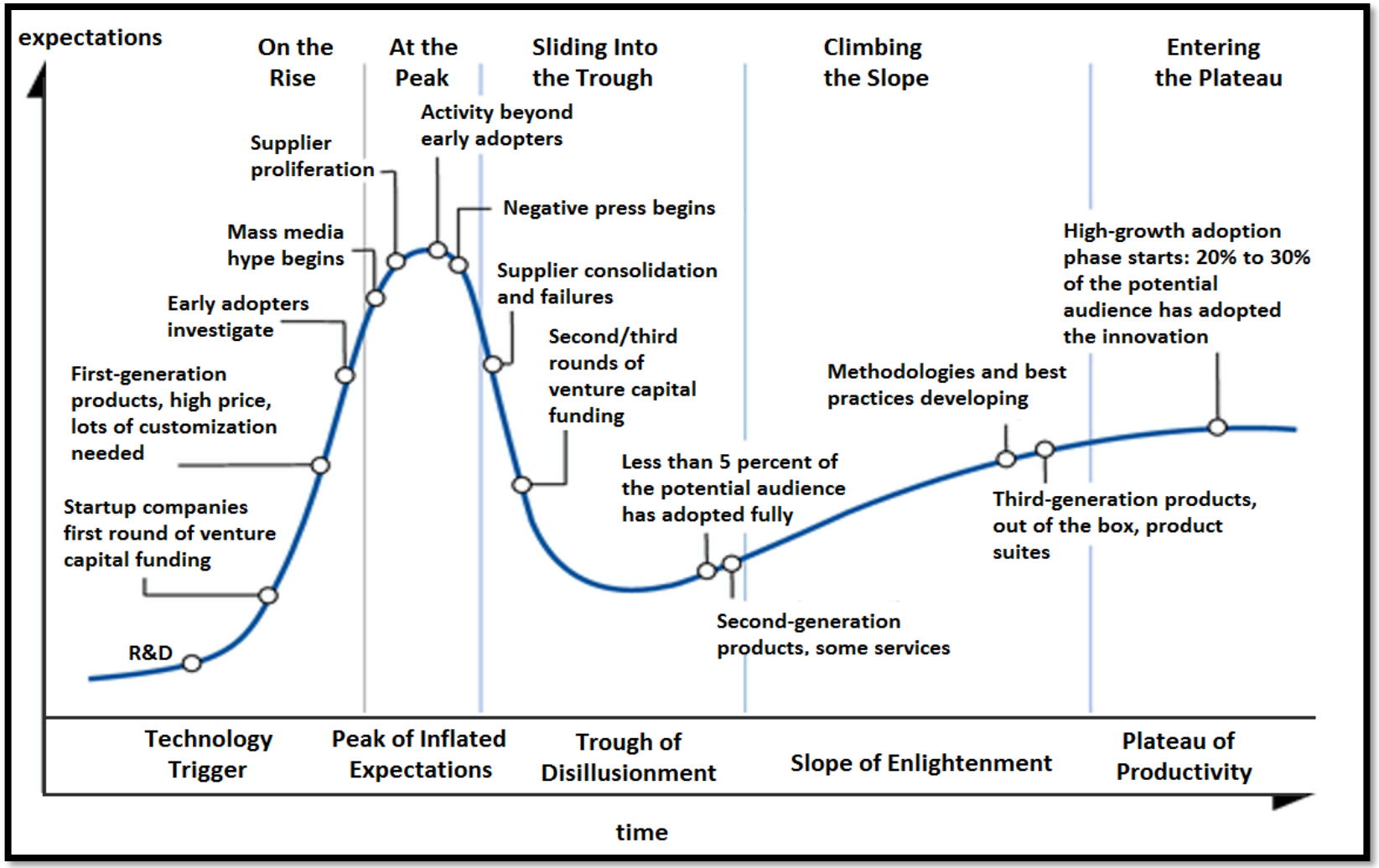
Why the Love/Hate Relationship?



