

# Machine Learning in Information Security

---



Prof. Sukumar Nandi

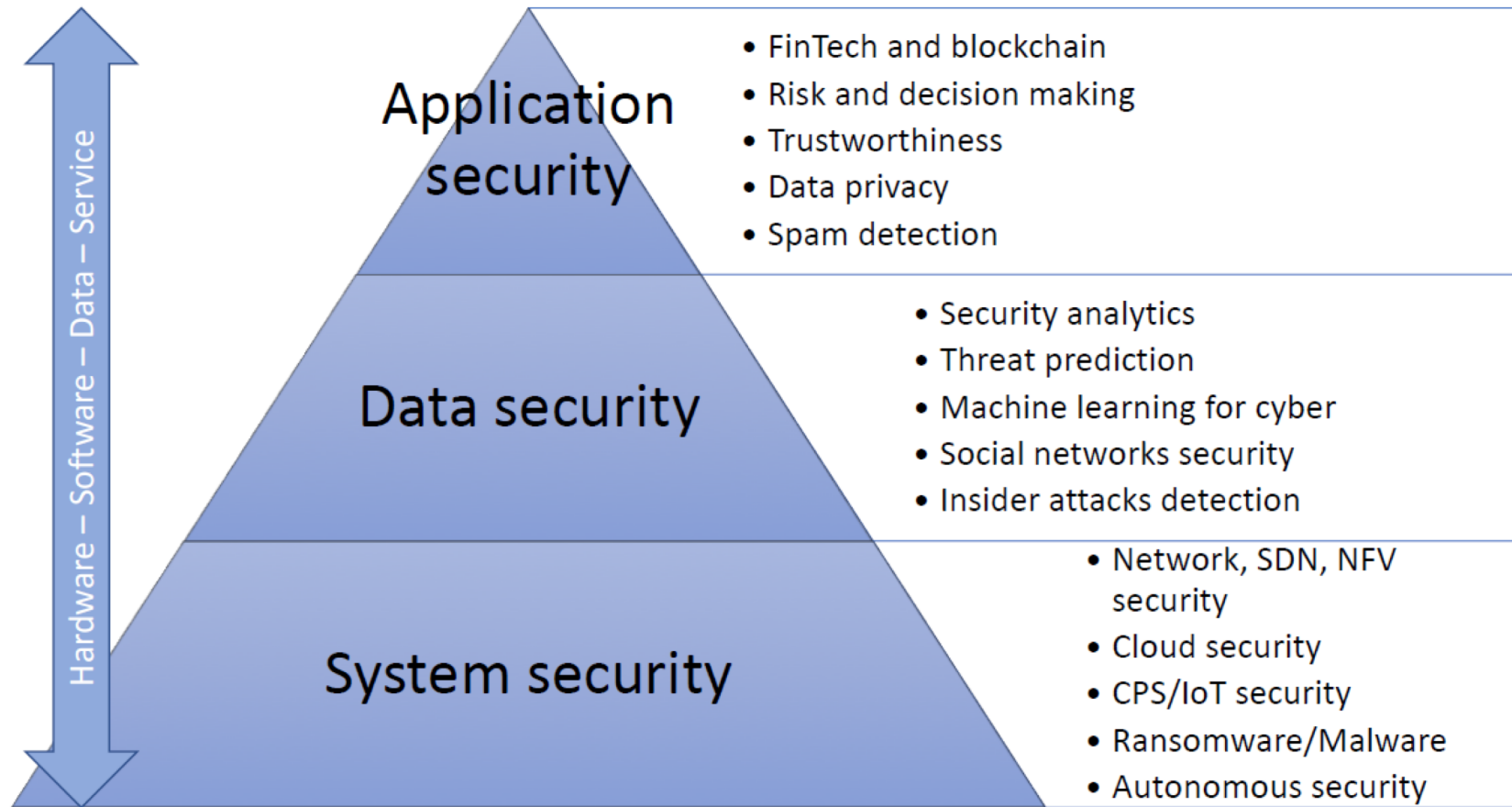
Department of Computer Science and Engineering  
Indian Institute of Technology Guwahati  
Guwahati, Assam

