Siber Güvenlikte Makine Öğrenmesi

Doç. Dr. **Cihangir Tezcan**

MIDDLE EAST TECHNICAL UNIVERSITY

Informatics Institute, Department of Cyber Security
MIDDLE EAST TECHNICAL UNIVERSITY, ANKARA, TURKEY

Boğaziçi Üniversitesi DataCamp'22
13 Kasım 2022

CyBOK - The Cyber Security Body of Knowledge

The Cyber Security Body of Knowledge

1. Human, Organisational & Regulatory Aspects
2. Attacks & Defences
3. Systems Security
4. Software and Platform Security
5. Infrastructure Security

https://www.cybok.org/knowledgebase1_1

CyBOK - The Cyber Security Body of Knowledge

1. Human, Organisational & Regulatory Aspects

1. Risk Management & Governance
2. Law & Regulation
3. **Human Factors**
4. **Privacy & Online Rights**

CyBOK - The Cyber Security Body of Knowledge

2. Attacks & Defences

1. **Malware & Attack Technologies**
2. Adversarial Behaviours
3. **Security Operations & Incident Management**
4. Forensics

Challenges in AI-based Malware Detection

Challenges in AI-based Malware Detection

1. Results generally valid only for the used dataset

Challenges in AI-based Malware Detection

Challenges in AI-based Malware Detection

1. Results generally valid only for the used dataset
2. Performance

Challenges in AI-based Malware Detection

Challenges in AI-based Malware Detection

1. Results generally valid only for the used dataset
2. Performance
3. Encrypted data

Challenges in AI-based Malware Detection

Challenges in AI-based Malware Detection

1. Results generally valid only for the used dataset
2. Performance
3. Encrypted data
4. Static analysis might not be enough

Challenges in AI-based Malware Detection

Challenges in AI-based Malware Detection

1. Results generally valid only for the used dataset
2. Performance
3. Encrypted data
4. Static analysis might not be enough
5. Malware can detect that it is inside a virtual machine

CyBOK - The Cyber Security Body of Knowledge

3. Systems Security

1. Cryptography
2. Operating Systems & Virtualisation Security
3. Distributed Systems Security
4. Formal Methods for Security
5. **Authentication, Authorisation & Accountability**

CyBOK - The Cyber Security Body of Knowledge

4. Software and Platform Security

1. **Software Security**
2. Web & Mobile Security
3. Secure Software Lifecycle

CyBOK - The Cyber Security Body of Knowledge

5. Infrastructure Security

1. Applied Cryptography
2. **Network Security**
3. **Hardware Security**
4. Cyber Physical Systems
5. Physical Layer and Telecommunications Security

Network Security Example: Anomaly-Based Intrusion Detection

TABLE 21. Accuracy and false alarm rate scores for both data sets and models.

Data Set	Model	Accuracy	False Alarm Rate
Institutional	Ensemble Learning with SVM	0.9768	0.0010
	Ensemble Learning with KNN	0.9931	0.0010
	Ensemble Learning with Naive Bayes	0.9806	0.0011
	Ensemble Learning with Logistic Regression	0.9790	0.0010
	CNN	0.9968	0.0008
UNSW-NB15	Ensemble Learning with SVM	0.9902	0.0051
	Ensemble Learning with KNN	0.9960	0.0051
	Ensemble Learning with Naive Bayes	0.9818	0.0049
	Ensemble Learning with Logistic Regression	0.9896	0.0037
	CNN	0.9885	0.0041

E. Tufan *et al.*: Anomaly-Based Intrusion Detection by Machine Learning: A Case Study on Probing Attacks

Challenges in Anomaly Detection

Challenges in Anomaly Detection

1. Cannot convert IDS to IPS

Challenges in Anomaly Detection

Challenges in Anomaly Detection

1. Cannot convert IDS to IPS
2. Cannot analyze encrypted data

Challenges in Anomaly Detection

Challenges in Anomaly Detection

1. Cannot convert IDS to IPS
2. Cannot analyze encrypted data
3. There are limited datasets

Challenges in Anomaly Detection

Challenges in Anomaly Detection

1. Cannot convert IDS to IPS
2. Cannot analyze encrypted data
3. There are limited datasets
4. Institutional data cannot be shared