The Uncanny Valley



The Uncanny Valley

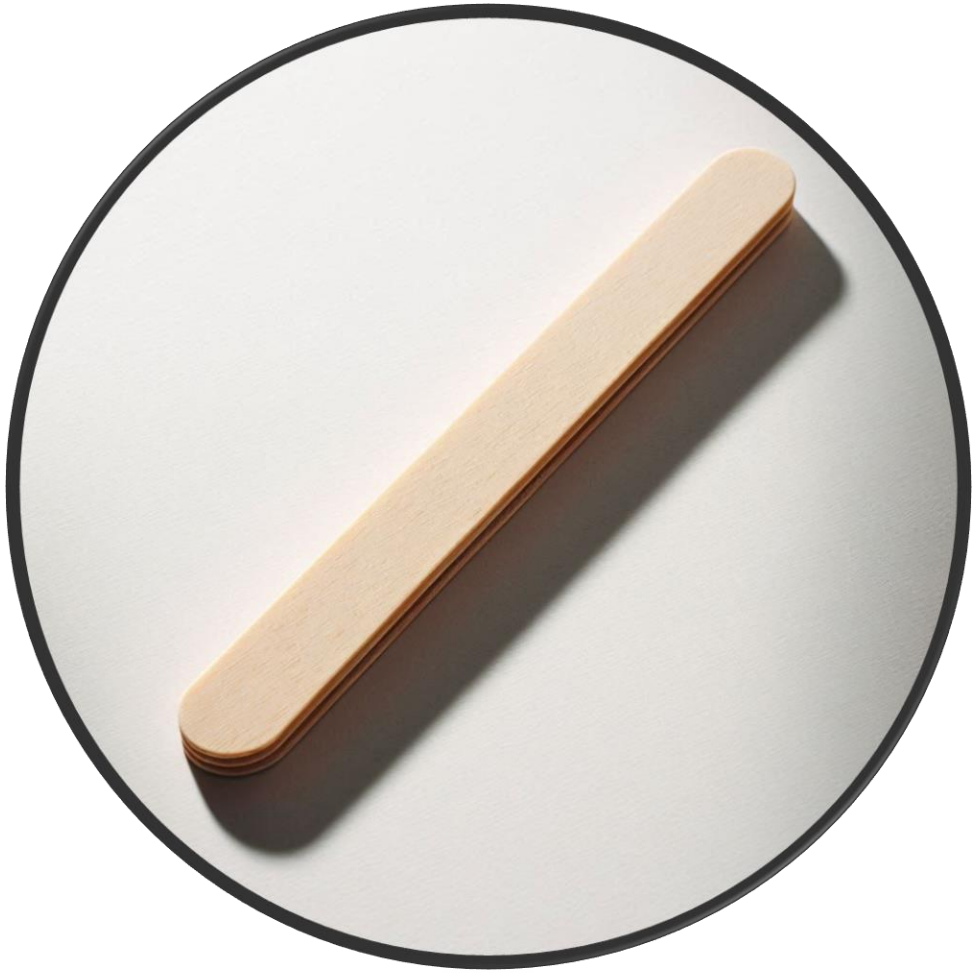


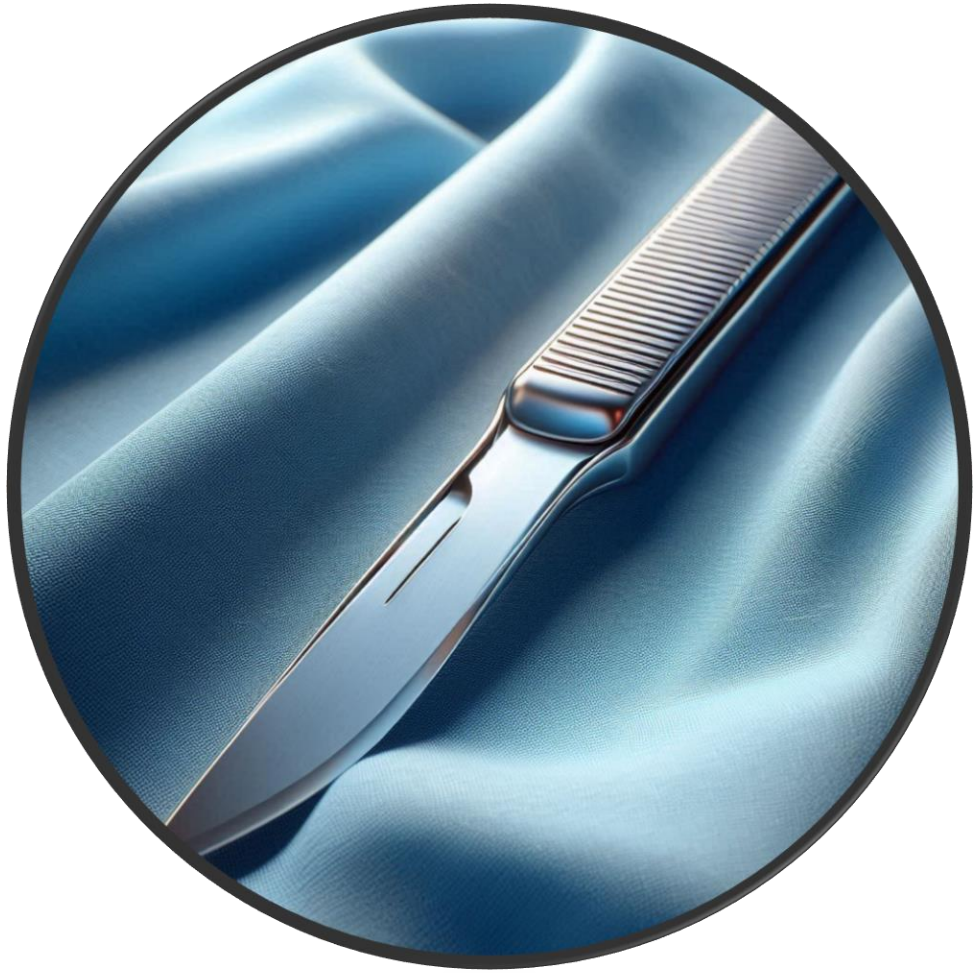
The Uncanny Valley

FOMO

FOBO

Tools Create Risk





