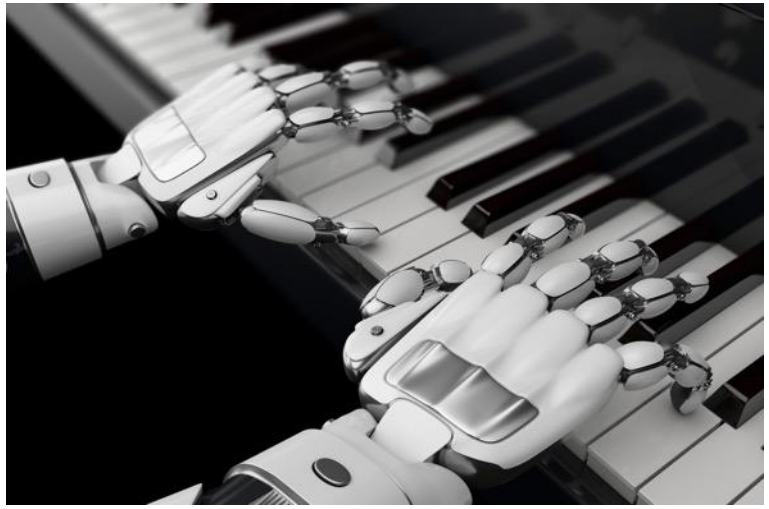
Responsible AI: From Bias and Privacy to Compliance and Risk Management

Anthony Mullen

Gartner®

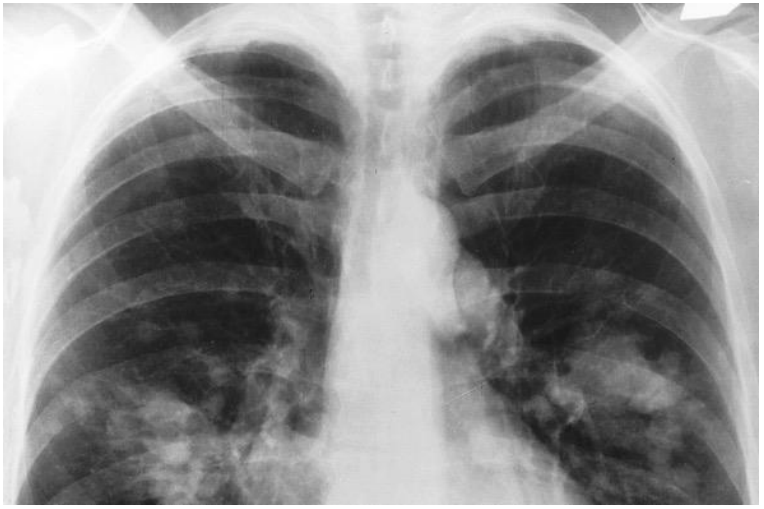
Gartner®

Cool!



Composing music for commercials, movies, video games

Aiva Technologies



Algorithms detect anomalies in real time.

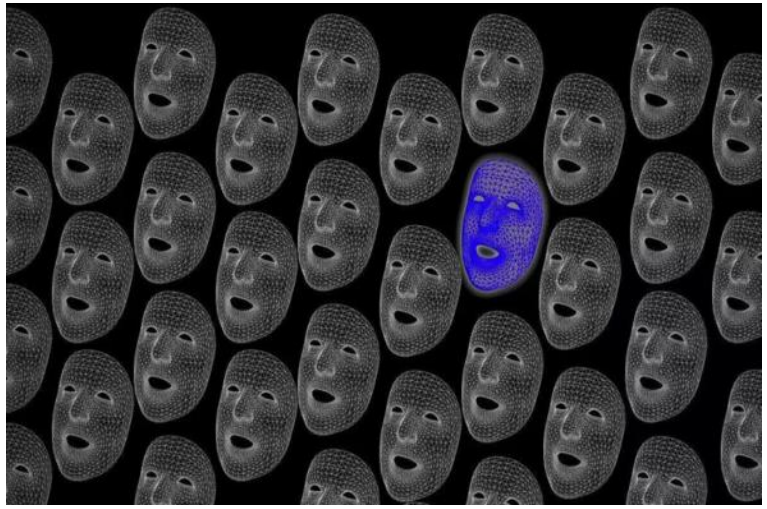
AGFA HealthCare

Domestic robots keeping people company

SoftBank Robotics

Gartner

Creepy!



