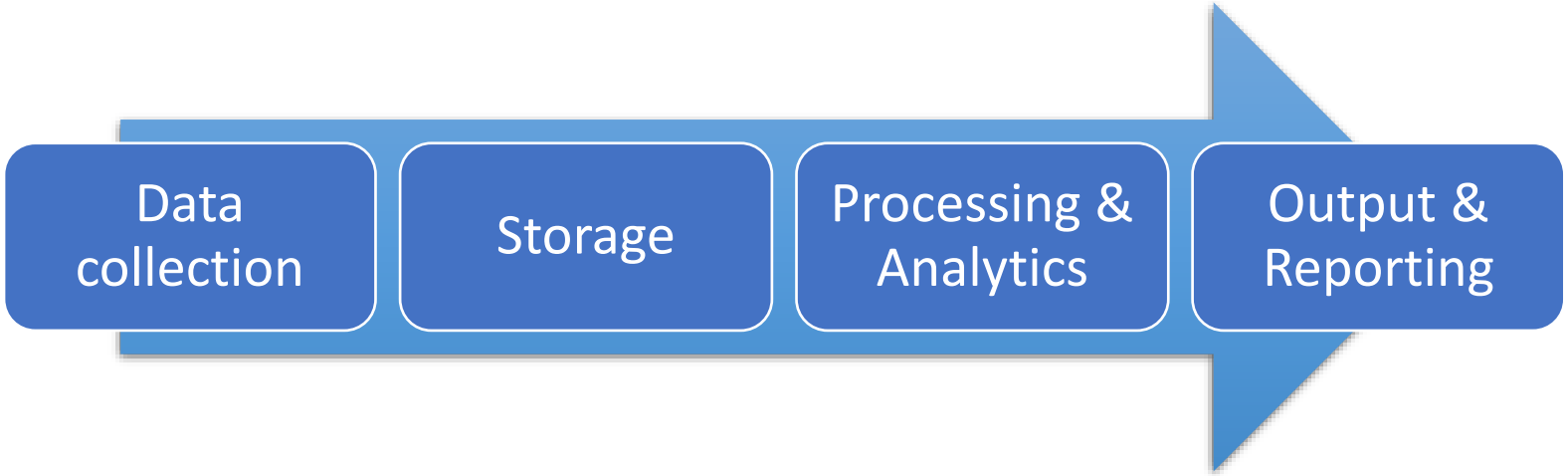
How Security Teams Can Help Build An AI Program

Mark Francis
Partner, New York | Data Strategy, Security & Privacy | Holland & Knight LLP

The Explosion of AI Use

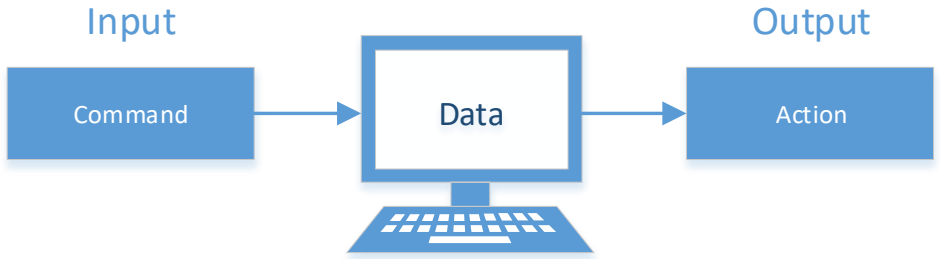
- AI as we are currently using it represents the next generation of big data analytics



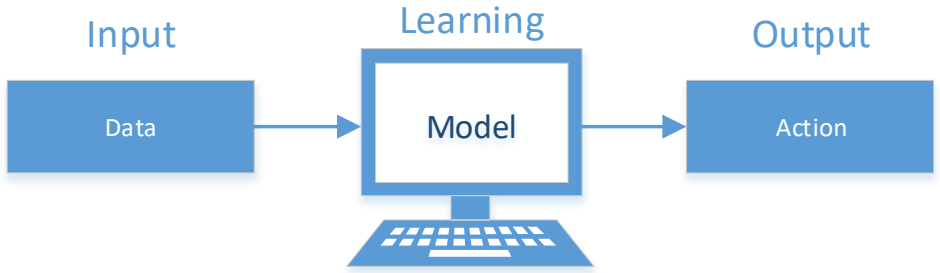
- Big data analytics relies on two key components:
  - Accumulation of big data sets
    +
  - Availability of cheap computing power

Transition to AI

- How does AI change the approach?
  - Switch from ***command***-driven analytics…
    - Need to know desired output
    - Need to write commands for desired output



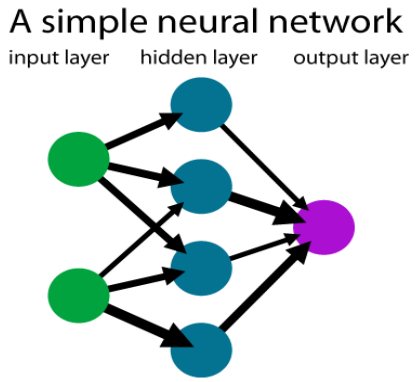
  - …to ***model***-driven analytics
    - Machine determines the output
    - Machine develops model to achieve that output



Advanced AI/ML

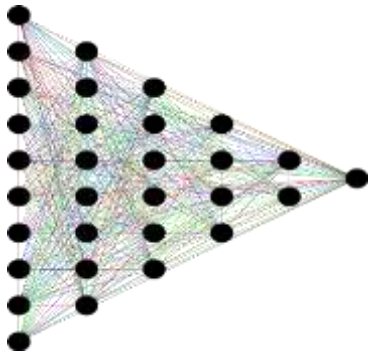
Artificial "Neural Networks"

- Designed to function more like the human brain where interconnected "neurons" can perform discrete data-related tasks, such as recognizing something, creating associations between information, or evaluating a relationship; neurons can adapt over time