How do we manage risk when intelligence is involved?

Ethics:

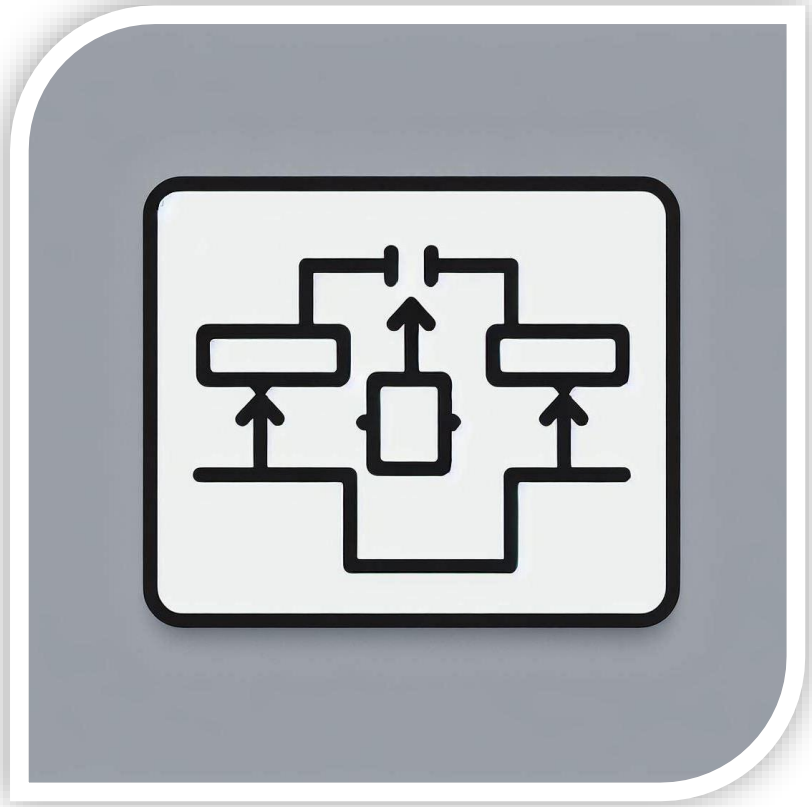
from ethos, that which is characteristic of the group