Facial recognition leads to wrongful imprisonment

Children's **privacy violation** by smart AI toy

Filter bubbles empowered by AI on social media

To reap the benefits of AI, we need to empower the cool and protect us from the creepy, by managing AI risks

Gartner®

From Responsible AI to AI Risk Management

AI Risk Management

Responsible AI

Gartner

Key Issues

1. Eight AI risks and how to mitigate them
2. Emerging practices for AI risk management

Key Issues

1. Eight AI risks and how to mitigate them
2. Emerging practices for AI risk management

1. Trust

AI Risks

- Invalid (wrong)
- Unreliable (inconsistent)

Key Mitigations

- Improve model explainability & transparency
- Guard rails
- Human-in-the-loop

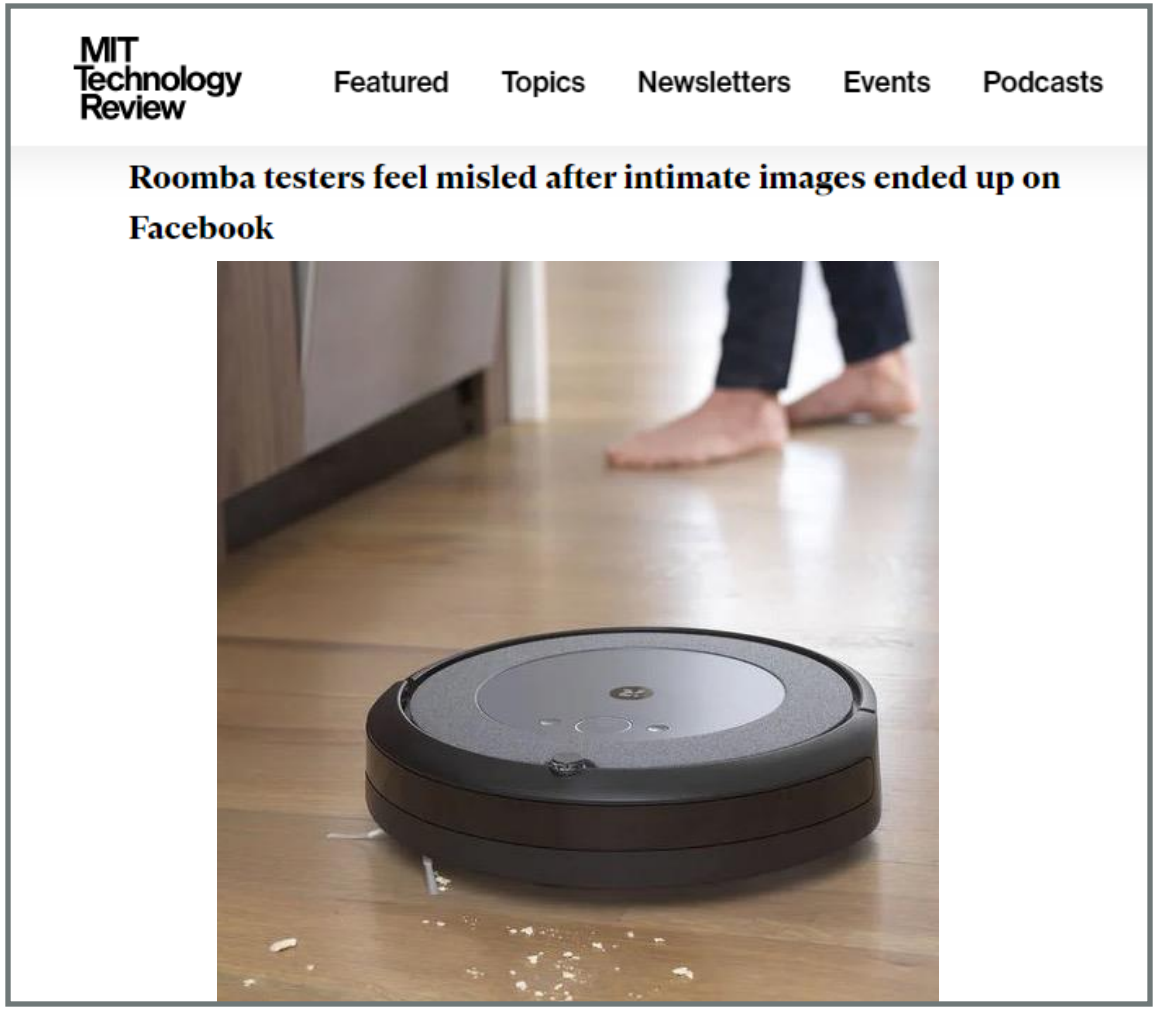
2. Privacy

AI Risks

- Unwarranted use of privacy sensitive data/personal identifiable information (PII)
- Unauthorized sharing of PII

Key Mitigations

- Assess intended use of AI & use data proportionally
- Apply privacy enhancing technologies (PET)
- Keep data secure



Source: MIT TechnologyReview

3. Fairness

AI Risks

- Bias against groups or classes of people, based on their color, ethnicity, gender, religion, age or any other classification protected by law.