<https://www.iitg.ac.in/sukumar/>

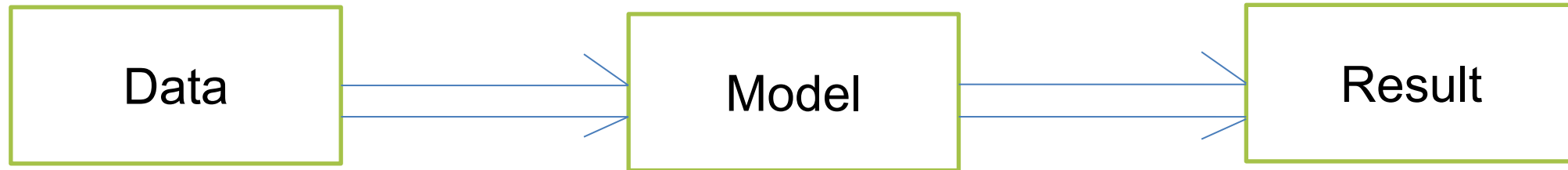
# Research on Information Security @ IIT Guwahati

- A team of professor with Ph.D. Scholar leading capabilities in cyber security
- We develop innovative defense mechanisms for securing cyberspace
- We work with industry and government to provide a security solution for protection of Cybersystems from major cyber security threats

# Core Capabilities @ IIT Guwahati



# Process

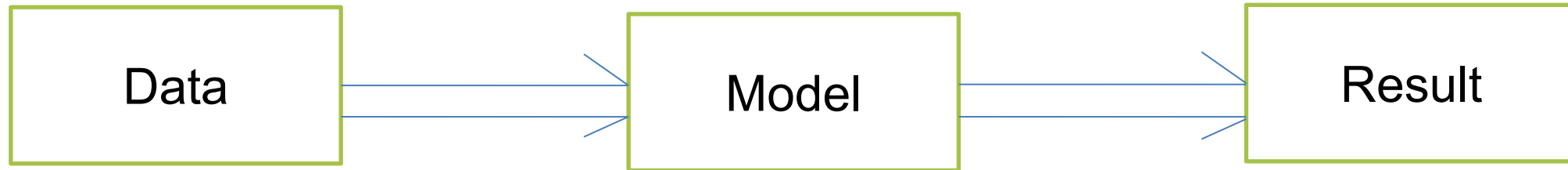




# Data

- Big Data --- sampling
- Incomplete --- multi-point acquisition
- Unbalanced --- Skewed --- Small sample size
- Life time --- stream
- context sensitive
- time series
- Unencrypted/encrypted/compressed
- multi modal

# Tuning



# How AI/ML can be Useful



- Improve Speed of Reasoning
- Improve Speed of Reaction  
(human latency vs. machine latency)
- Scale in Data (reasoning over huge amount of data)
- Scale Personnel (force multiplier)
- See Patterns and Connections – where humans can't

# AI/ML Milestones So Far



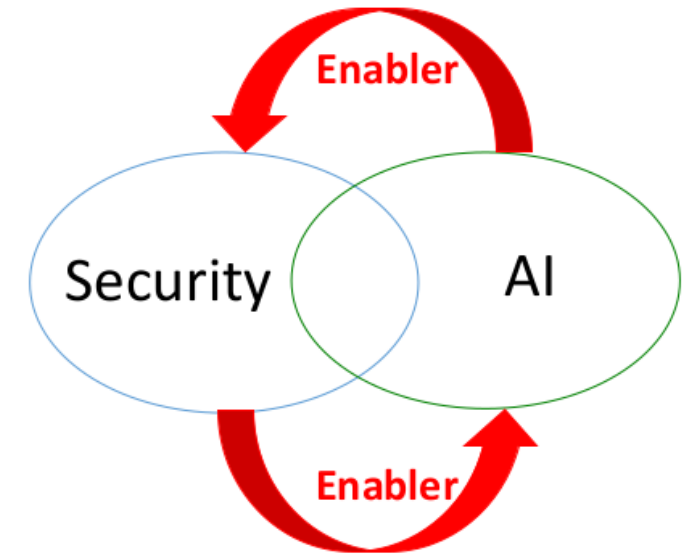
- [1997] Deep Blue defeats world chess champion Garry Kasparov
- [2005] The DARPA Grand Challenge (race for Autonomous Vehicles)
- [2011] IBM Watson's Jeopardy! Victory
- [2012] Power of deep learning – computers learn to identify cats
- [2015] Machines “see” better than humans (annual ImageNet challenge)
- [2016] AlphaGo defeats world Go champion Lee Sedol
- [2018] Self-driving cars hit the roads
  - (The EFF maintains a [page](#): “AI Progress Measurement”)