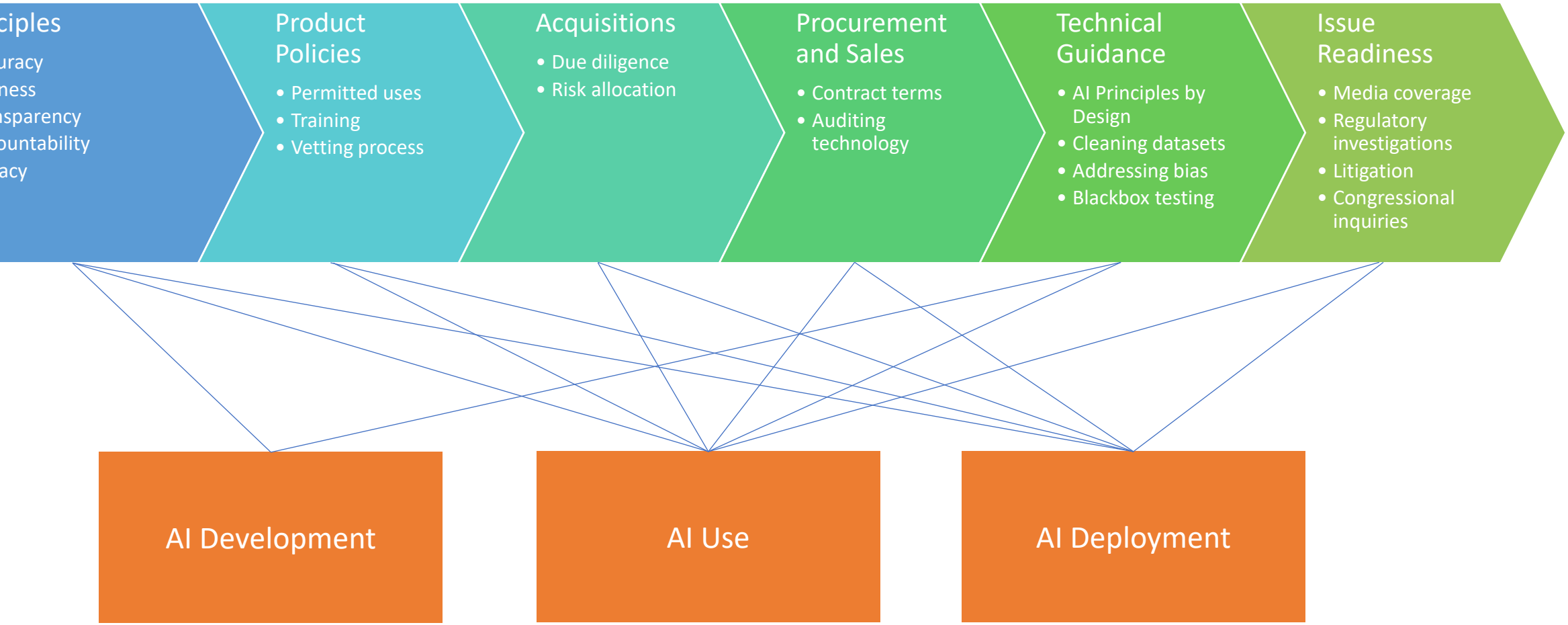


"Deep Learning"

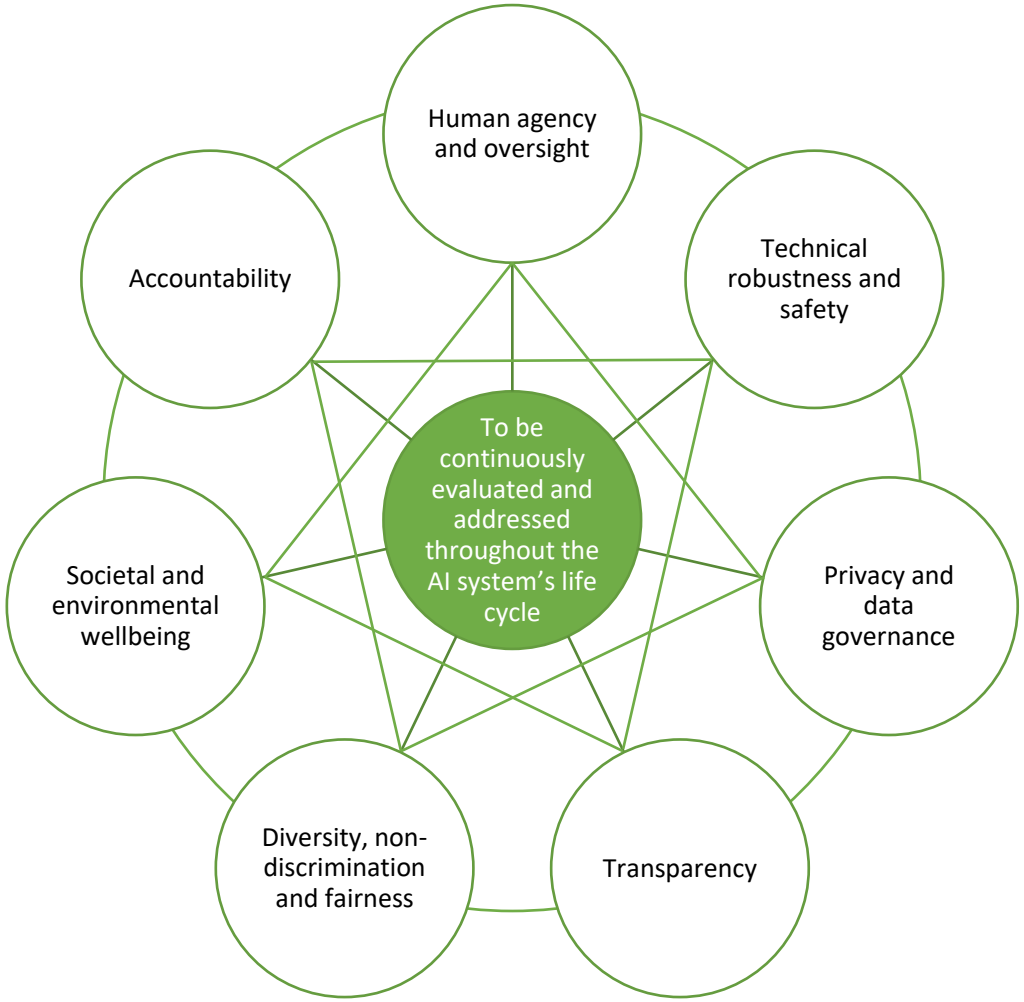
- Multiple layers of neural networks, with more algorithms applied to perform more complex simulations – these tools can learn from mistakes and, over time, are able to produce results with increasing accuracy and precision



How You Can Help Build An AI Program



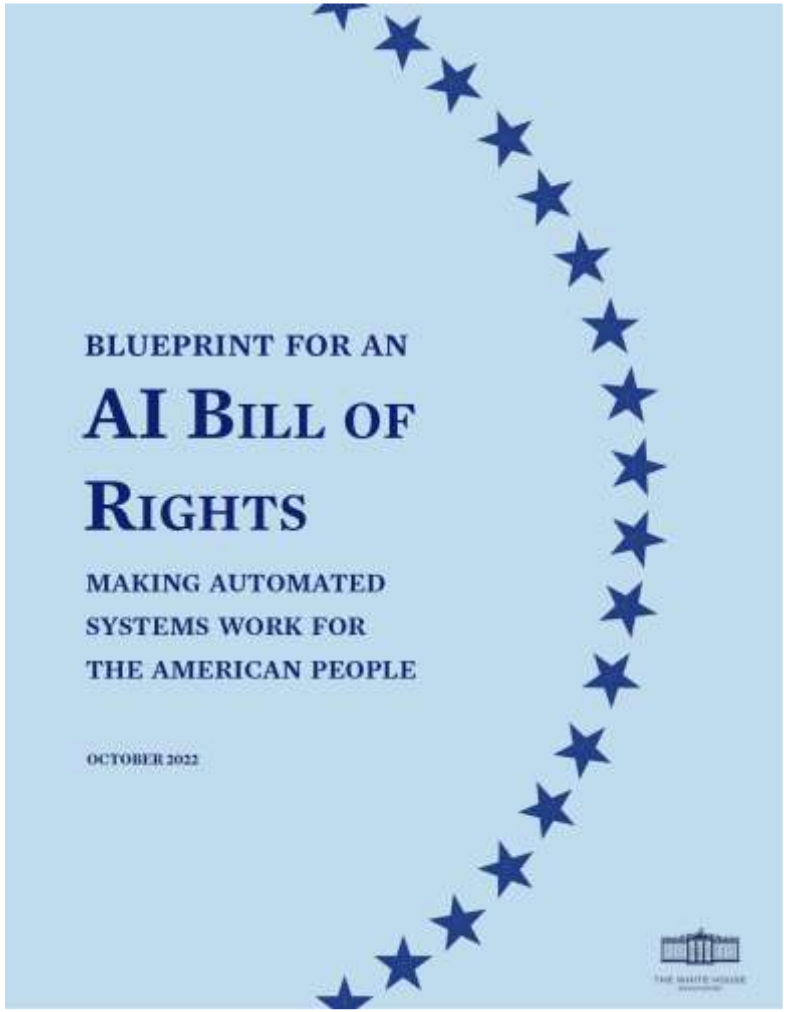
Development of Industry Standards



E.C. AI Ethics Guidelines
Figure 2

Interrelationship of the seven requirements: all are of equal importance, support each other, and should be implemented and evaluated throughout the AI system's lifecycle