Key Mitigations

- Balance diversity in data (and in team)
- Use fairness tools/frameworks
- Include fairness in model monitoring

A.I. Bias Caused 80% Of Black Mortgage Applicants To Be Denied

Kori Hale Contributor

I'm the CEO of CultureBanx, redefining business news for minorities.

Follow

Source: A.I. Bias Caused 80% of Black Mortgage Applicants to Be Denied, Forbes Media.

Gartner

4. Security

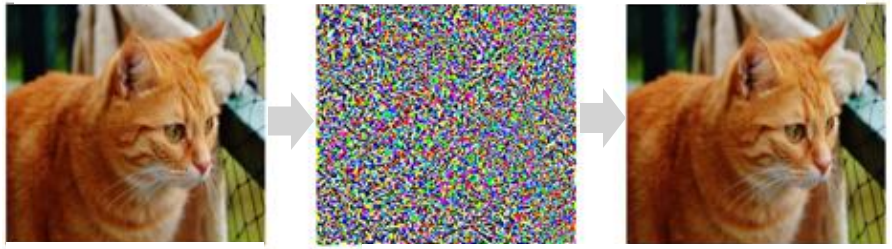
AI Risks

- Data poisoning
- Model theft
- Model deception

Key Mitigations

- AI — cybersecurity collaboration
- Secure training data
- Harden AI models

Manipulated Digital Images



Add perturbations to digital image to fool AI model

Source: iProov; DARTS: Deceiving Autonomous Cars With Toxic Signs, arXiv.

5. Safety

AI Risks

- Unsafe outcomes, beyond intended or tested use
- Amplification of harmful output through automation


Chess robot grabs and breaks finger of seven-year-old opponent

Moscow incident occurred because child 'violated' safety rules by taking turn too quickly, says official

Key Mitigations

- Monitor & alert for unintended consequences
- Prefilter outliers or untested cases
- Frequent drift monitoring