- There have been many reports of AI showing “gender bias” and “racial bias”
- Instances of incorrect face detection
- Fatality involving self-driving cars
- Wrong recommendations & predictions in the fields like Healthcare & Sports

# Two Aspects of AI in Security



- AI for Security
  - AI enables security applications
- Security for AI
  - Security enables better AI
    - *Integrity*: Produces intended / correct results (adversarial machine learning)
    - *Confidentiality/Privacy*: Does not leak user's private data
    - Preventing *misuse* of AI



# AI in the presence of Attacker

- Attack AI
  - Cause learning systems to not produce intended / correct results
  - Cause learning systems to produce targeted outcome designed by the attacker
  - Learn sensitive information about individuals
  - Need security in learning systems

# AI in the presence of Attacker (contd.)

- Misuse AI
  - Misuse AI to attack other systems
    - Find vulnerabilities in other systems
    - Target attacks
    - Devise attacks
  - Need security in other systems



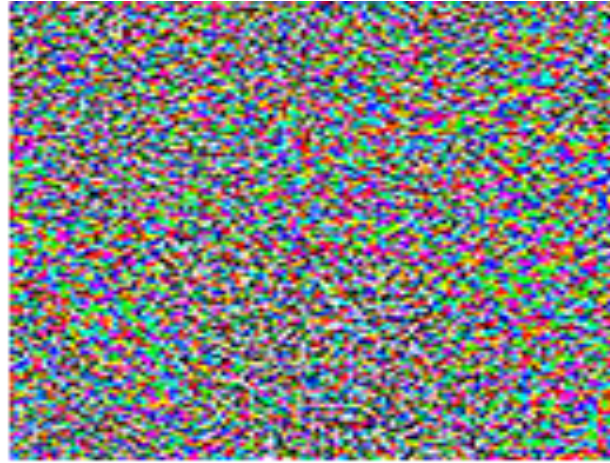
# Adversarial Machine Learning



“panda”

57.7% confidence

+  $\epsilon$



=



“gibbon”

99.3% confidence

Adversarial examples are inputs to machine learning models that an attacker has intentionally designed to cause the model to make a mistake.


# Adversarial Machine Learning (contd.)

- Learning in the presence of adversaries
- Adversarial example fools learning system
- Attacker poisons training dataset (eg. poisons labels) to fool learning system to learn wrong model
- Attacker selectively shows learner training data points (even with correct labels) to fool learning system to learn wrong model

# Adversarial Machine Learning (contd.)

- Data poisoning is particularly challenging with crowd-sourcing & insider attack
- Difficult to detect when the model has been poisoned
- Adversarial machine learning particularly important for *security critical systems*

# With Cyberspace comes Cybercrime

- Red and Blue AI 
  - Red AI is AI based Cybercrime
    - Next Level Threats based Cyber attacks
    - Difficult to Assess the level of this Risk
  - Blue AI is Defensive AI
    - Best Defence against Red AI
    - Mostly uses a mix of different technologies, such as – Classification, Anomaly Detection, Cluster Analysis

# Some Uses of Red AI

- Malware Creation
- Smart Botnets
- Advanced Spear Phishing
- Counter Threat Intelligence
- Unauthorised Access
- Poisoning ML Engines
- Using AI to Classify victims & Optimize RoI
- Condition based Cyber attacks (eg. cyber attacks using blockchain based smart contracts)

# Some Uses of Blue AI

- Malware Detection
- Anti-Spam
- Intrusion Detection & Prevention
- Vulnerability Management
- User and Entity Behaviour Analysis (UEBA)
- Data Classification (ease compliance with data privacy and data protection regulations)
- Cyber Threat Intelligence
- New generation of Honeypots

# Few Areas of ML in Cybersecurity



- Identify Anomalies
- Check Suspicious or Unusual Behaviour
- Detect & Correct known Vulnerabilities
- Suspicious behaviour & Zero-day Attacks
- Resource Optimization
- Improve Accuracy & Effectiveness of the response to an Attack

# Few ML Apps in Cybersecurity

- Malware Detection
- Intrusion Detection
- Fraud Detection & Prevention
- Credit Scoring
- Phishing Prevention
- Spam Filtering
- Botnet Detection



# Few ML Apps in Cybersecurity (contd.)

- Cyber Ratings
- Incident Forecasting
- User Authentication

# AI/ML in Finance

- Fraud Detection
- Identity & Access Management
- Monitoring & Preventing Threats
- Insider Threats
- Anti Money Laundering
- Risk Management
- Customer Protection
- Regulatory Compliance