or

an accustomed place

How do we agree on AI ethics?

- Algorithm

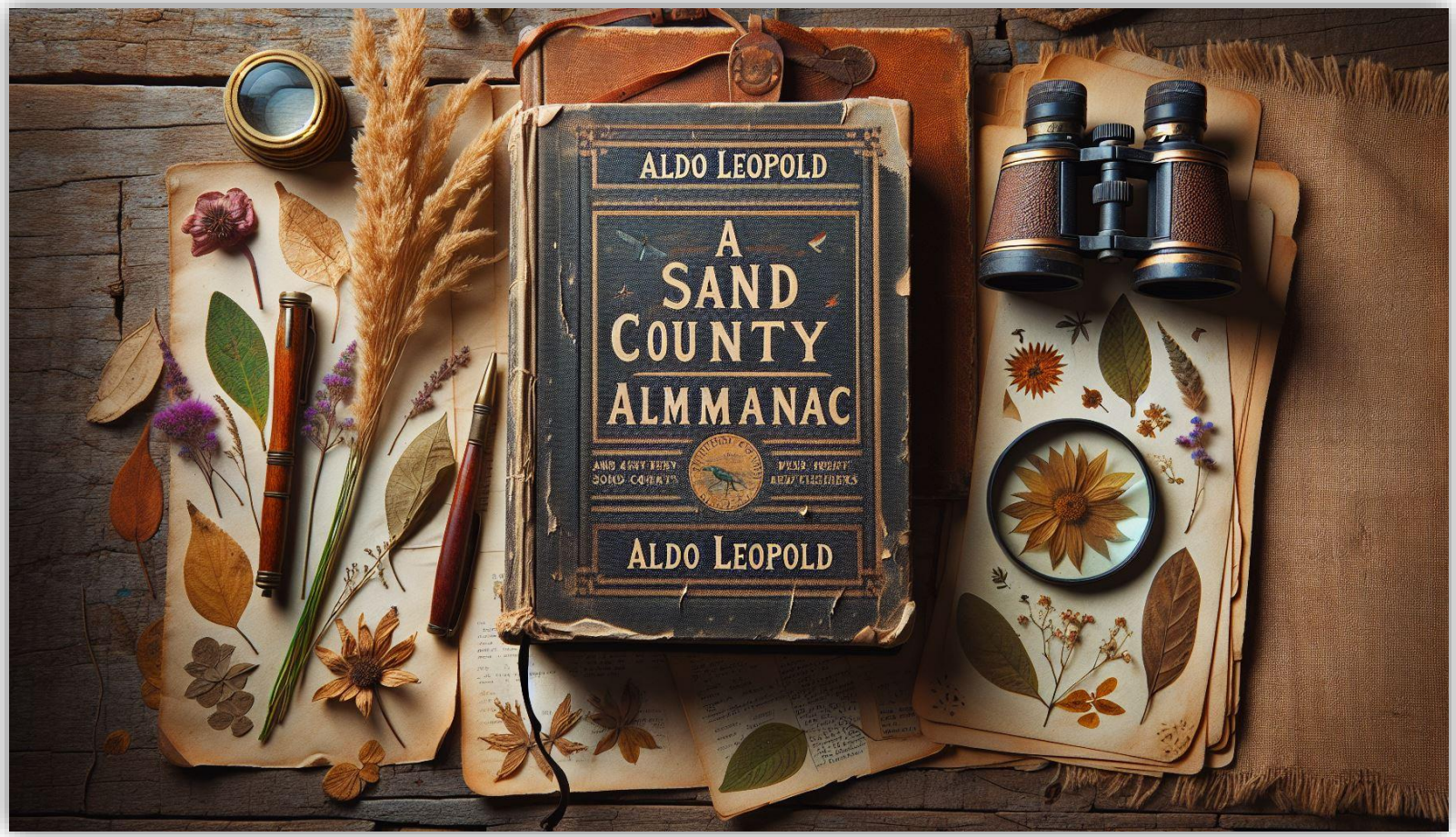


- Heuristic

How do we agree on AI Ethics?

- Algorithmic Risk, *e.g.*, risks arising from automated decision making
  - Automation bias
  - Biased inputs, biased outputs
  - Static, instantly obsolete
- Heuristic Risk, *e.g.*, risks from taking short cuts
  - Short cuts are hard to erase