Challenges in Anomaly Detection

Challenges in Anomaly Detection

1. Cannot convert IDS to IPS
2. Cannot analyze encrypted data
3. There are limited datasets
4. Institutional data cannot be shared
5. Human crafted datasets are different than the real world data

Challenges in Anomaly Detection

Challenges in Anomaly Detection

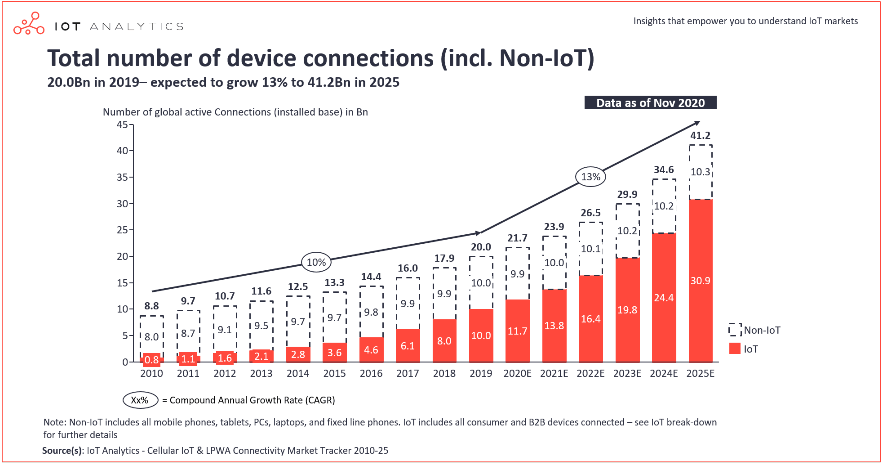
1. Cannot convert IDS to IPS
2. Cannot analyze encrypted data
3. There are limited datasets
4. Institutional data cannot be shared
5. Human crafted datasets are different than the real world data
6. Use AI to simulate user behavior

Challenges in Anomaly Detection

Challenges in Anomaly Detection

1. Cannot convert IDS to IPS
2. Cannot analyze encrypted data
3. There are limited datasets
4. Institutional data cannot be shared
5. Human crafted datasets are different than the real world data
6. Use AI to simulate user behavior
7. AI Wars: Attackers also train their AI according to your AI

Who is Responsible for IoT Security



IoT connections, surpassing non-IoT for the first time:
`https://iot-analytics.com/state-of-the-iot-2020-12-billion-iot-connections-surpassing-non-iot-for-the-first-time`
Instead of securing every device, current approach is the collect all data, process it and then detect problems

Big Data and AI/ML

Big Data

"Big data is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software" - *Wikipedia*

Big Data and AI/ML

Big Data

"Big data is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software" - *Wikipedia*

AI/ML for Pattern Recognition and Scoring

Artificial Intelligence and Machine Learning techniques can be used on big data

- to obtain patterns that are impossible to detect by human inspection

Big Data and AI/ML

Big Data

"Big data is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software" - *Wikipedia*

AI/ML for Pattern Recognition and Scoring

Artificial Intelligence and Machine Learning techniques can be used on big data

- to obtain patterns that are impossible to detect by human inspection
- to categorize data owners

Big Data and AI/ML

Big Data

"Big data is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software" - *Wikipedia*

AI/ML for Pattern Recognition and Scoring

Artificial Intelligence and Machine Learning techniques can be used on big data

- to obtain patterns that are impossible to detect by human inspection
- to categorize data owners
- to score data owners

The Instagram Ads Facebook Won't Show You

The Instagram Ads Facebook Won't Show You

- Technology is not built for you, it is built for your data

The Instagram Ads Facebook Won't Show You

The Instagram Ads Facebook Won't Show You

- Technology is not built for you, it is built for your data
- "Data is the new oil" - *Clive Humby*

