AI GOVERNANCE

LAW 371
SANTA CLARA U. LAW SCHOOL

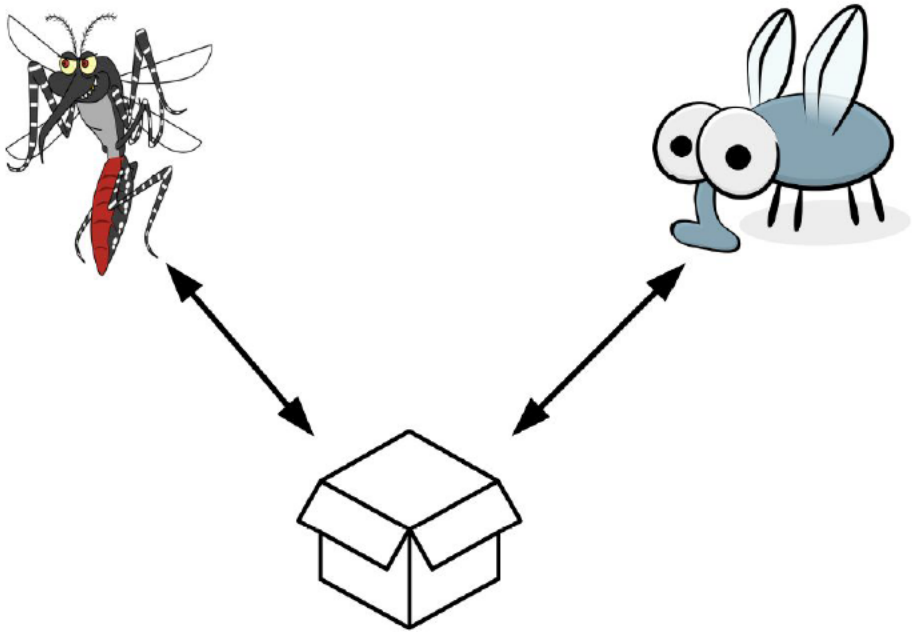
Agenda

1. CRISP-DM
2. Explainability
3. Cybersecurity
4. Privacy
5. Public Policy
6. Example Regulations