# Threats in Healthcare

- Limited Cybersecurity Budget
- Black market for Electronic Health Records
- Vulnerable IoT and IoMT devices
- Unsecured Mobile Devices
- Ransomware
- Malware & Phishing
- Cloud Security
- Under-trained Staff

# Potential Attacks in Connected Vehicles

- Attacks on Connected & IoT Devices
- OTA Updates
- Engine Control Units
- Remote Hijacking or Vehicle Theft
- Ransomware
- DDoS
- Attacks on the Cloud automotive service provider

# Ethics in AI

- Automation & the resulting loss of human jobs
- Predictive Cybersecurity (eg. accused are implicated in crimes that have yet to be committed)
- Algorithmic Transparency
- Poor quality and/or inadequate quantity of data on which predictions are based (bias)
- Predictive capability of the algorithm used
- Some of the information learned might be private or confidential (think GDPR regulations)

# Compliance & Regulatory Challenges

- Privacy & Confidentiality
- Accountability & Liability
- Governance
- Bias & Discrimination
- Algorithmic Transparency

# Research Area – AI for Security

- Enhance Trustworthiness of Systems
- Resilient & Autonomous Cyber Actions
- Attacker-oriented Cyber Defence
- Predictive Analytic for Security

# Research Area – Security of AI

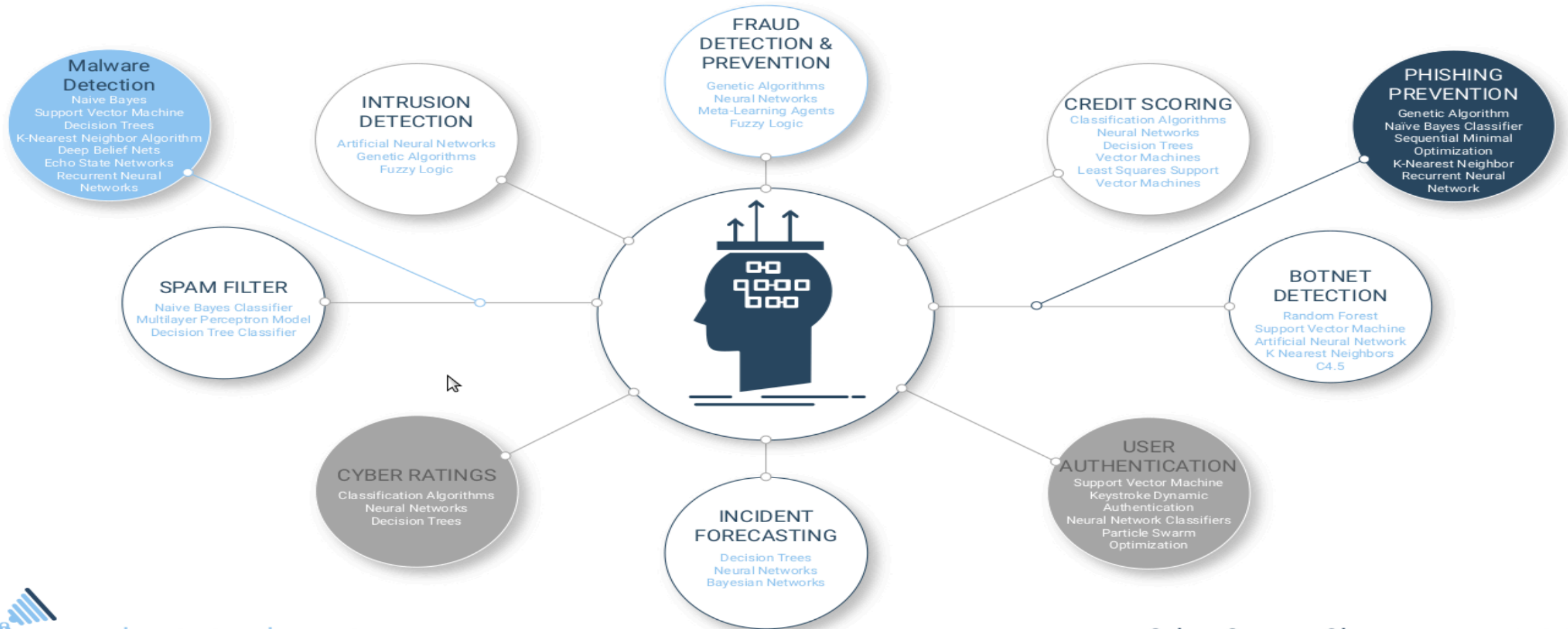
- Specification & Validation of AI Systems
- Trustworthy AI Decision Making
- Trustworthy Machine Learning
- Engineering Trustworthy AI/ML-augmented Systems