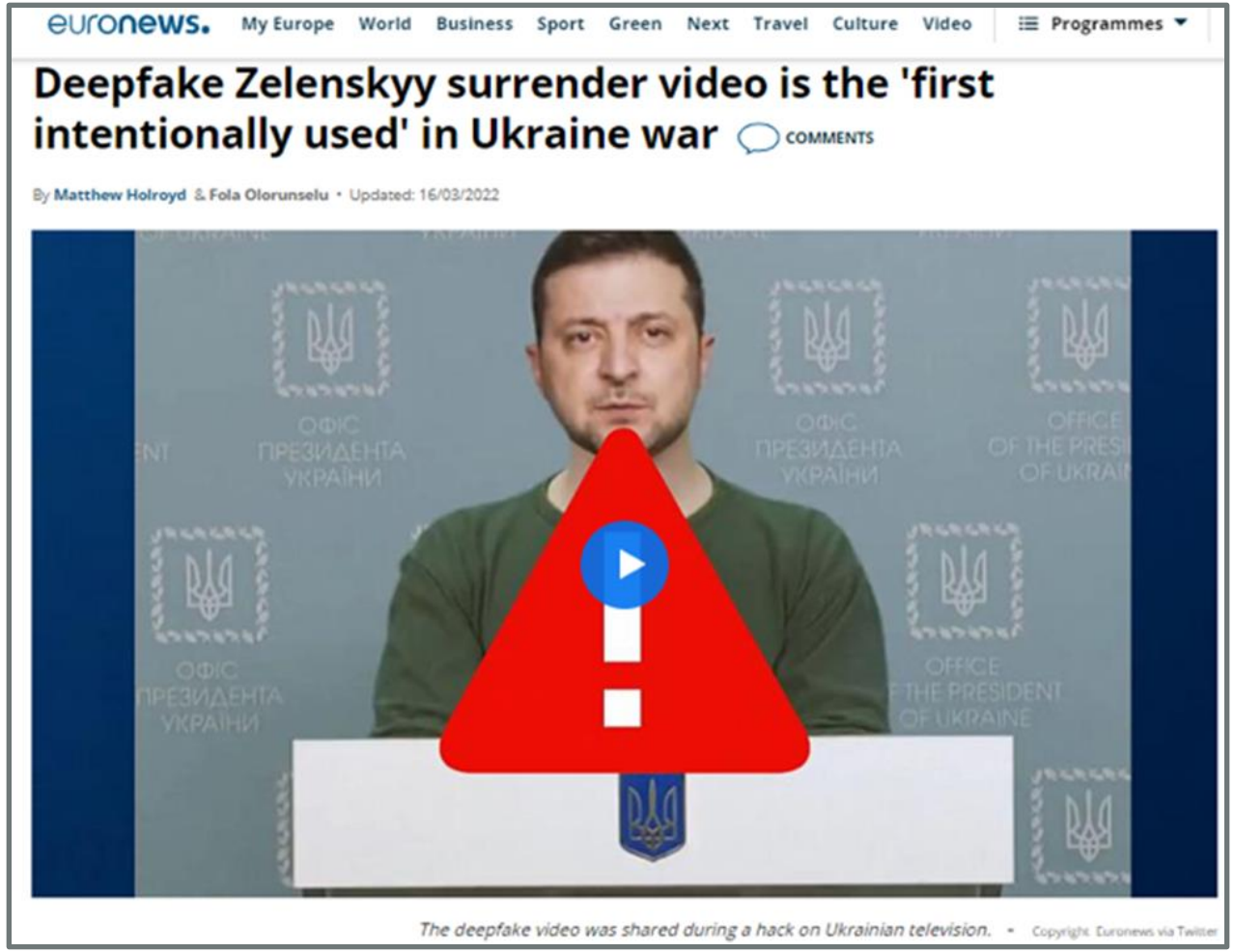
6. Societal Impact

AI Risks

- Negative impact on human well-being
- Misinformation and social bubble amplification

Key Mitigations

- Apply human-centric design principles
- Verify content & detect fakes
- Limit content personalization



Source: Deepfake Zelenskyy Surrender Video Is the "First Intentionally Used" in Ukraine War, Euronews.