  - The “unreasonable reliability of data”; using a “brilliantly stupid” tool
  - Getting the right answer for the wrong reason is not always right

How do we agree on AI Ethics?



How do we agree on AI Ethics?

Asimov's three laws of robotics:

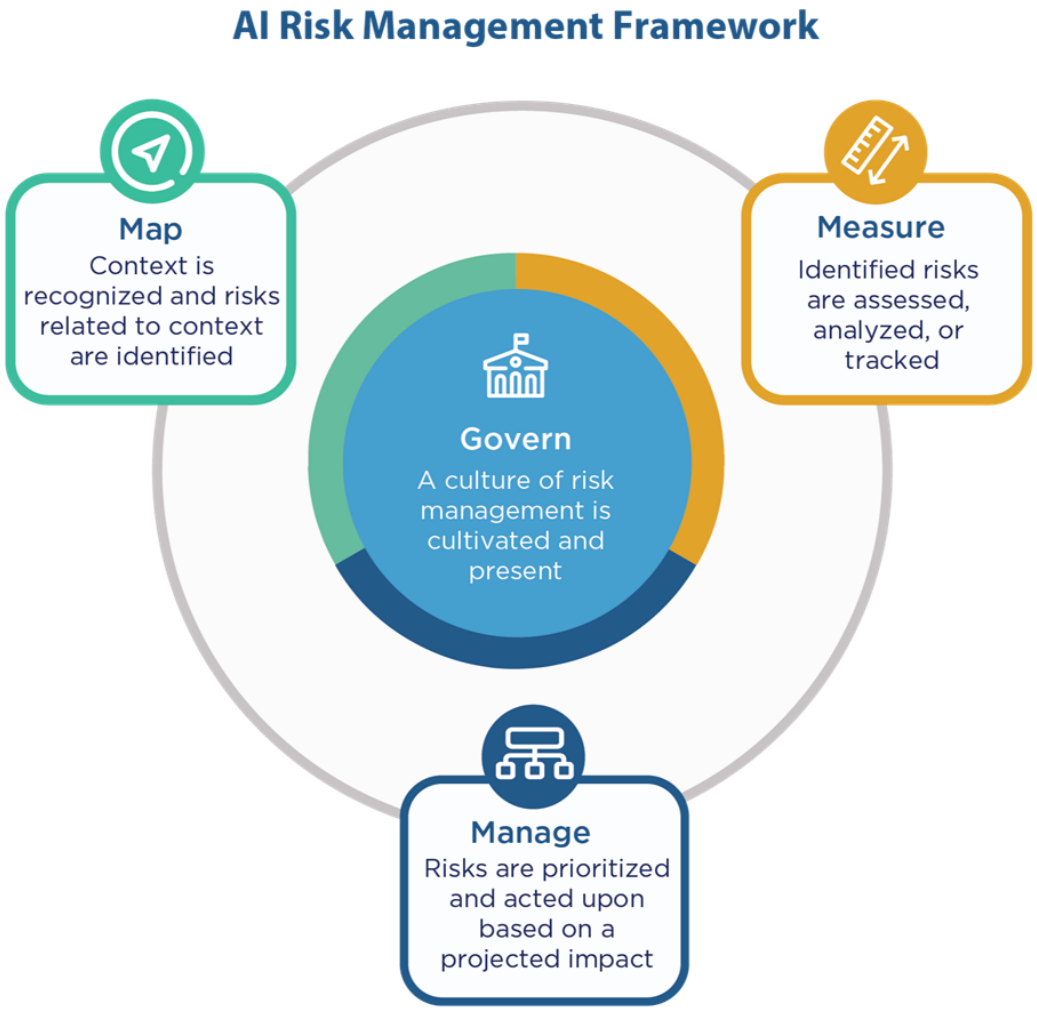
1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2. A robot must obey orders given it by human beings except where such orders would conflict with the First Law.