The Instagram Ads Facebook Won't Show You

The Instagram Ads Facebook Won't Show You

- Technology is not built for you, it is built for your data
- "Data is the new oil" - *Clive Humby*
- Facebook collects everything about you from FB, Instagram, WhatsApp to sell visibility into people and their lives

The Instagram Ads Facebook Won't Show You

The Instagram Ads Facebook Won't Show You

- Technology is not built for you, it is built for your data
- "Data is the new oil" - *Clive Humby*
- Facebook collects everything about you from FB, Instagram, WhatsApp to sell visibility into people and their lives
- Advertiser's buy access to your personal data

The Instagram Ads Facebook Won't Show You

The Instagram Ads Facebook Won't Show You

- Technology is not built for you, it is built for your data
- "Data is the new oil" - *Clive Humby*
- Facebook collects everything about you from FB, Instagram, WhatsApp to sell visibility into people and their lives
- Advertiser's buy access to your personal data
- In May 2021, people at Signal bought some Instagram ads

The Instagram Ads Facebook Won't Show You

The Instagram Ads Facebook Won't Show You

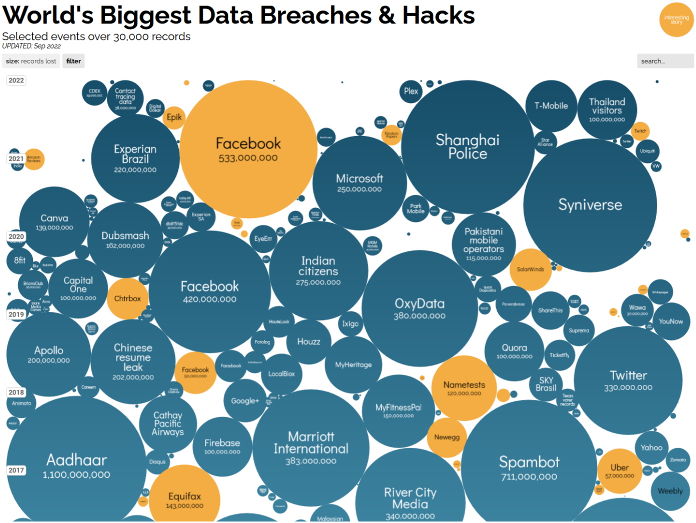
- Technology is not built for you, it is built for your data
- "Data is the new oil" - *Clive Humby*
- Facebook collects everything about you from FB, Instagram, WhatsApp to sell visibility into people and their lives
- Advertiser's buy access to your personal data
- In May 2021, people at Signal bought some Instagram ads
- As advertisements, they showed the user's personal data back to them

The Instagram Ads Facebook Won't Show You

The Instagram Ads Facebook Won't Show You

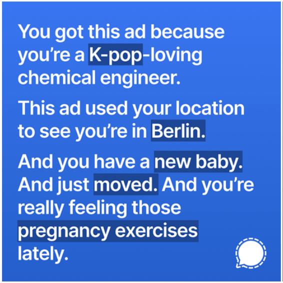
- Technology is not built for you, it is built for your data
- "Data is the new oil" - *Clive Humby*
- Facebook collects everything about you from FB, Instagram, WhatsApp to sell visibility into people and their lives
- Advertiser's buy access to your personal data
- In May 2021, people at Signal bought some Instagram ads
- As advertisements, they showed the user's personal data back to them
- But this transparency got them banned