# Blue AI Emerging Usages

- **Automated Remediation** (recommendations that carry the lowest risk to the company, or that offers the highest benefit)
- **Orchestration Framework**
- **Intrusion Detection & Prevention – Autonomous Incident Response**
- **Intrusion Detection & Prevention – Automatic Self-assessment & Remediation**
- **Automated Penetration Testing**
- **Blockchain** (could enhance data integrity, digital identities, enable safer IoT devices to prevent DDoS attacks etc.)

# AI Algorithms Used in Cybersecurity Applications



# Third Wave of AI



## Symbolic AI

Logic rules represent knowledge

No learning capability and poor handling of uncertainty



## Statistical AI

Statistical models for specific domains training on big data

No contextual capability and minimal explainability



## Explainable AI

Systems construct explanatory models

Systems learn and reason with new tasks and situations

## Factors driving rapid advancement of AI



GPUs , On-chip  
Neural Network



Data  
Availability

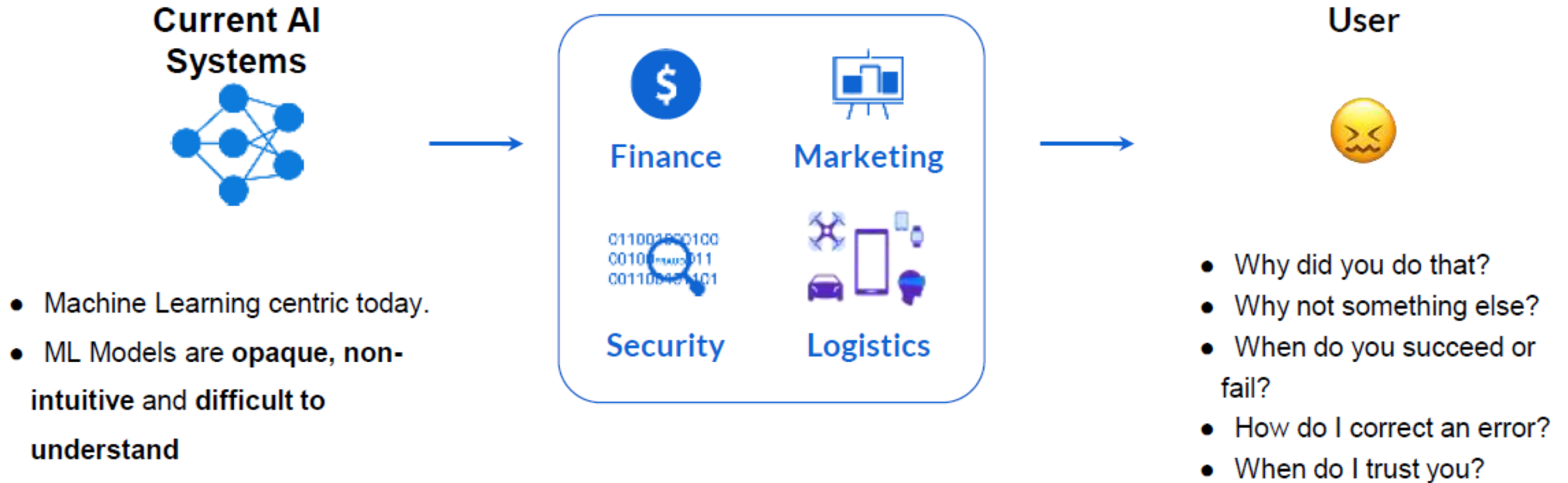


Cloud  
Infrastructure



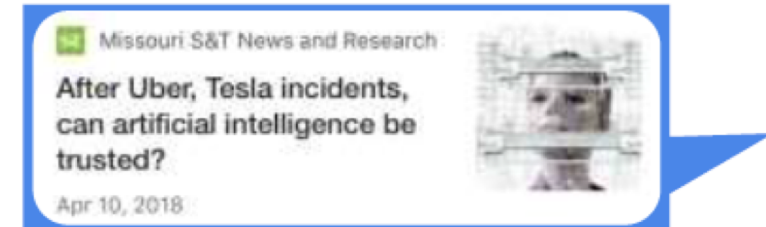
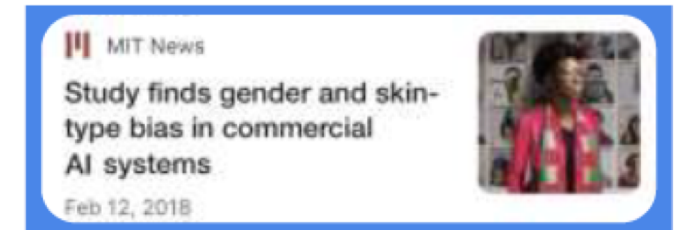
New  
Algorithms