7. Sustainability

AI Risks

- Growing energy consumption and carbon footprint of AI

Key Mitigations

- Apply emerging sustainable AI practices to reduce energy consumption
- Proactively apply AI to improve sustainability in business operations and society

Source: AI Power Consumption Exploding, Semiconductor Engineering.

8. Accountability

AI Risks

- Insufficient attention to preventing or mitigating AI failures
- Noncompliance with growing number of complex regulations

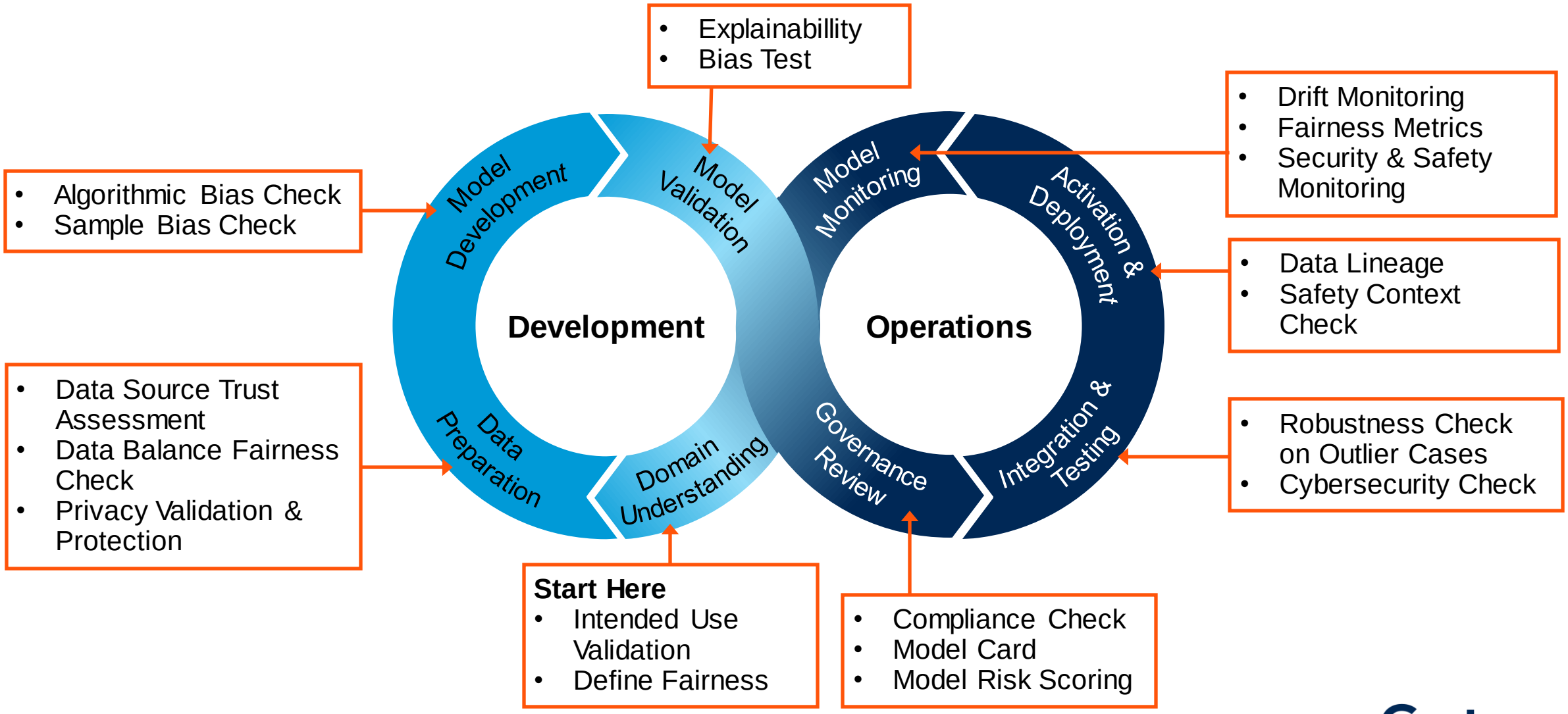
Key Mitigations

- Create own set of principles and actionable guidelines (but do not reinvent the wheel)
- Partner with legal and compliance experts
- AI risks & ethics board