Development of Industry Standards



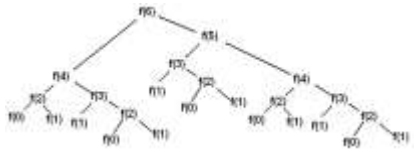
- Safe and Effective Systems
- Algorithmic Discrimination Protections
- Data Privacy
- Notice and Explanation
- Human Alternatives, Consideration, and Fallback

AI Terminology

Data feed
(Training dataset)

Transformed data
(Scrubbed data)

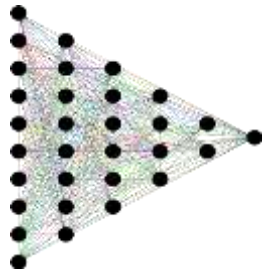
Untrained Model
(Algorithms)



Development

SaaS / AI Engine

Trained model

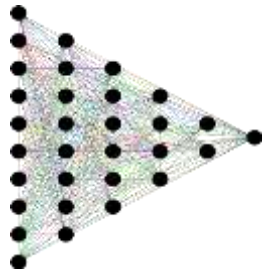


Derived Data
(Output)

Deployment

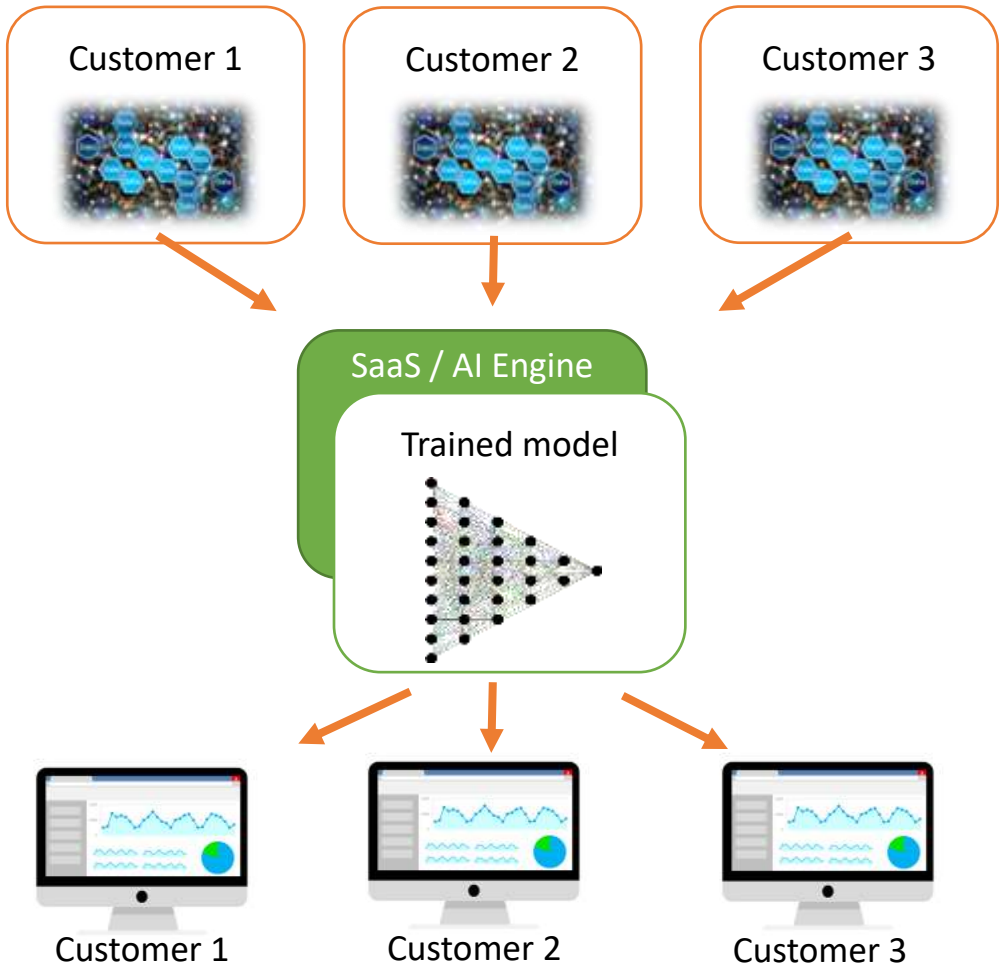
Training

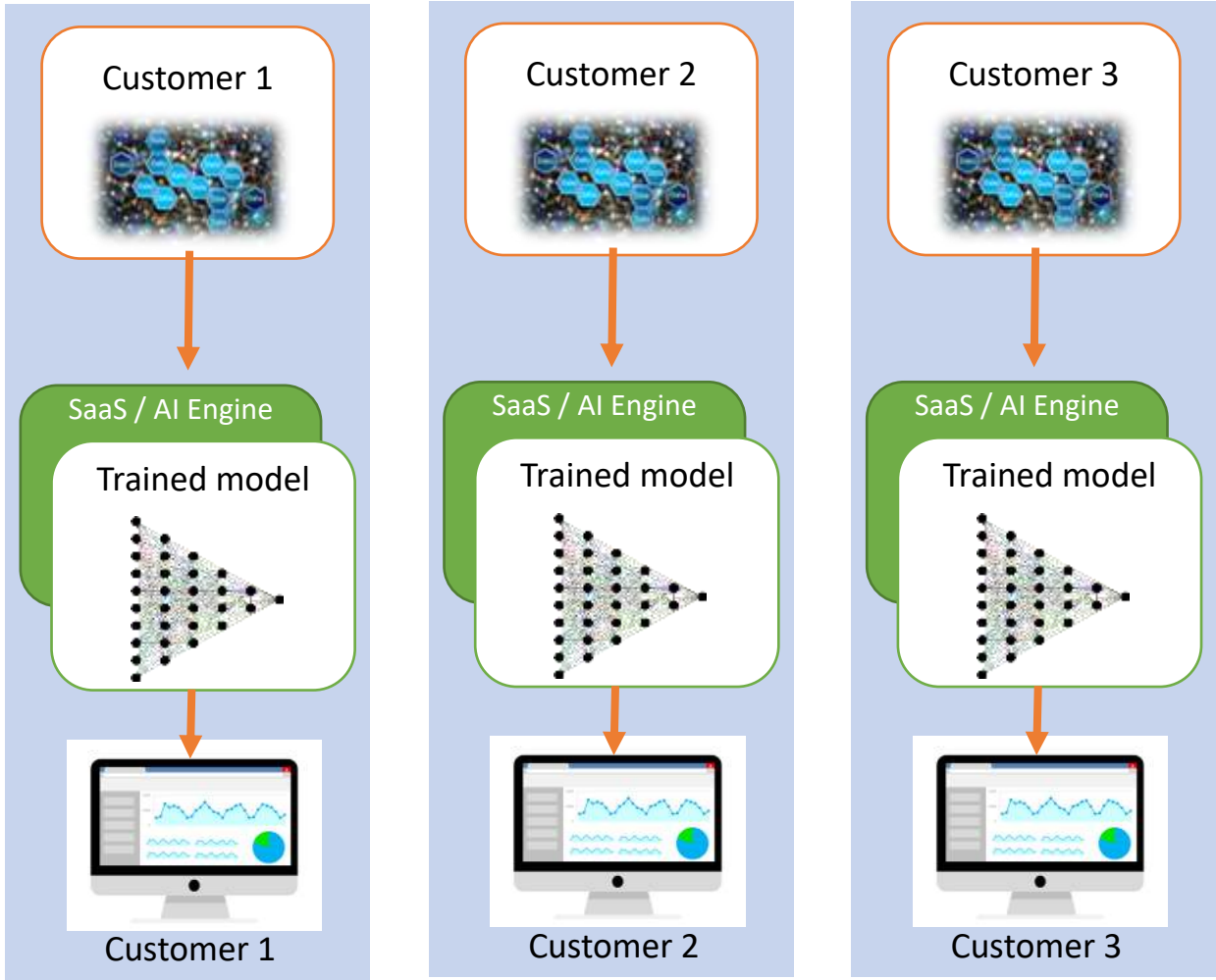
Trained model



Dev Environment

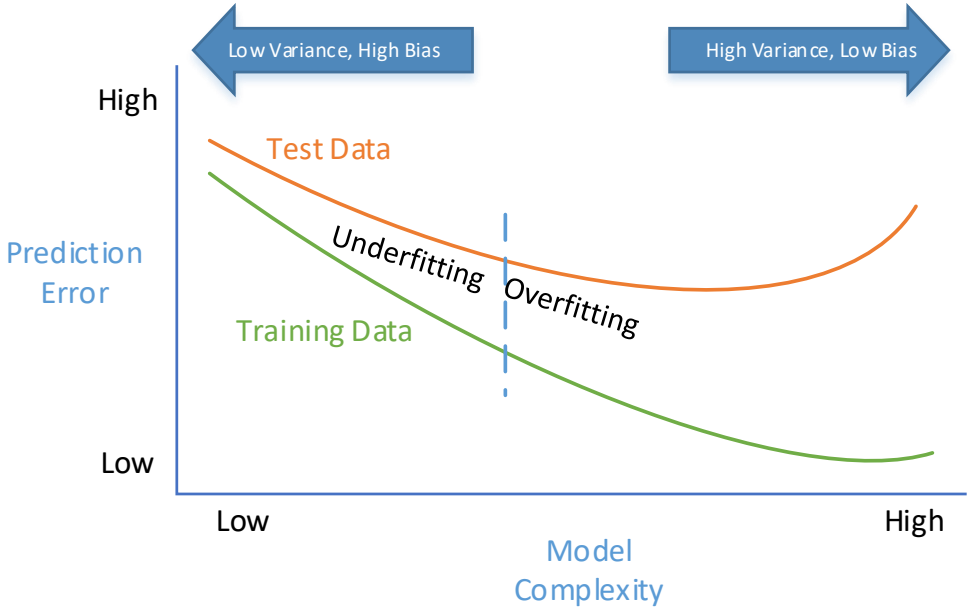
Different customer paradigms



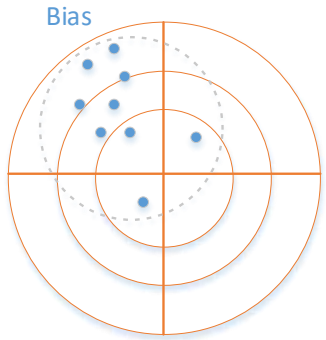


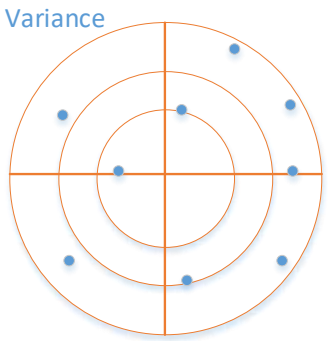
Understanding AI Challenges

Technical Bias refers to a gap between a predicted value and actual value—such as where errors tend to skew in a certain direction



Variance refers to how concentrated or scattered the predicted values are