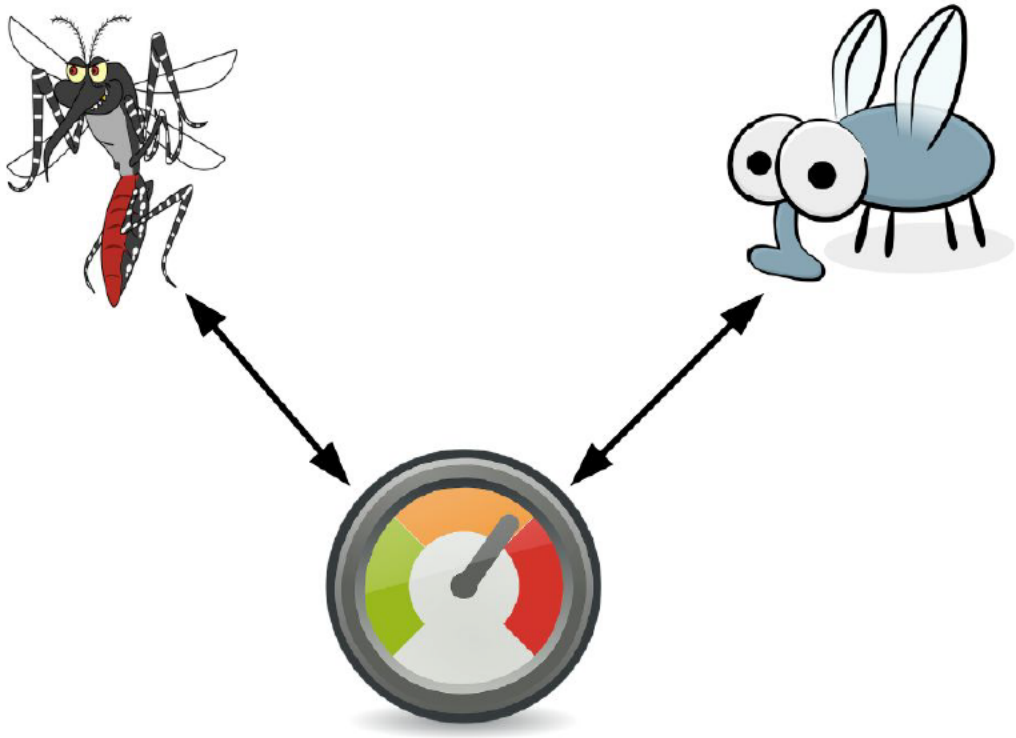
CRISP-DM


Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Data

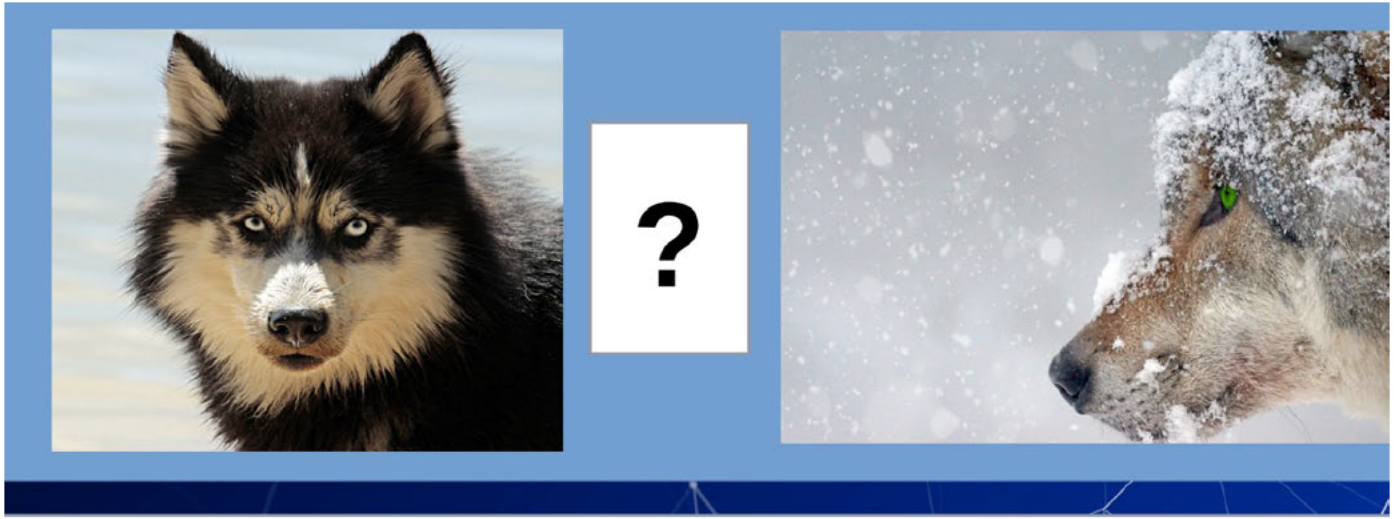
CRISP-DM (biz purpose)



CRISP-DM (biz purpose)