Key Issues

1. Eight AI risks and how to mitigate them
2. Emerging practices for AI risk management

Full Life Cycle AI Risk Management



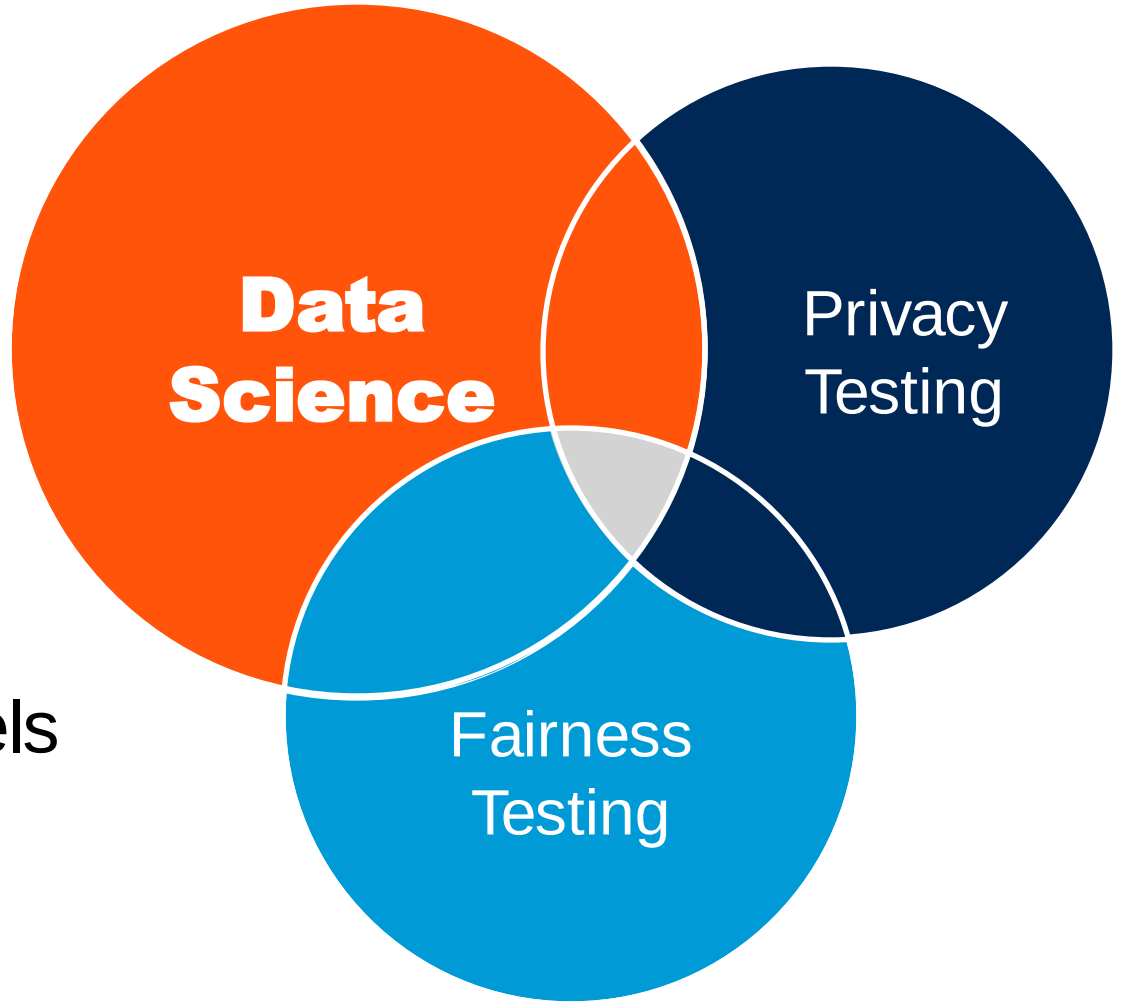
Gartner®

Adding an AI Model Validator Role

1/3

model validator **model developers**

- Not involved in model development
- Internal or external role
- Test for bias and other risks in models
- Audit for compliance



Gartner®

Create an Awareness and Training Program for AI Risks & Ethics

- Every month, send around a dilemma. Discuss! What would you do?
- It trains being comfortable with ambiguity, and ethical thinking.

Mikkel Muller, CIO

DR

The Daily Telegraph

U.K. 2016: The Daily Telegraph installed motion sensors under the desks of employees to improve ener
sustainability. The sensor boxes, branded OccupEye and made by Cad-Capture, contain motion and ten
link back wirelessly to central network points, but do not monitor the desk areas themselves. It was, how
the sensors could be used to monitor desk occupancy.

Sources: Hacks Rebel After Bosses Secretly Install Motion Sensors Under Desks, The Register, Jan. 2016
National Union of Journalists Criticizes Daily Telegraph for Monitoring Journalists' Desk Time, The Guardian, Jan. 2016

Arguments Pro

- Such instruments provide accurate occupancy data to building managers, who can then use it for maintaining temperature, reallocating unutilized desk space and other energy- and space-efficient measures.

Arguments Con

- If the intent was to gather occupancy and temperature
could have identified alternative appropriate locations a
than installing the device beneath the desks of employe
suspicious.

S05 — Insight Versus Privacy

Many digital initiatives start with the best of intentions — in this case, introducing sensors as part of a smart building sustainability initiative. But thi
speculation and interpretation of the story — that they were meant to monitor employees. This demonstrates how highly sensitive people can be w