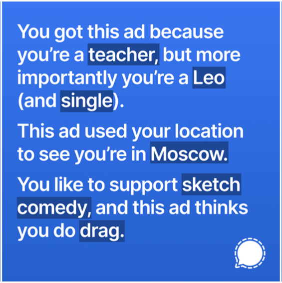
Big Data Means Big Data Breaches

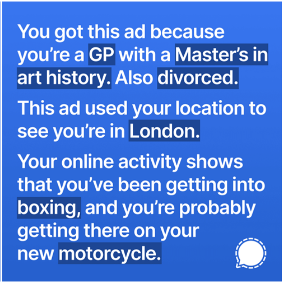


https://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks

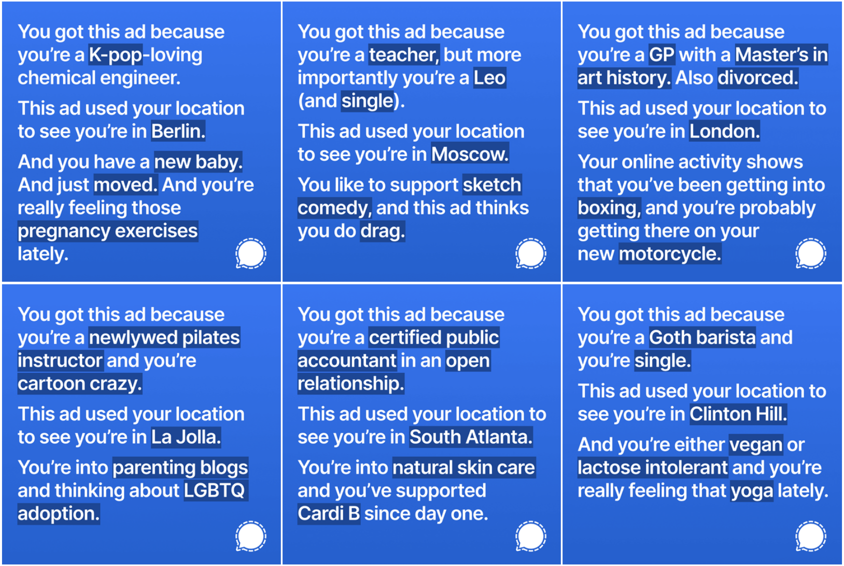
The Instagram Ads Facebook Won't Show You





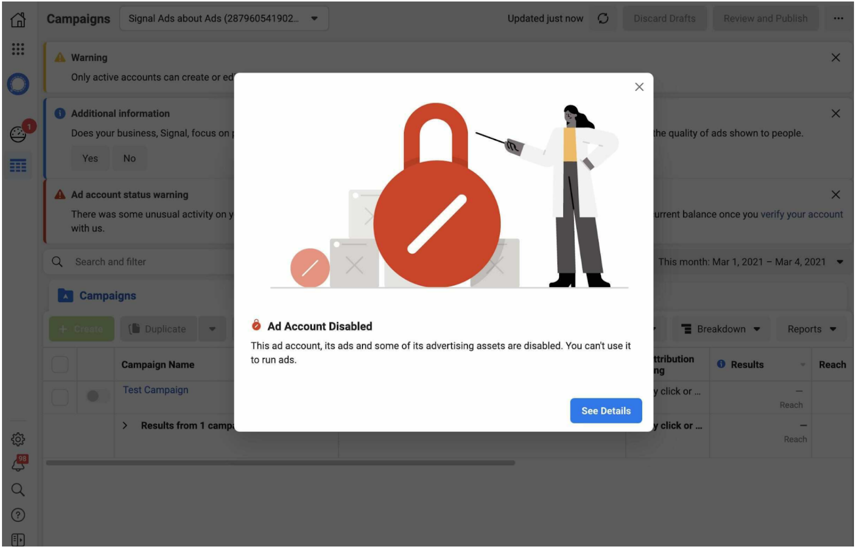


The Instagram Ads Facebook Won't Show You



https://signal.org/blog/the-instagram-ads-you-will-never-see

The Instagram Ads Facebook Won't Show You



https://signal.org/blog/the-instagram-ads-you-will-never-see

AI Explainability

AI Explainability

You should be able to explain design choices of the system and rationale for deploying it

AI Explainability

AI Explainability

You should be able to explain design choices of the system and rationale for deploying it

Example: Wrong Design Choices

- In December 2021, the Dutch Data Protection Authority announced a fine of 2.75 million euros against the Dutch Tax and Customs Administration