# Need for Explainable AI



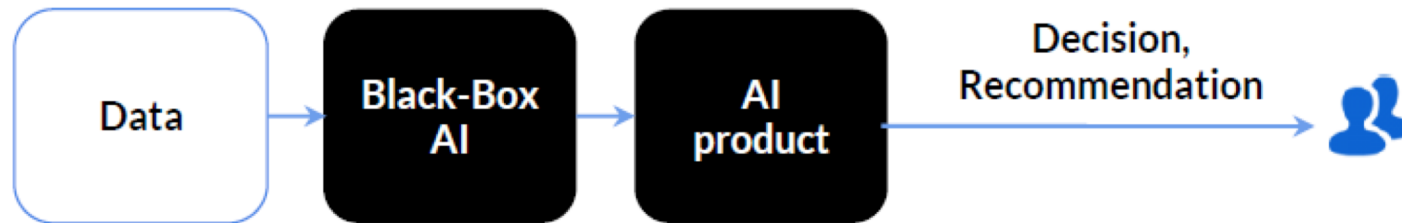
**Explainable AI and ML** is essential for future customers to understand, trust, and effectively manage the emerging generation of AI applications

# Black-box AI creates business risk for Industry



# What is Explainable AI?

## Black Box AI

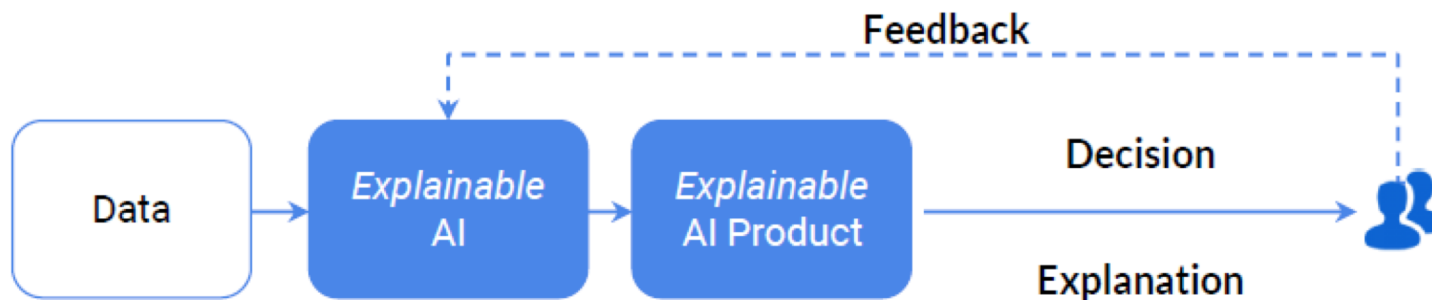


## Confusion with Today's AI Black Box

- Why did you do that?
- Why did you not do that?
- When do you succeed or fail?
- How do I correct an error?

---

## Explainable AI



## Clear & Transparent Predictions

- I understand why
- I understand why not
- I know why you succeed or fail
- I understand, so I trust you



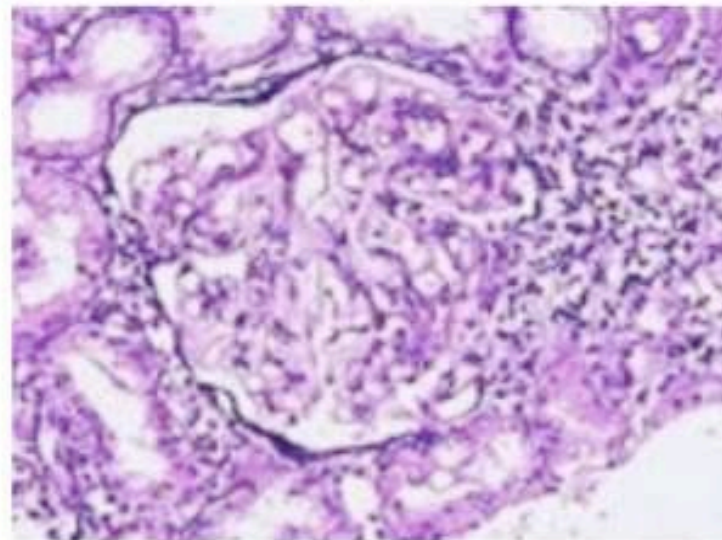
# Why Explainability: Verify the ML Model / System