Legal Bias refers to a decision that discriminates based on association with a legally-protected class (race, religion, gender, age, sexual preference)

When an AI uses data associated with a protected class (or reflective of a protected class), it can produce outcomes that may be:

- ***Programmatically* right**
- **But *legally* wrong**

Current Legal Framework

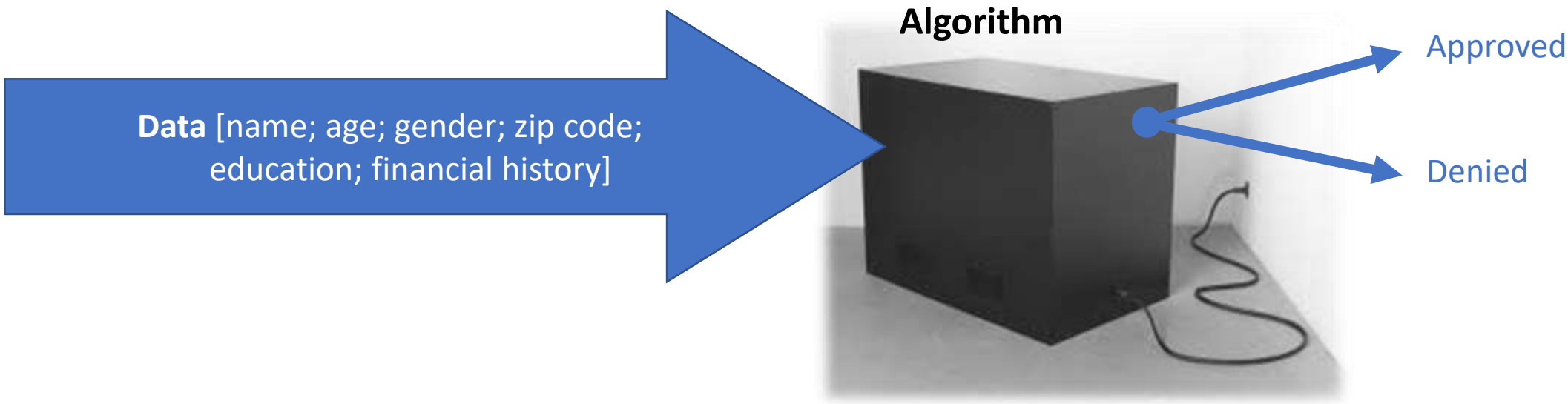
- How the law governs evolving AI technology is still unsettled

- Most experts believe that laws already in place for human activity that AI replaces can equally apply to developing technologies, for example:
  - Labor laws (AI as labor replacement)
  - Copyright and intellectual property
  - Data privacy and cybersecurity
  - Consumer protection and non-discrimination
  - Product liability / strict liability
  - Tort law (negligence and malpractice)
  - Contractual obligations (government, commercial)
  - CFPB, OCC, FTC and AG regulatory pronouncements

- 100s of bills addressing AI introduced in Congress (and more at the state level), but legislators and regulators are reluctant to act too fast or adversely impact beneficial development or use of AI