issues they may see as affecting their privacy, even if this is not the case in practice.

Gartner for IT Leaders Toolkit
Notes accompany this presentation. Please select Notes Page view to examine the Notes text.
63 CONFIDENTIAL AND PROPRIETARY I © 2016 Gartner, Inc. and/or its affiliates. All rights reserved.

Gartner

Create and Involve AI Risks & Ethics Board

- AI risk management is not a checklist.
- Ethical dilemmas should not be left to project teams to figure out, causing stagnation or risky decisions.
- Board decides/advices on guidelines and project dilemmas.
- AI champion, business owners, external experts, other stakeholder representatives.

Organizations That Manage Their AI Risks Realize More Value From AI

Recommendations

- Prioritize and contextualize AI risks for your organization. At least include privacy, fairness and accountability.
- Not all regulations and ethics are globally the same. Take regional and cultural differences into account.
- Make guidelines actionable. Do not reinvent the wheel: leverage existing policies, frameworks and tools.
- Partner with AI governance, legal, cybersecurity and other stakeholders.
- AI and AI risk management are about people. Build awareness and skills.

Recommended Gartner Research

- **A Comprehensive Guide to Responsible AI**
  Farhan Choudhary and Svetlana Sicular (G00764905)
- **AI Ethics: Use 5 Common Principles as Your Starting Point**
  Frank Buytendijk, Erick Brethenoux and Others (G00774103)
- **Applying AI — Governance and Risk Management**
  Avivah Litan, Svetlana Sicular and Others (G00745080)
- **Digital Ethics by Design: A Framework for Better Digital Business**
  Frank Buytendijk, Jim Hare and Lydia Clougherty Jones (G00421878)
- **Market Guide for AI Trust, Risk and Security Management**
  Avivah Litan, Jeremy D'Hoinne and Others (G00758388)

Gartner®