Wrong decisions can be costly and dangerous

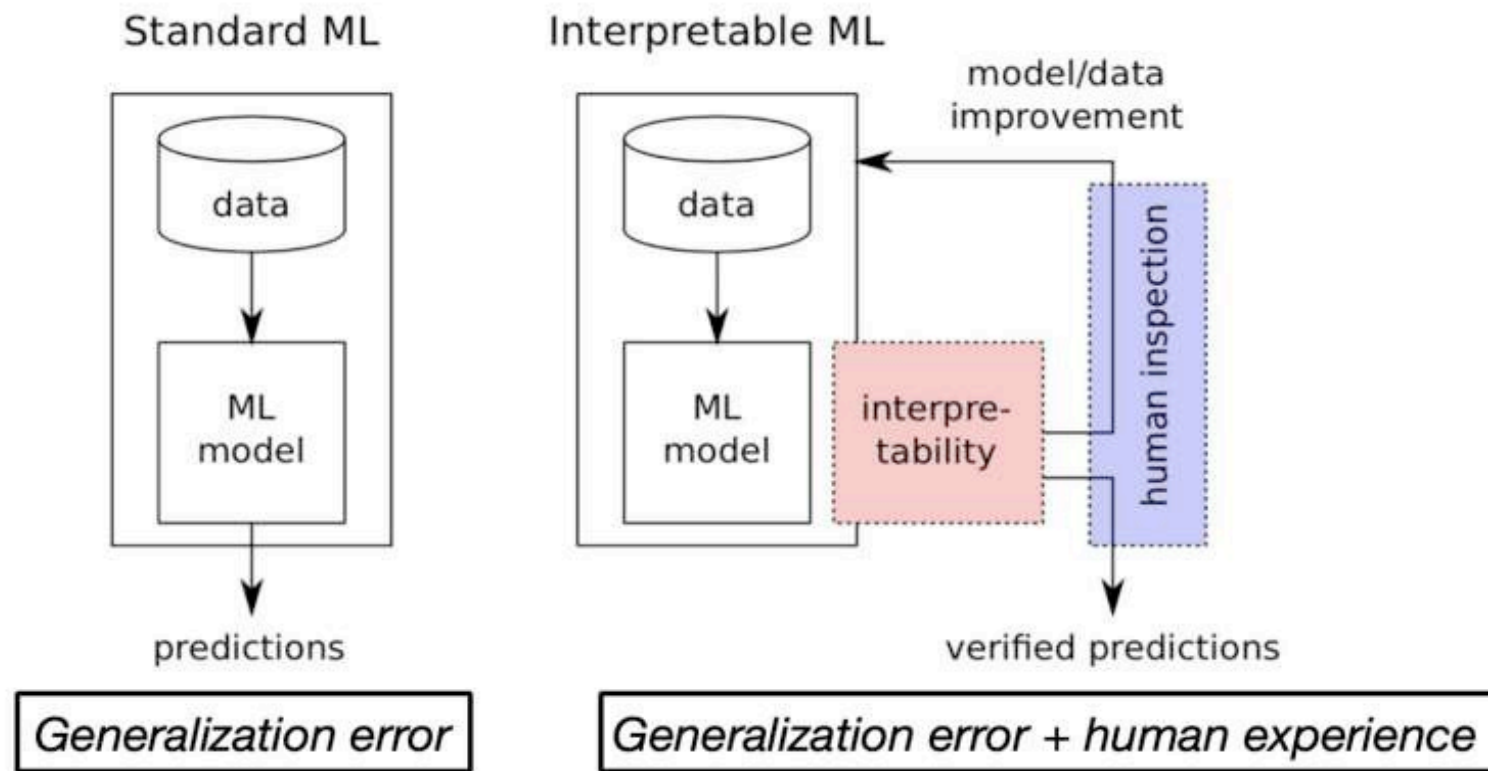
*“Autonomous car crashes, because it wrongly recognizes ...”*