Explainability

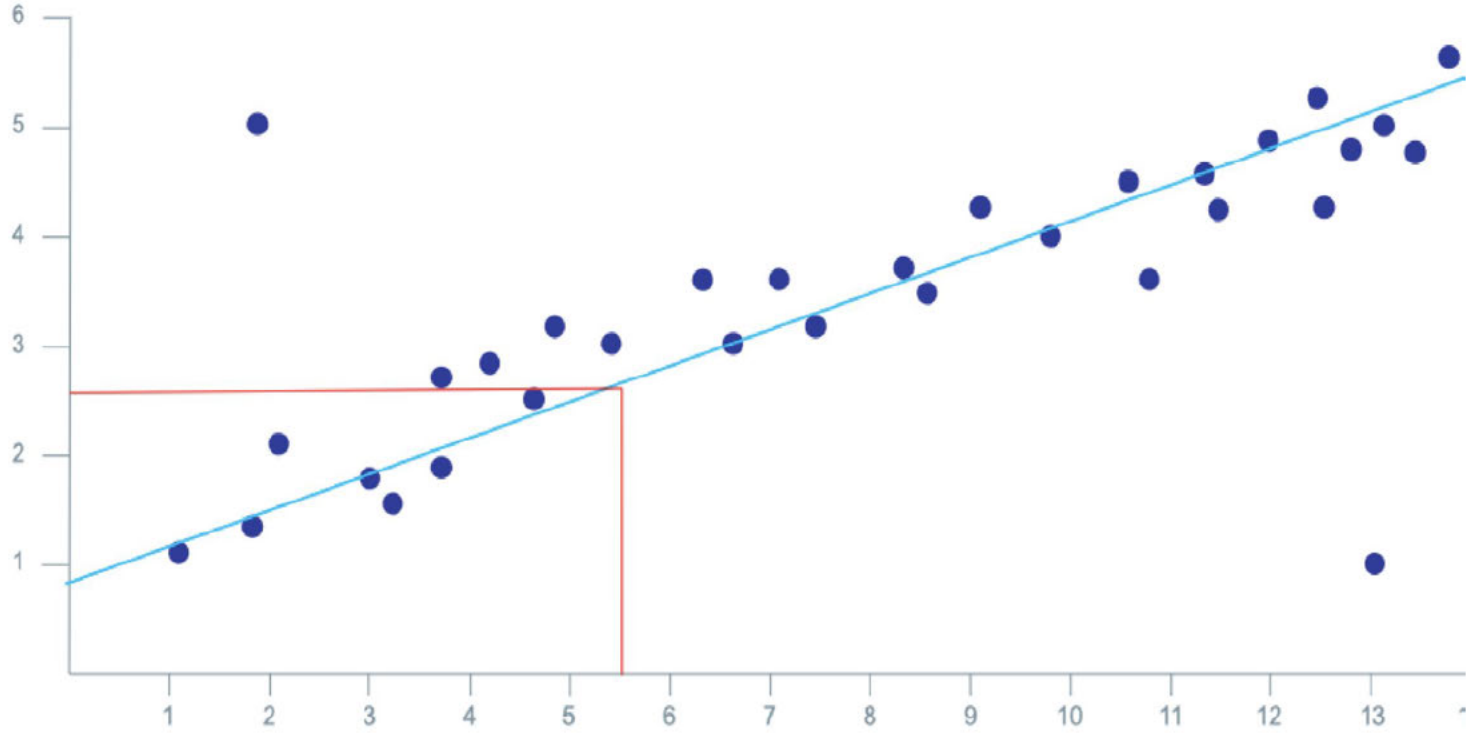


- Thousands of pictures of huskies
- Thousands of pictures of wolves
- Model learns from pictures to build model to classify
- 90% accuracy. Yes, but...