Issue Spotting: Risk of Legal Bias

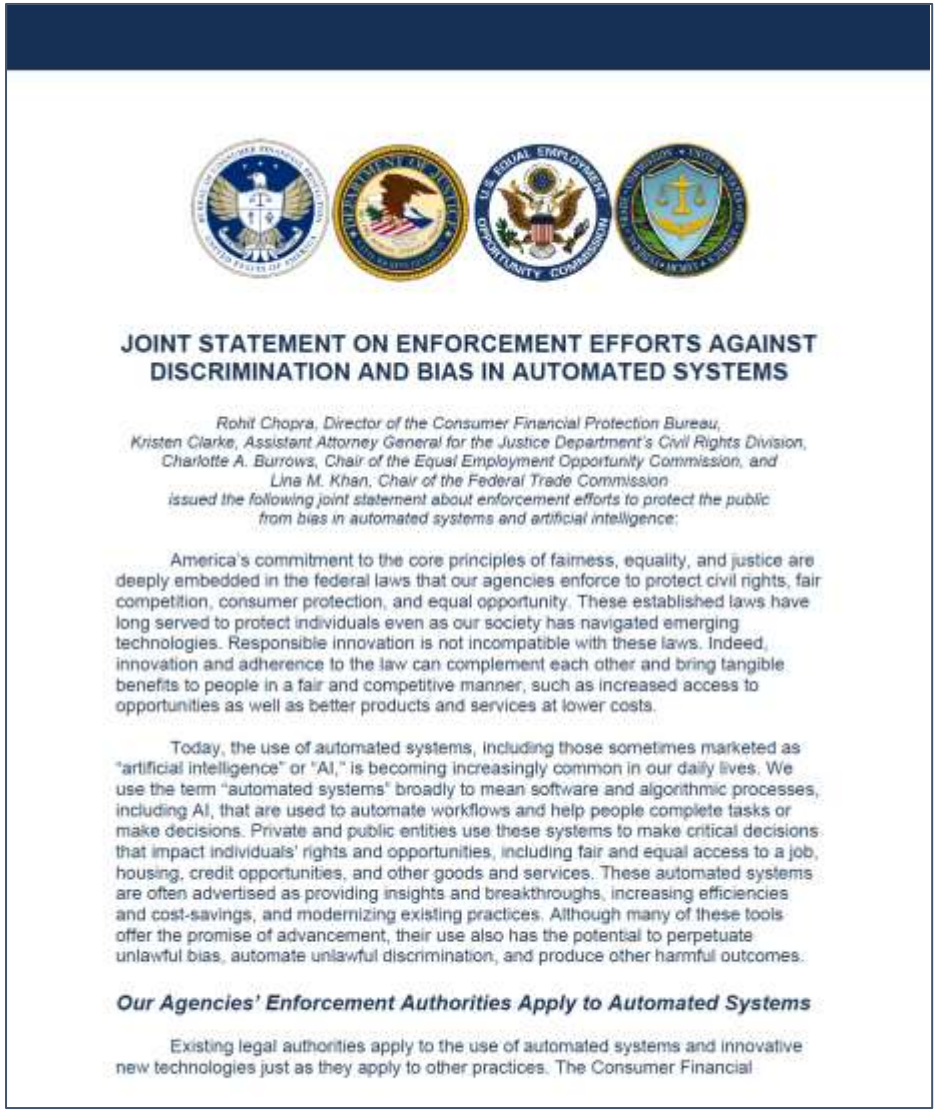
- "Digital Redlining" – disparate impact based on decision making that is biased (creates or perpetuates inequality)
  - *E.g., HUD Charges Facebook With Housing Discrimination Over Company's Targeted Advertising Practices (March 28, 2019).*



Regulatory Interest and Guidance

"Existing legal authorities apply to the use of automated systems and innovative new technologies just as they apply to other practices. The Consumer Financial Protection Bureau, the Department of Justice's Civil Rights Division, the Equal Employment Opportunity Commission, and the Federal Trade Commission are among the federal agencies responsible for enforcing civil rights, non-discrimination, fair competition, consumer protection, and other vitally important legal protections. We take seriously our responsibility to ensure that these rapidly evolving automated systems are developed and used in a manner consistent with federal laws, and each of our agencies has previously expressed concern about potentially harmful uses of automated systems."

https://www.consumerfinance.gov/about-us/newsroom/cfpb-federal-partners-confirm-automated-systems-advanced-technology-not-an-excuse-for-lawbreaking-behavior/



JOINT STATEMENT ON ENFORCEMENT EFFORTS AGAINST DISCRIMINATION AND BIAS IN AUTOMATED SYSTEMS

Rohit Chopra, Director of the Consumer Financial Protection Bureau,
Kristen Clarke, Assistant Attorney General for the Justice Department's Civil Rights Division,
Charlotte A. Burrows, Chair of the Equal Employment Opportunity Commission, and
Lina M. Khan, Chair of the Federal Trade Commission
issued the following joint statement about enforcement efforts to protect the public from bias in automated systems and artificial intelligence:

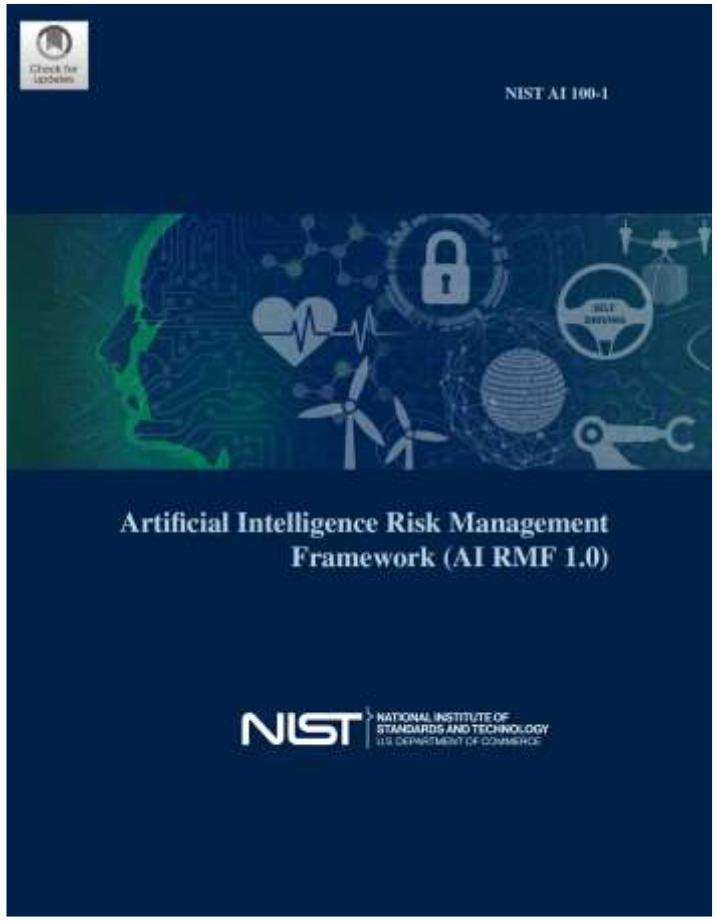
America's commitment to the core principles of fairness, equality, and justice are deeply embedded in the federal laws that our agencies enforce to protect civil rights, fair competition, consumer protection, and equal opportunity. These established laws have long served to protect individuals even as our society has navigated emerging technologies. Responsible innovation is not incompatible with these laws. Indeed, innovation and adherence to the law can complement each other and bring tangible benefits to people in a fair and competitive manner, such as increased access to opportunities as well as better products and services at lower costs.

Today, the use of automated systems, including those sometimes marketed as "artificial intelligence" or "AI," is becoming increasingly common in our daily lives. We use the term "automated systems" broadly to mean software and algorithmic processes, including AI, that are used to automate workflows and help people complete tasks or make decisions. Private and public entities use these systems to make critical decisions that impact individuals' rights and opportunities, including fair and equal access to a job, housing, credit opportunities, and other goods and services. These automated systems are often advertised as providing insights and breakthroughs, increasing efficiencies and cost-savings, and modernizing existing practices. Although many of these tools offer the promise of advancement, their use also has the potential to perpetuate unlawful bias, automate unlawful discrimination, and produce other harmful outcomes.