3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

How do we agree on AI ethics?

NIST AI RMF



For informational purposes only.
Does not constitute legal advice.

NIST AI RMF

NIST CSF	NIST AI RMF
• Identify • Protect • Detect • Respond • Recover • Govern	• Govern • Map • Measure • Manage

NIST AI RMF

- NIST is a federal agency, and the AI RMF is written from that perspective
  - Assumes critical mass to address AI risk, relevant SMEs, as well as time (and budget) to do the required work
  - 71 subcategories, 19 categories, 5 functions
    - To be executed in a non-sequential fashion
    - Repeated as necessary
    - Based on establishment of KPIs, as well as measurement and analysis of those KPIs

NIST AI RMF

- Adapt, adopt, improve?
  - None of us will have a textbook application of NIST AI RMF
  - The right application is the one that moves the needle for your organization in the right direction:
    - More visibility, more diligence, more accountability *i.e.*, documentation
  - Pick and choose, or start on a high-level first
    - If all you can address are the 19 Categories, you are probably off to a good start

NIST AI RMF

- Emulate
  - Watch this space: https://www.nist.gov/itl/ai-risk-management-framework
  - Most recent update, Gen AI RMF Profile: https://airc.nist.gov/docs/NIST.AI.600-1.GenAI-Profile.ipd.pdf
  - Implement a maturity model: https://ieeeusa.org/product/a-flexible-maturity-model-for-ai-governance/
  - Watch others in your industry
  - Ask vendors for their RMF compliance documentation