AI Explainability

AI Explainability

You should be able to explain design choices of the system and rationale for deploying it

Example: Wrong Design Choices

- In December 2021, the Dutch Data Protection Authority announced a fine of 2.75 million euros against the Dutch Tax and Customs Administration
- It was based on a GDPR-violation for processing the nationality of applicants by an ML algorithm in a discriminatory manner

AI Explainability

AI Explainability

You should be able to explain design choices of the system and rationale for deploying it

Example: Wrong Design Choices

- In December 2021, the Dutch Data Protection Authority announced a fine of 2.75 million euros against the Dutch Tax and Customs Administration
- It was based on a GDPR-violation for processing the nationality of applicants by an ML algorithm in a discriminatory manner
- The algorithm had identified double citizenship systematically as high-risk, leading to marking claims by those individuals more likely as fraudulent

AI Fairness and Bias

AI Fairness

- Fairness means that personal data needs to be handled in ways people would reasonably expect and not use it in ways that have unjustified adverse effects on them

AI Fairness and Bias

AI Fairness

- Fairness means that personal data needs to be handled in ways people would reasonably expect and not use it in ways that have unjustified adverse effects on them
- A practice will be considered unfair if it causes more harm than good

AI Fairness and Bias

AI Fairness

- Fairness means that personal data needs to be handled in ways people would reasonably expect and not use it in ways that have unjustified adverse effects on them
- A practice will be considered unfair if it causes more harm than good

Bias

- Bias can be addressed prior to training the algorithm, during model training, and post-processing

AI Fairness and Bias

AI Fairness

- Fairness means that personal data needs to be handled in ways people would reasonably expect and not use it in ways that have unjustified adverse effects on them
- A practice will be considered unfair if it causes more harm than good

Bias

- Bias can be addressed prior to training the algorithm, during model training, and post-processing
- It is unclear how to avoid bias in practice

Why Some AI/ML Algorithms Act Racist or Sexist?

Why Some AI/ML Algorithms Act Racist or Sexist?

- Many minority's credit application got rejected with a statement indicating that there was nothing wrong with their application

Why Some AI/ML Algorithms Act Racist or Sexist?

Why Some AI/ML Algorithms Act Racist or Sexist?

- Many minority's credit application got rejected with a statement indicating that there was nothing wrong with their application
- But it got rejected because people like "them" failed to pay their dues in the past

Why Some AI/ML Algorithms Act Racist or Sexist?

Why Some AI/ML Algorithms Act Racist or Sexist?

- Many minority's credit application got rejected with a statement indicating that there was nothing wrong with their application
- But it got rejected because people like "them" failed to pay their dues in the past
- So if two people with same credit scores apply for a credit, the results might depend on if they are white or black

Why Some AI/ML Algorithms Act Racist or Sexist?

Why Some AI/ML Algorithms Act Racist or Sexist?

- Many minority's credit application got rejected with a statement indicating that there was nothing wrong with their application
- But it got rejected because people like "them" failed to pay their dues in the past
- So if two people with same credit scores apply for a credit, the results might depend on if they are white or black
- But it is illegal to ask race in applications

Why Some AI/ML Algorithms Act Racist or Sexist?

Why Some AI/ML Algorithms Act Racist or Sexist?

- Many minority's credit application got rejected with a statement indicating that there was nothing wrong with their application
- But it got rejected because people like "them" failed to pay their dues in the past
- So if two people with same credit scores apply for a credit, the results might depend on if they are white or black
- But it is illegal to ask race in applications
- So the design choices are not racist, but then why the AI act racist?

Why Some AI/ML Algorithms Act Racist or Sexist?

Why Some AI/ML Algorithms Act Racist or Sexist?

- Many minority's credit application got rejected with a statement indicating that there was nothing wrong with their application
- But it got rejected because people like "them" failed to pay their dues in the past
- So if two people with same credit scores apply for a credit, the results might depend on if they are white or black
- But it is illegal to ask race in applications
- So the design choices are not racist, but then why the AI act racist?
- Because AI uses location info and the wealth is not equally distributed among the black and white people

Why Some AI/ML Algorithms Act Racist or Sexist?