Our Agencies' Enforcement Authorities Apply to Automated Systems

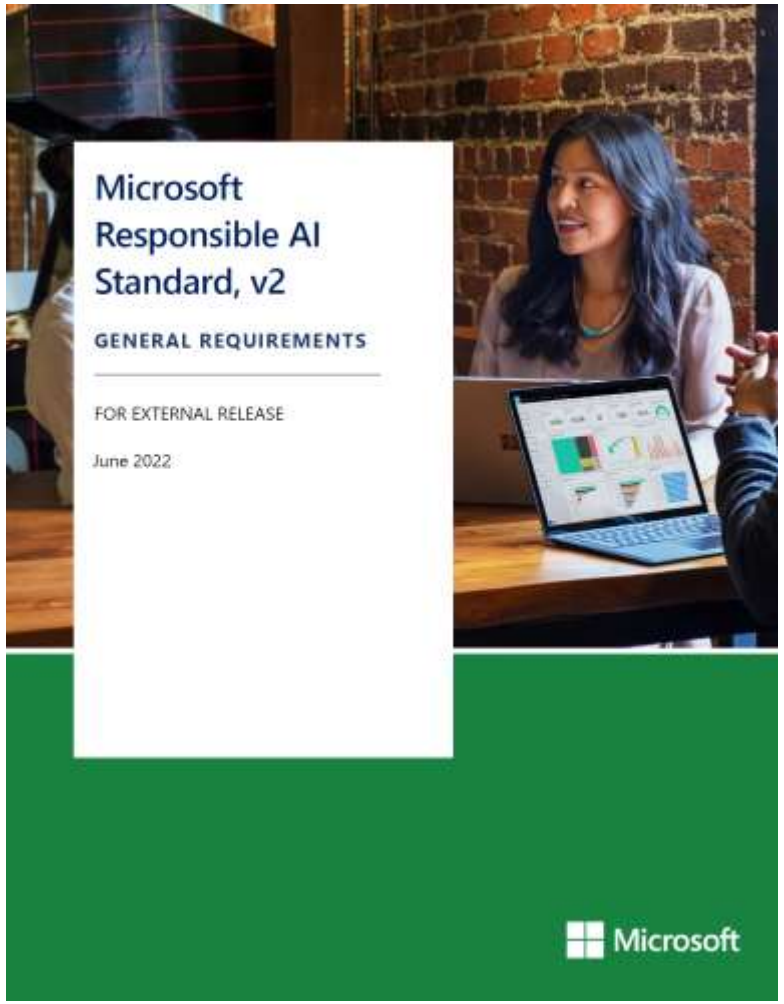
Existing legal authorities apply to the use of automated systems and innovative new technologies just as they apply to other practices. The Consumer Financial

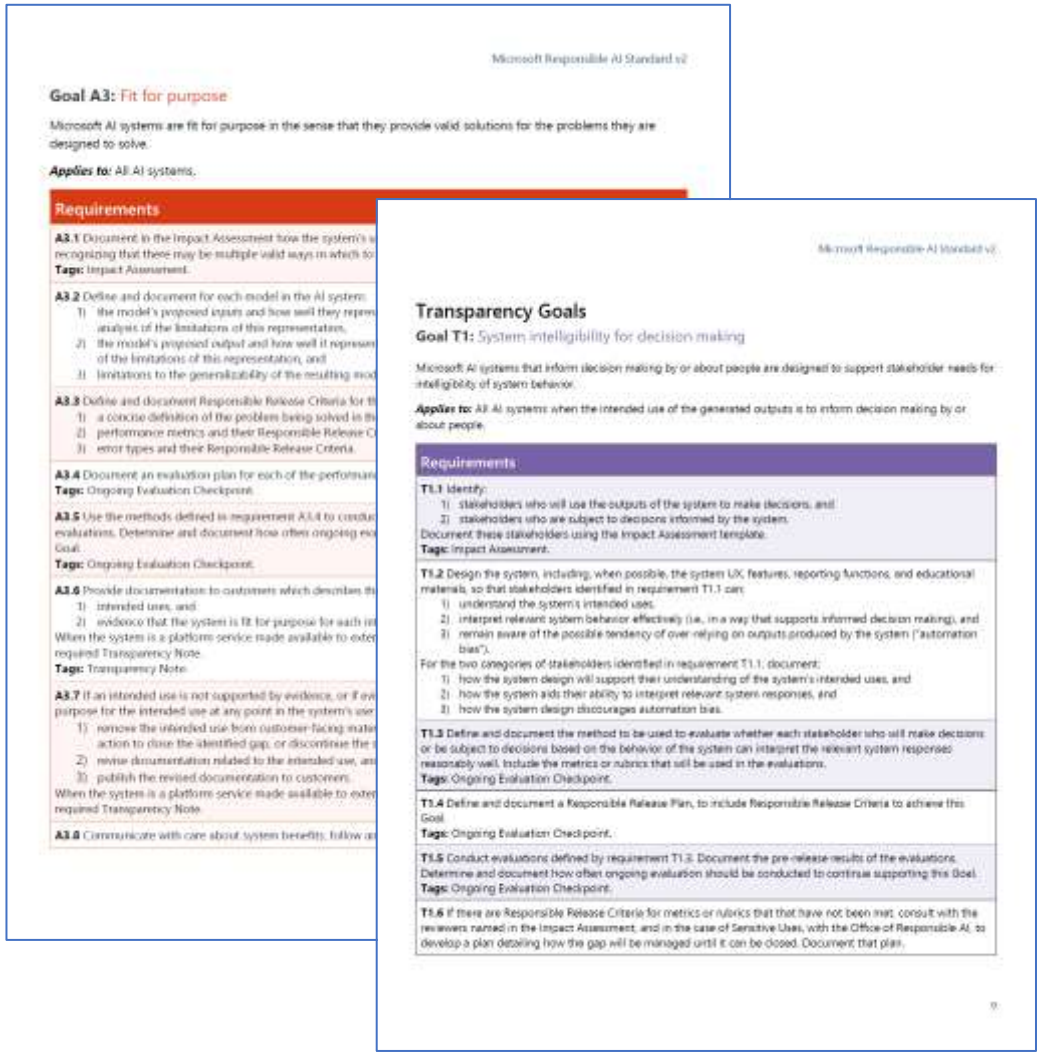
Development of Industry Standards



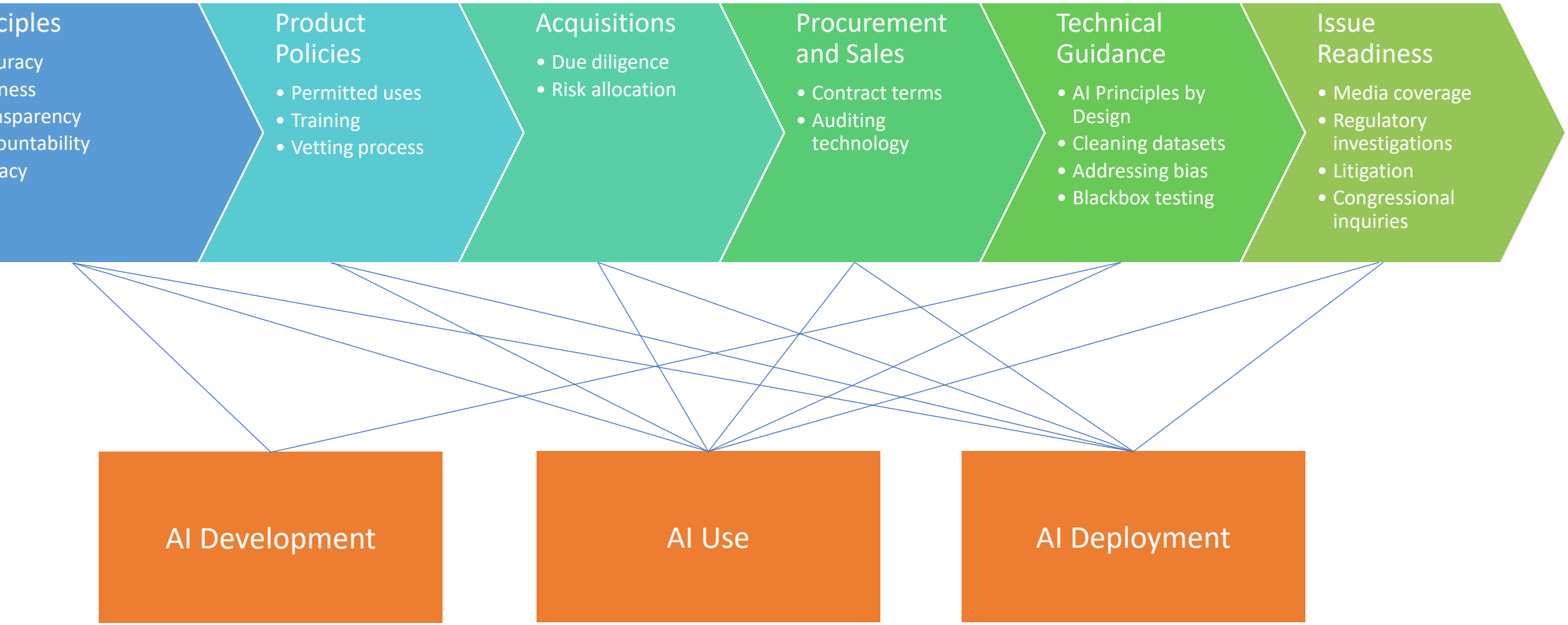
Key Dimensions	Application Context	Data & Input	AI Model	AI Model	Task & Output	Application Context	People & Planet
Lifecycle Stage	Plan and Design	Collect and Process Data	Build and Use Model	Verify and Validate	Deploy and Use	Operate and Monitor	Use or Impacted by
TEVV	TEVV includes audit & impact assessment	TEVV includes internal & external validation	TEVV includes model testing	TEVV includes model testing	TEVV includes integration, compliance testing & validation	TEVV includes audit & impact assessment	TEVV includes audit & impact assessment
Activities	Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations.	Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations.	Create or select algorithms; train models.	Verify & validate, calibrate, and interpret model output.	Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience.	Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations.	Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights.
Representative Actors	System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators.	Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts.	Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts.		System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts,	System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impact-ed individuals/commu-nities; evaluators.	End users, operators, and practitioners; impacted individu-als/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers.