*“AI medical diagnosis system misclassifies patient’s disease ...”*



# Why Explainability: Improve ML Model





## Motivation

### Two Mega Trends Impacting Cyber Security

#### Growth of Malware

With over **12 million** new malicious threats created every month by hackers around the globe, we have exceeded the capacity of the threat research community to identify, research and write signatures for each threat



#### Growth of Devices

With **50B devices** being connected to public and **private networks by 2020**, the attack surface has increase exponentially



# Three Cyber Security Problems Ripe for Artificial Intelligence

## Malware Detection

Leveraging the power of machine learning to detect and prevent zero-day and polymorphic malware across multiple threat vectors and platforms

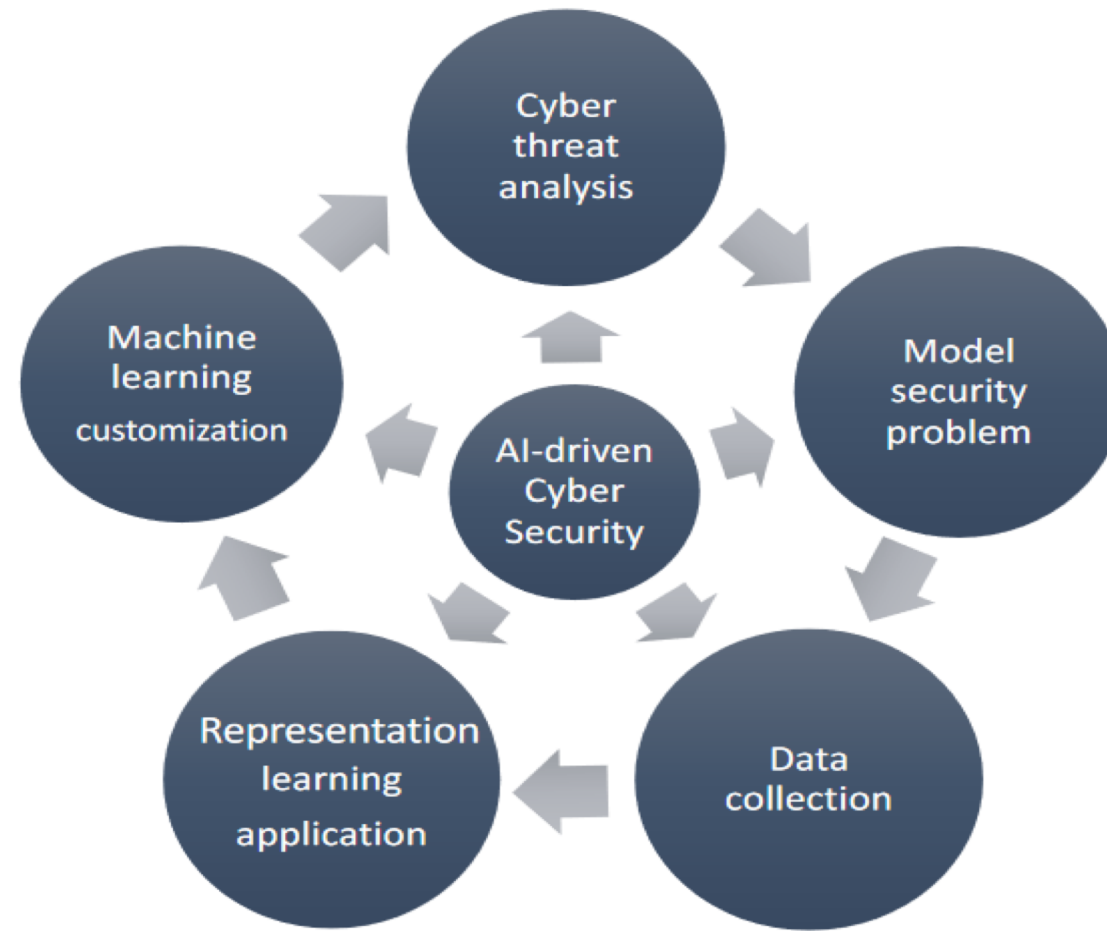
## Stream Analysis

Leveraging the power of machine learning to identify anomalous activity in system logs and network activity

## Threat Intelligence

Leveraging the power of machine learning and Natural Language Processing (NLP) to prioritize alerts and provide automated threat research

# Research Methodology



# Thank You



**Any Questions**