Why Some AI/ML Algorithms Act Racist or Sexist?

- Many minority's credit application got rejected with a statement indicating that there was nothing wrong with their application
- But it got rejected because people like "them" failed to pay their dues in the past
- So if two people with same credit scores apply for a credit, the results might depend on if they are white or black
- But it is illegal to ask race in applications
- So the design choices are not racist, but then why the AI act racist?
- Because AI uses location info and the wealth is not equally distributed among the black and white people
- Thus, AI act racist because the data is biased

Privacy-Preserving Solutions

Privacy-Preserving Solutions

1. Use fully homomorphic encryption

Privacy-Preserving Solutions

Privacy-Preserving Solutions

1. Use fully homomorphic encryption
   - Currently you can evaluate small neural network efficiently in encrypted domain

Privacy-Preserving Solutions

Privacy-Preserving Solutions

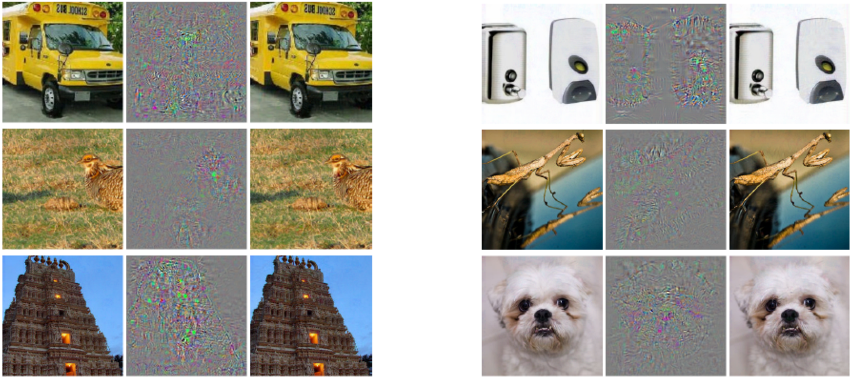
1. Use fully homomorphic encryption
   - Currently you can evaluate small neural network efficiently in encrypted domain
2. Use multi-party computation

Privacy-Preserving Solutions

Privacy-Preserving Solutions

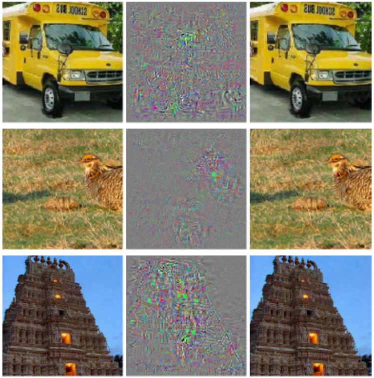
1. Use fully homomorphic encryption
   - Currently you can evaluate small neural network efficiently in encrypted domain
2. Use multi-party computation
3. Move your data from cloud to local solutions

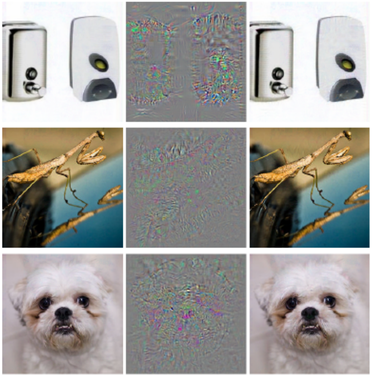
Adversarial Models (Data Poisoning)



Szegedy *et al.* *Intriguing properties of neural networks*, 2nd International Conference on Learning Representations (ICLR) 2014

Adversarial Models (Data Poisoning)





Szegedy *et al.* *Intriguing properties of neural networks*, 2nd International Conference on Learning Representations (ICLR) 2014
Sağdaki resimleri yapay zeka modelleri Deve Kuşu zannediyor

Adversarial Models

https://www.wired.com/story/machine-learning-backdoors

Teşekkürler

Teşekkürler