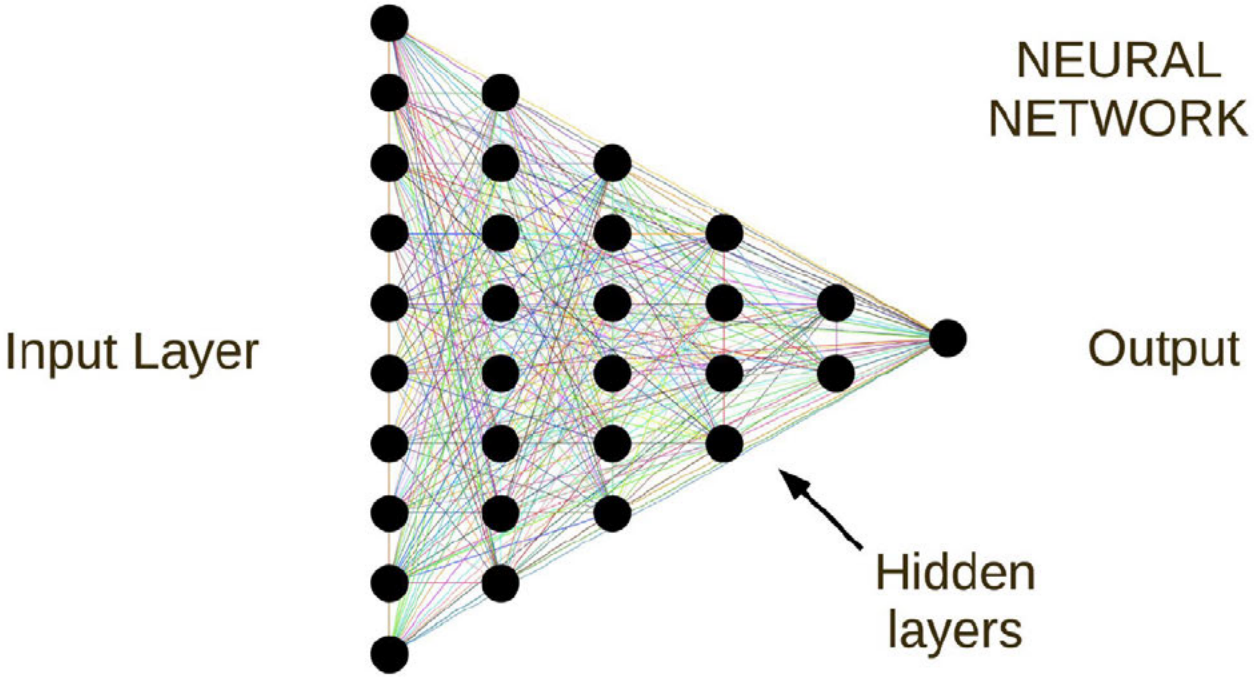
Explainability

- **Thousands of pictures of huskies**
- **Thousands of pictures of wolves**
- **Model learns from pictures to build model to classify**
- **90% accuracy. Yes, but...model uses snow as differentiator!**