GDPR Privacy Principles

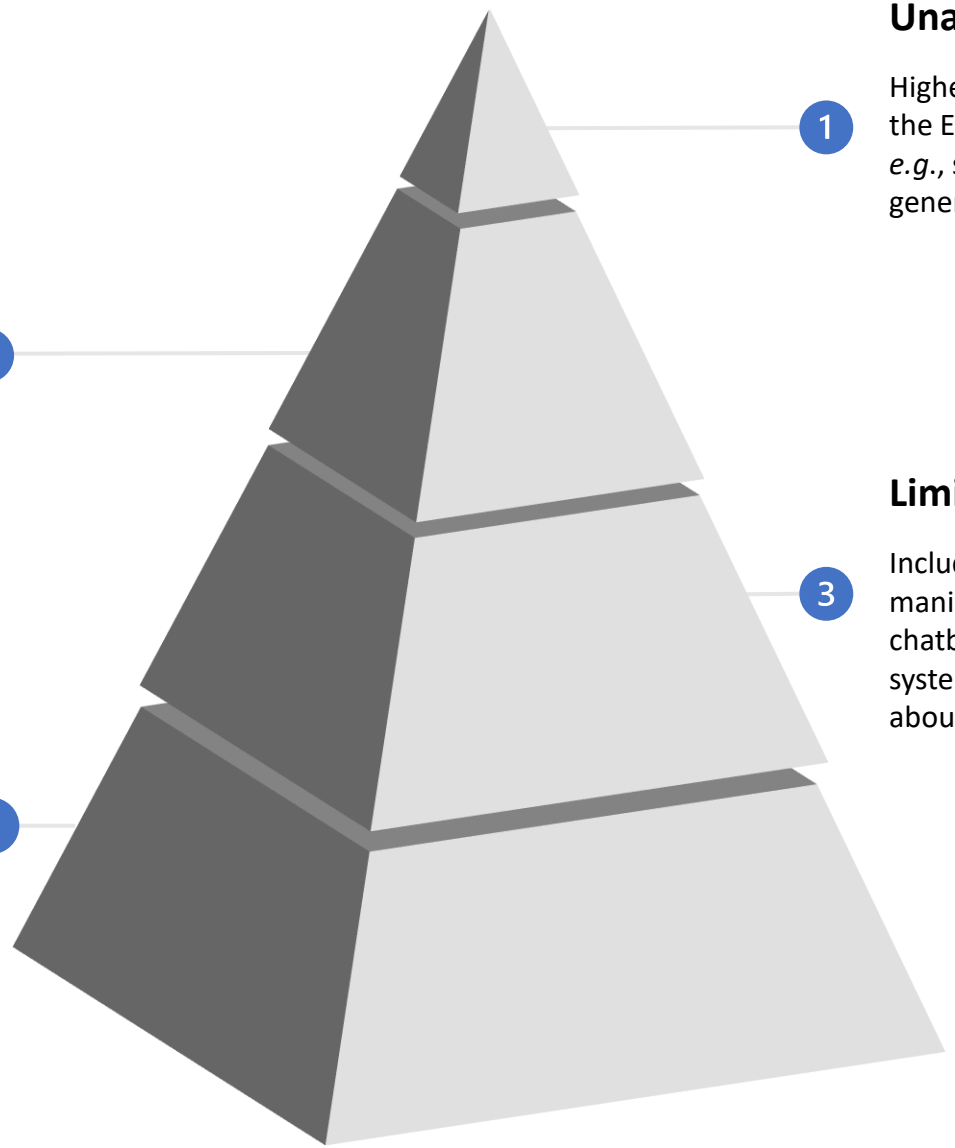
- Lawfulness, fairness, transparency
- Process limitation
- Data minimization
- Accuracy
- Storage limitation
- Integrity and Confidentiality (and don’t forget Security)
- Accountability

GDPR Privacy Principles

DPIA

AIIA

EU AI Act



Unacceptable Risk

Highest level of risk prohibited in the EU. Includes AI systems using *e.g.*, subliminal manipulation or general social scoring.

High Risk

Most regulated AI systems, as these have the potential to cause significant harm if they fail or are misused, *e.g.*, if used in law enforcement or recruiting.

Limited Risk

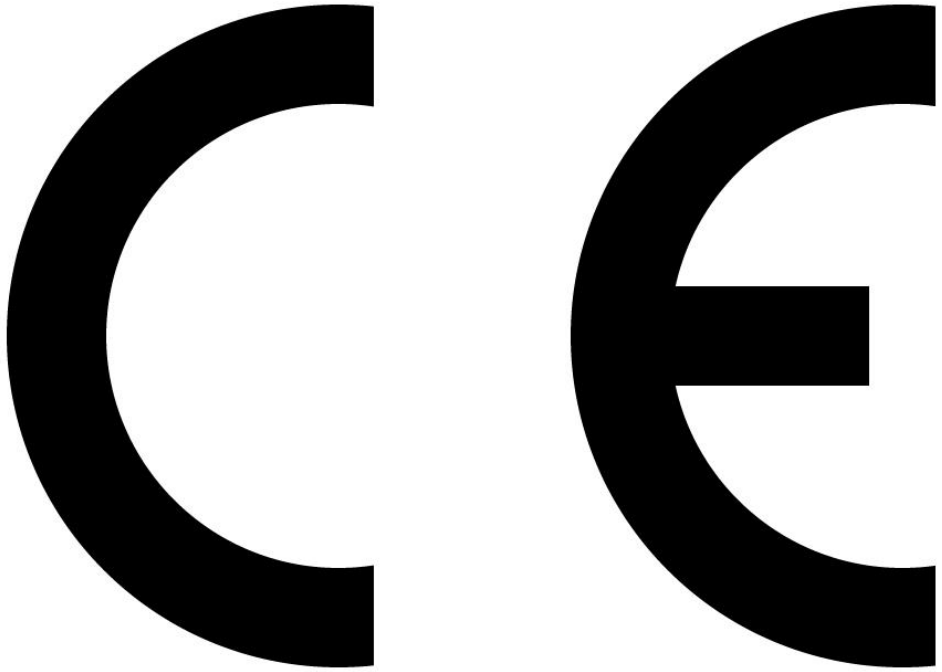
Includes AI systems with a risk of manipulation or deceit, *e.g.*, chatbots or emotion recognition systems. Humans must be informed about their interaction with the AI.