Development of Industry Standards





How You Can Help Build An AI Program



Adopt a set of suitable AI principles

Principles
- Accuracy
- Fairness
- Transparency
- Accountability
- Privacy

Product Policies
- Permitted uses
- Training
- Vetting process

Acquisitions
- Due diligence
- Risk allocation

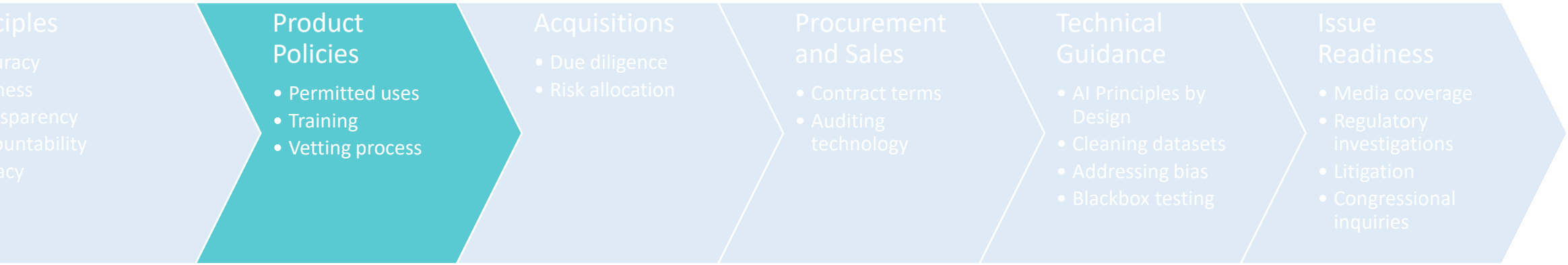
Procurement and Sales
- Contract terms
- Auditing technology

Technical Guidance
- AI Principles by Design
- Cleaning datasets
- Addressing bias
- Blackbox testing

Issue Readiness
- Media coverage
- Regulatory investigations
- Litigation
- Congressional inquiries

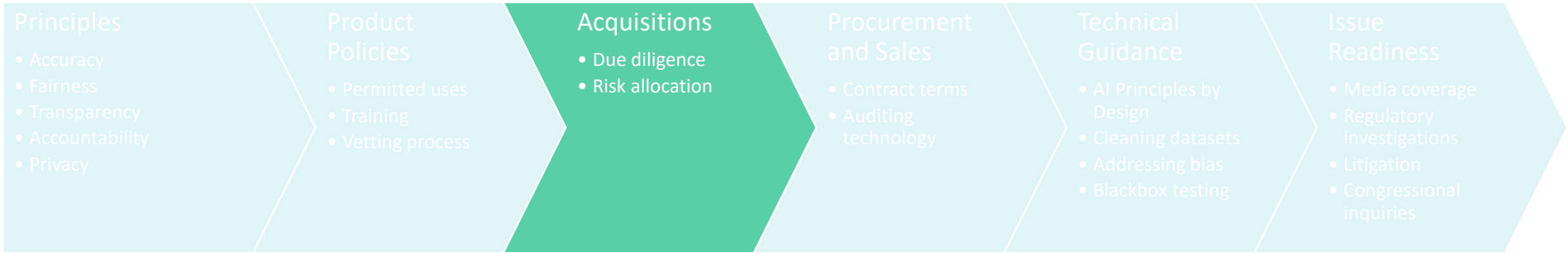
- ❑ Identify corporate goals for the use of AI (develop, use, sell)
- ❑ Evaluate the potential impact of such use (risks and rewards)
- ❑ Establish principles for the use of AI that is aligned with the corporate mission
- ❑ Consider existing AI principles and standards for relevance
- ❑ Principles should guide the development, deployment and use of AI internally and in the marketplace

Use the principles to develop a company policy



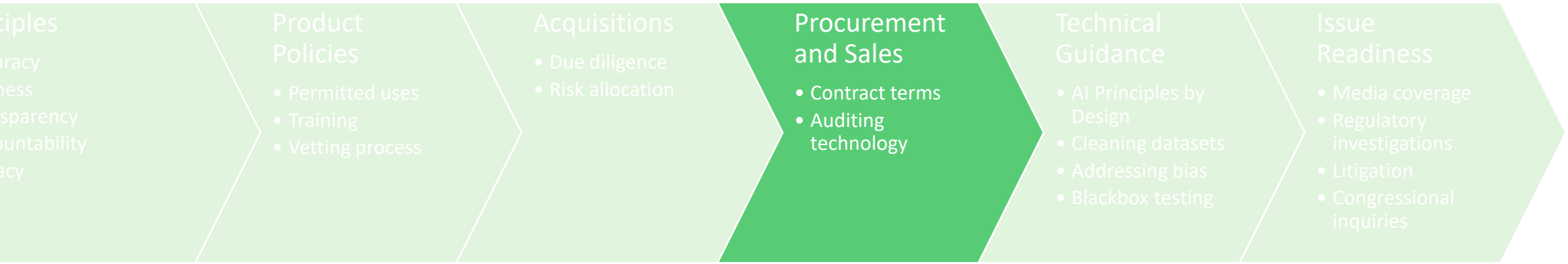
- ❑ Establish governance and accountability to address legal and ethical concerns
- ❑ Identify internal and external stakeholders for AI development and use
- ❑ Identify permissible and prohibited uses
- ❑ Establish use guidelines aligned with the AI principles
- ❑ Communicate AI policies *internally* for adoption
- ❑ Communicate AI policies *externally* for disclosure and transparency