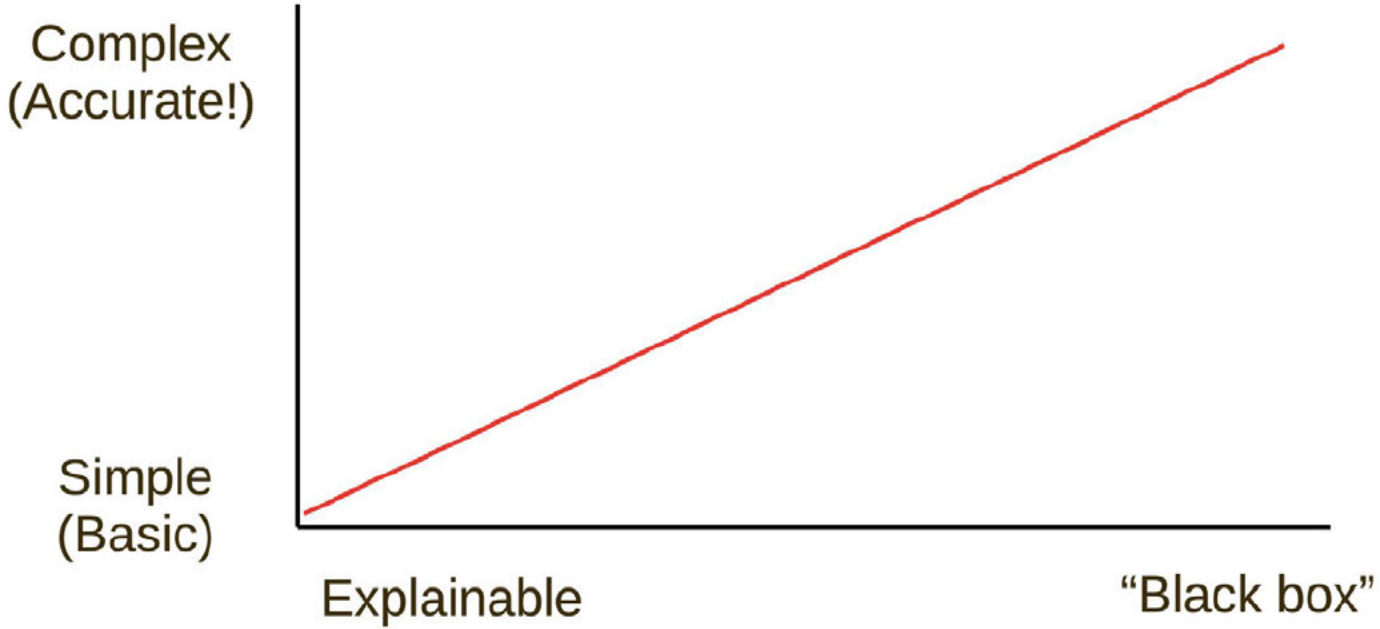
Explainability



Explainability



Explainability

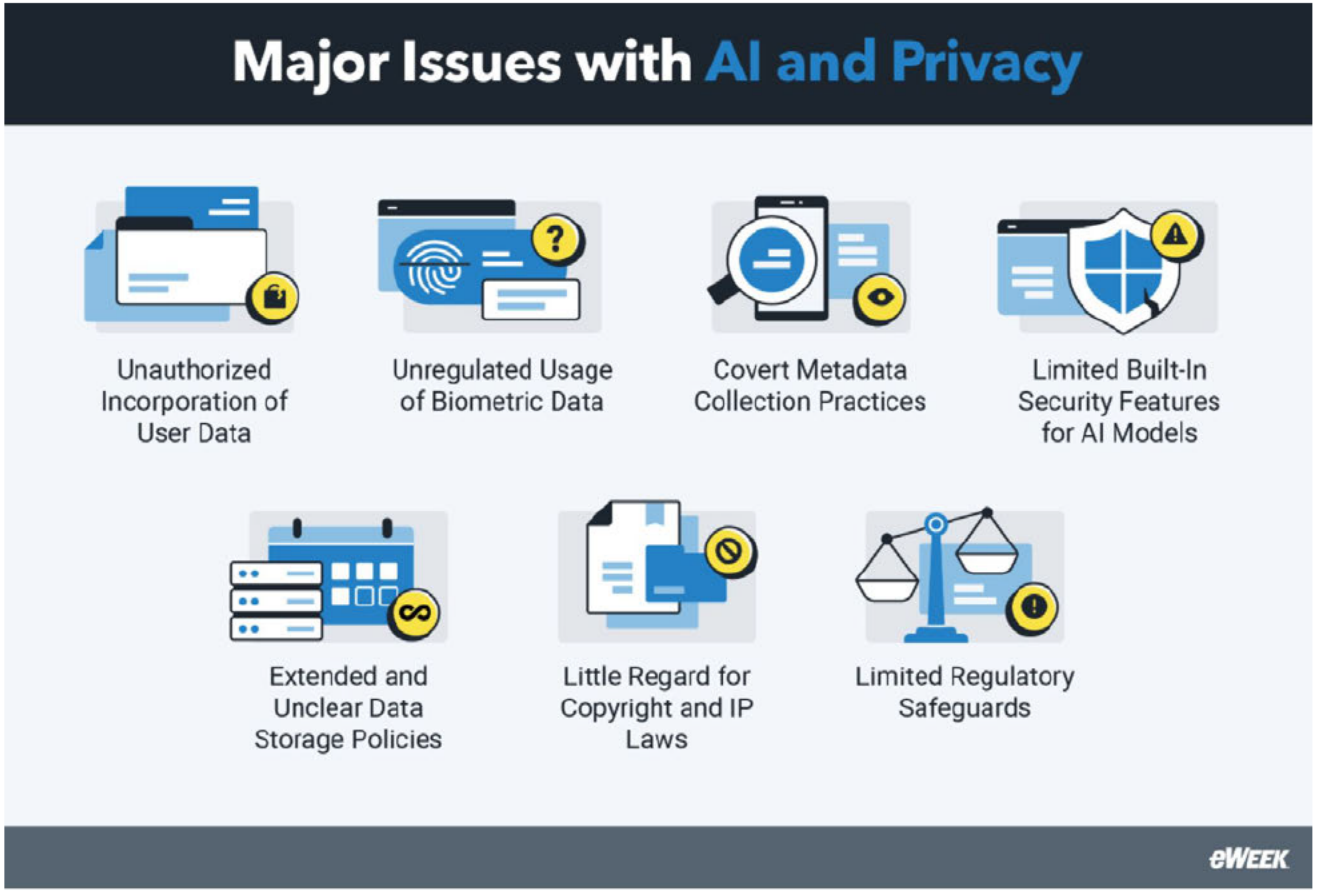


Cybersecurity

Privacy

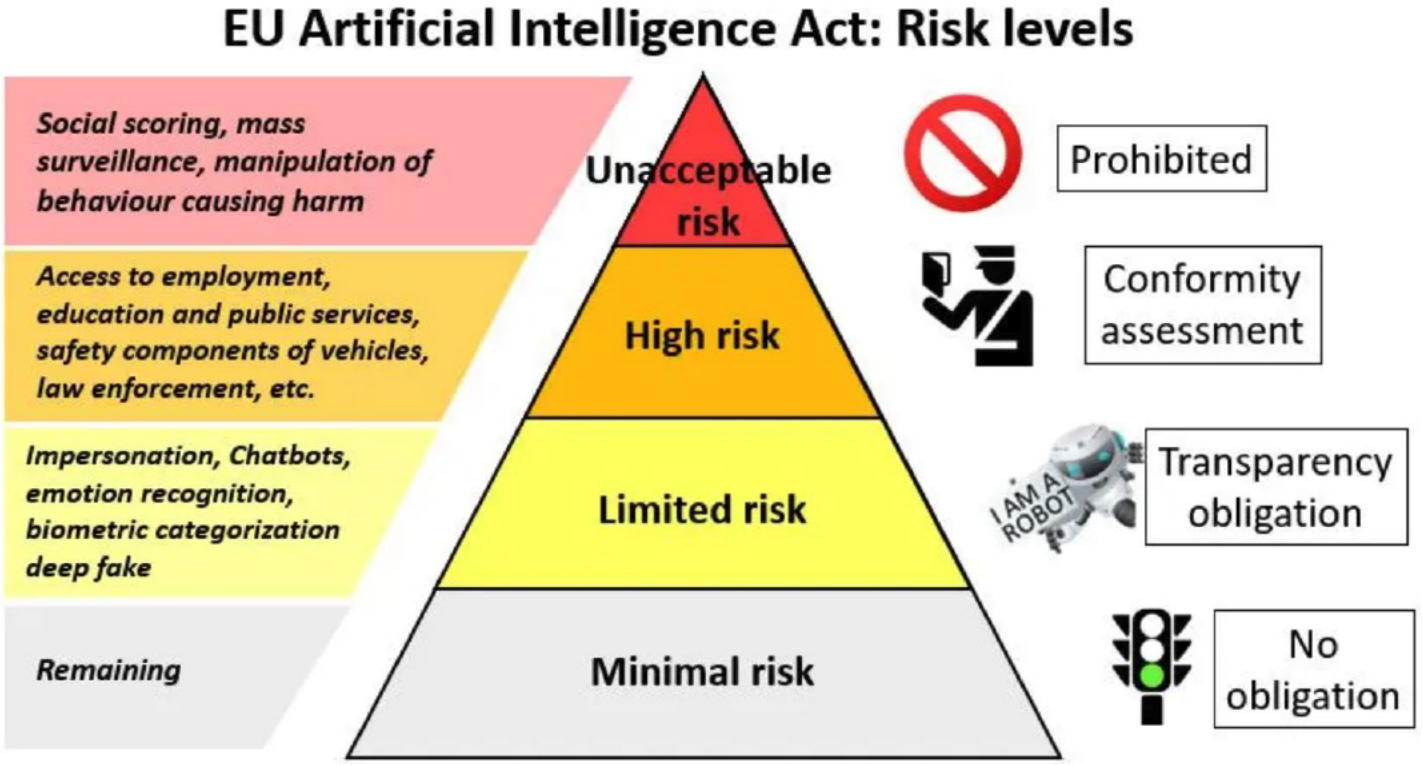

The 7 key GDPR principles
Lawfulness, fairness, and transparency
Purpose limitation
Data minimization
Accuracy
Storage limitation
Integrity and confidentiality
Accountability