Minimal Risk

All other AI systems, *e.g.*, a spam filter, which can be deployed without additional restrictions.

For informational purposes only.
Does not constitute legal advice.

EU AI Act



What’s already in your toolbox?

- Data Classification; Record of Processing Activities
- Risk Assessment/Data Protection Impact Assessments
- Policy/Procedure Development
- Vulnerability Assessments/Red Teaming
- Incident Response Planning
- Table Top Exercises
- Third Party/Supply Chain Risk Management
- Codes of Conduct

Case study: facial recognition

- Commercially available solution
- Pilot at one location
  - Scope defined
  - Volunteers sourced
  - Disclosures to pilot team
  - KPIs defined
  - Signage
  - Pilot launched
- Not anticipated: brand impact, news coverage, regulatory attention
- Best defense: the accountability principle

Case study: facial recognition

- Do you need consent?
  - GDPR: lawful basis always required
    - Can be consent or legitimate interest, among other things
  - In the US, biometrics can constitute “sensitive personal data”
    - Consent required for processing of SPD (VT, CT, NJ, MD, DE)
    - Otherwise, right to restrict processing (CA)
  - How can you balance consent with a need to catch the bad guys?
    - Limited caveats under each law (with the organization bearing the burden of proof):

> Nothing in this chapter shall be construed to restrict a controller's or processor's ability to . . . [p]revent, detect, protect against, or respond to security incidents, identity theft, fraud, harassment, malicious or deceptive activities, or any illegal activity; preserve the integrity or security of systems; or investigate, report, or prosecute those responsible for any such action

Where to start

- Build a diverse team
- Bring your ethical toolkit with you – find the externalities and manage them
- Map, map, and then map again
- Start small, *e.g.*, GenAI usage
- Realize the limitations of a generalized framework
- Make the framework your own
- Benchmark, and look for guidance
- Ask dumb questions, and watch out for the hype cycle
- Repeat!

Cautionary Tales

- Don’t rely on the AI platform to tell you if it’s reliable

  Mata v. Avianca, Inc.

- Don’t use AI alone to draft your policies

  In re bitFlyer

- Don’t assume that bias is not an issue

  Big Data, A Tool for Inclusion or Exclusion

The technology ethic

Ethics = authenticity (aka Don't be creepy!)

- In the end, perception of fairness and ethics comes down to what is familiar, but not artificially so
- Authenticity will matter: human involvement necessary
- Trust but verify: ignoring AI is impossible, but jumping in blindly is exactly that: jumping in blindly
- Hype ≠ authenticity

THANK YOU!

www.linkedin.com/in/fgreene

For informational purposes only.
Does not constitute legal advice.

Thank you!

Have thoughts about SIA Education@ISC?

Scan the QR Code on the left to provide your feedback on SIA Education@ISC Sessions at ISC West

In partnership with SIA EDUCATION at ISC

Built by RX In the business of building businesses