Vet acquisitions for AI values and risks



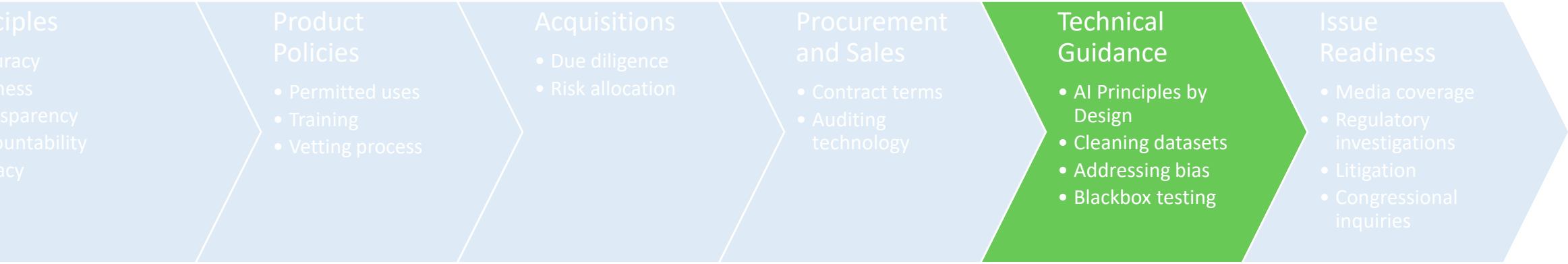
- ❑ Understand ownership and IP rights for AI models and data
- ❑ Question how AIs were developed (and data sets used for training)
- ❑ Apply a scrutiny similar to current cybersecurity and privacy vetting of data assets (e.g., due diligence questions and document requests)
- ❑ What commitments is the seller making regarding its AI technology
- ❑ Assess value of the AI relative to the deal, and the risks associated with a “bad AI” coming to light

Address AI in third party risk management process



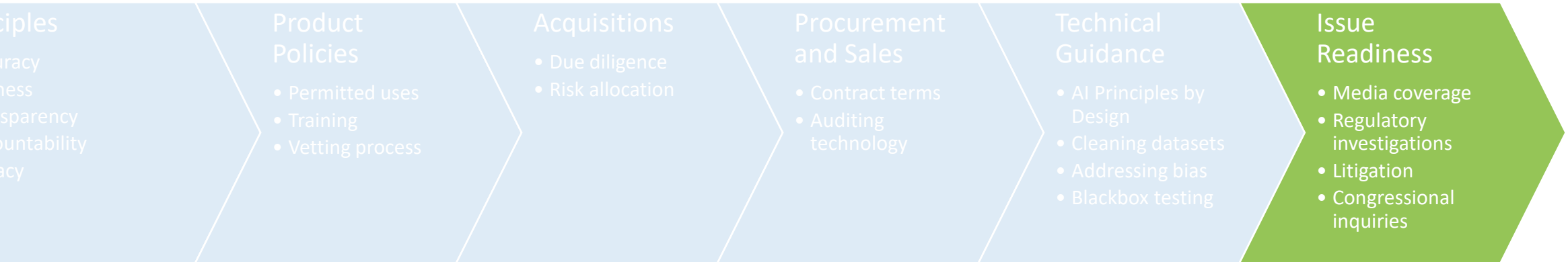
- ❑ Understand what data and AI technology is relevant to the services offered/received:
  - (1) What AI technologies and methodologies are being used? How? Who owns the trained model?
  - (2) Who owns datasets used to train the AI (initially and iterative)? Who owns derived data?
  - (3) What measures are taken to test for, and mitigate, bias, inaccuracy and other risks?
- ❑ Extend (cyber) due diligence to AI technology
- ❑ Understanding the allocation of responsibility and liability

Bridge the gap between policy and technology



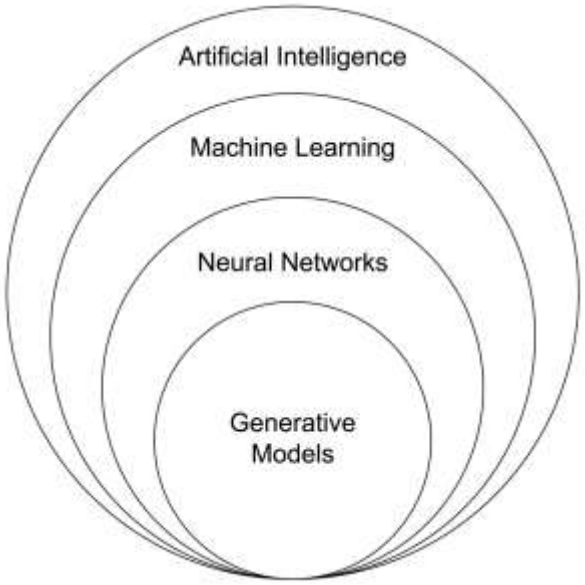
- ❑ Establish clear process for implementing AI, including use of open source or other third-party technology
- ❑ Guidance on data quality standards
- ❑ Maintain privacy of individuals within datasets
- ❑ Establish permissible “use cases” for AI solutions (consider “off label” risks)
- ❑ Audit outcomes to protect against legal bias, disparate impact and unexpected results

Be prepared: AI can fail, quickly and publicly



- ❑ Maintain transparency about the use of AI, particularly with respect to adverse/impactful decisions
- ❑ Establish complaint and response process for customers (i.e., alert system)
- ❑ Maintain an "incident response plan" for AI-related incidents (and unfavorable media attention)
- ❑ Keep a human touch and perspective despite AI insights and decisioning
- ❑ Program materials may be essential in responding to legal, business or reputational exposure (i.e., principles, policies, risk assessments, product vetting, etc.)

Example: ChatGPT



Generative AI is capable of generating text, images, or other media in response to prompts.

The Generative AI models learn the patterns and structure of their input training data by applying neural network machine learning techniques, and then generate new data that has similar characteristics.

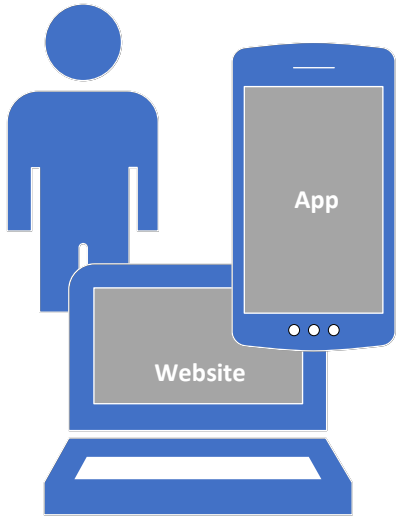
Potential risks:

- All output from generative AI is suspect and unverified – *it can produce completely false results*
- Confidential information input into the AI (e.g., personal information, proprietary code) may be used for training and *be output to other recipients*
- Results can be discriminatory, biased or violate applicable laws
- IP rights associated with generative AI are uncertain and still developing

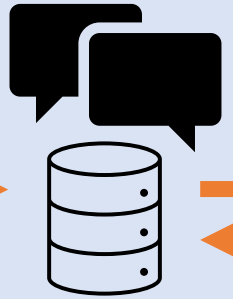
Example: Chatbots

Company Systems

End user

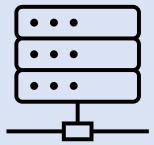


Chatbot tool



AI-powered Natural Language Processing (NLP) – *speech-to-text, text-to-speech*

Backend



AI-powered generative AI (e.g., ChatGPT), **or** Preset knowledgebase

Any decision-making?

Discussion and Q&A

Mark H. Francis
Partner | Data Strategy, Security & Privacy
Holland & Knight
31 West 52nd Street | New York, New York 10019
o. 212.513.3572
mark.francis@hklaw.com

https://www.hklaw.com/en/professionals/f/francis-mark-h
https://www.linkedin.com/in/markhfrancis/