Privacy

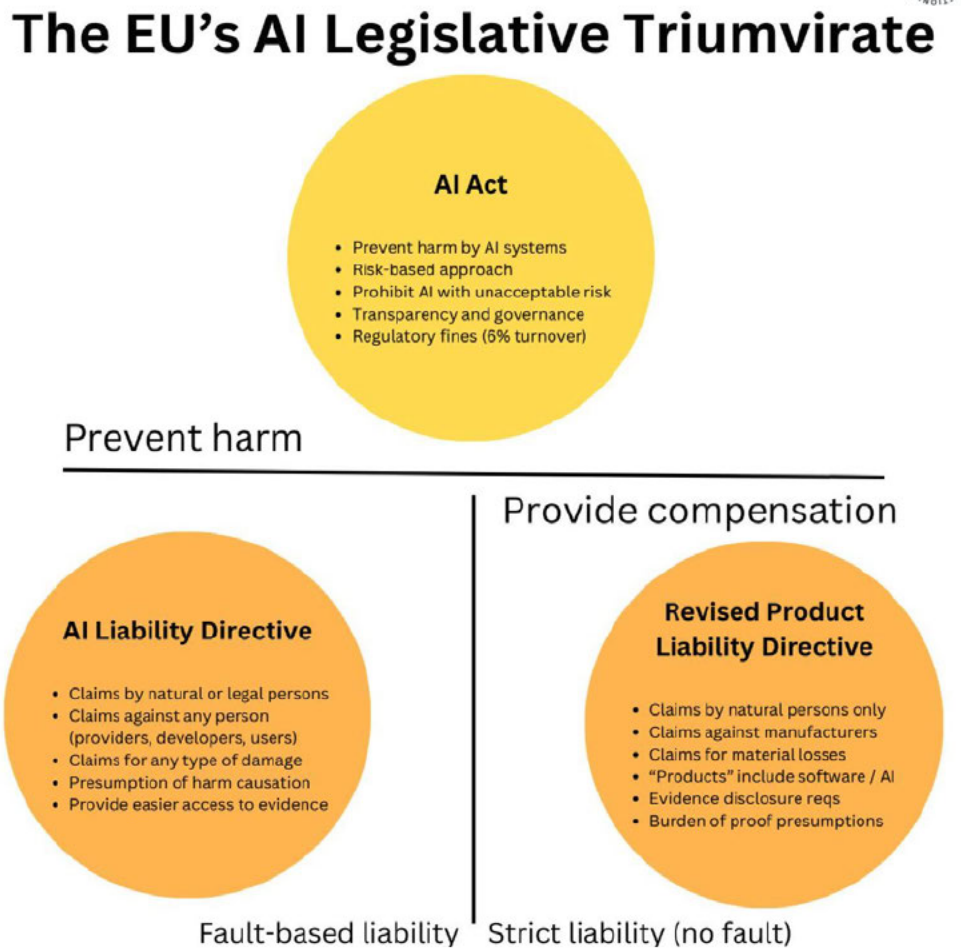

Major Issues with AI and Privacy
Unauthorized Incorporation of User Data
Unregulated Usage of Biometric Data
Covert Metadata Collection Practices
Limited Built-In Security Features for AI Models
Extended and Unclear Data Storage Policies
Little Regard for Copyright and IP Laws
Limited Regulatory Safeguards
eWEEK

Public Policy

Examples


EU Artificial Intelligence Act: Risk levels
Social scoring, mass surveillance, manipulation of behaviour causing harm
Access to employment, education and public services, safety components of vehicles, law enforcement, etc.
Impersonation, Chatbots, emotion recognition, biometric categorization deep fake
Remaining
Unacceptable risk
High risk
Limited risk
Minimal risk
Prohibited
Conformity assessment
I AM A ROBOT
Transparency obligation
No obligation

Examples


The EU's AI Legislative Triumvirate
AI Act
Prevent harm by AI systems
Risk-based approach
Prohibit AI with unacceptable risk
Transparency and governance
Regulatory fines (6% turnover)
Prevent harm
Provide compensation
AI Liability Directive
Claims by natural or legal persons
Claims against any person (providers, developers, users)
Claims for any type of damage
Presumption of harm causation
Provide easier access to evidence
Revised Product Liability Directive
Claims by natural persons only
Claims against manufacturers
Claims for material losses
"Products" include software / AI
Evidence disclosure reqs
Burden of proof presumptions
Fault-based liability
Strict liability (no fault)