

Data Science for Cybersecurity Analysis

Julián Darío Miranda-Calle



Universidad
Pontificia
Bolivariana
SECCIONAL BUCARAMANGA

DPhi Data Science
Bootcamp





DPhi Data Science
Bootcamp

About me



Julián Darío
Miranda-Calle

Internal Auditor ISO 27001:2013, Cybersecurity Specialist, Computer Science Engineer, and Electronics Engineer. Currently, graduate and undergraduate professor at the Faculty of Computer Science Engineering.

Scrum Master, developer and researcher in cryptography, steganography, and steganalysis using Data Science, Machine Learning and Deep Learning techniques. Leader and Coach of the teams that will participate in the XXXIV ACIS/REDIS National Programming Contest 2020.



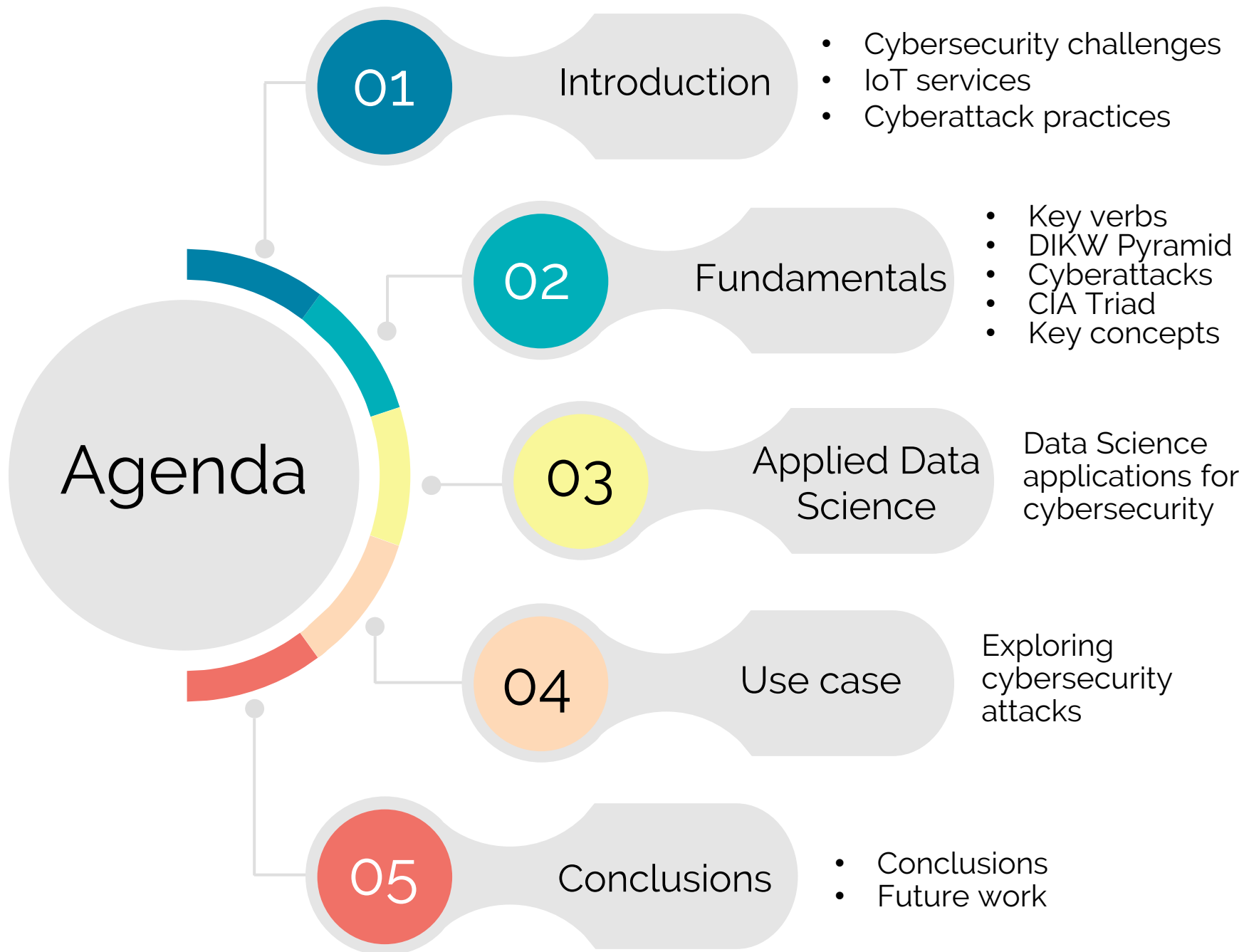
[linkedin.com/in/juliandariomiranda/](https://www.linkedin.com/in/juliandariomiranda/)



[0000-0002-7580-2361](https://orcid.org/0000-0002-7580-2361)



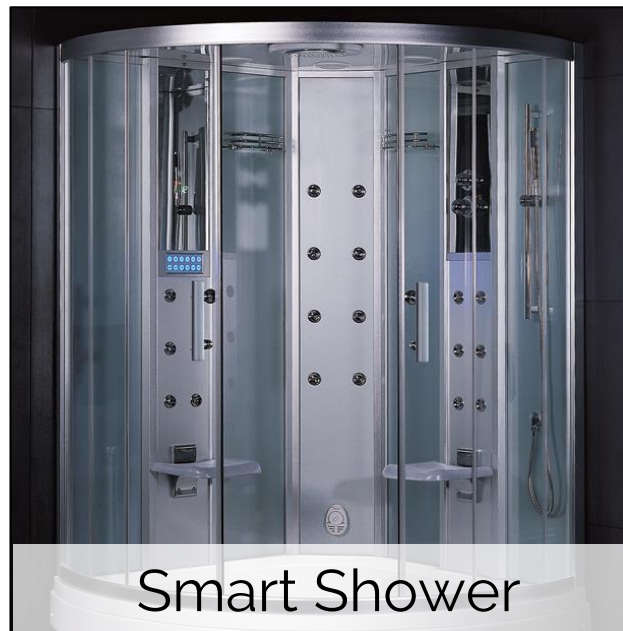
https://www.researchgate.net/profile/Julian_Miranda2





Introduction

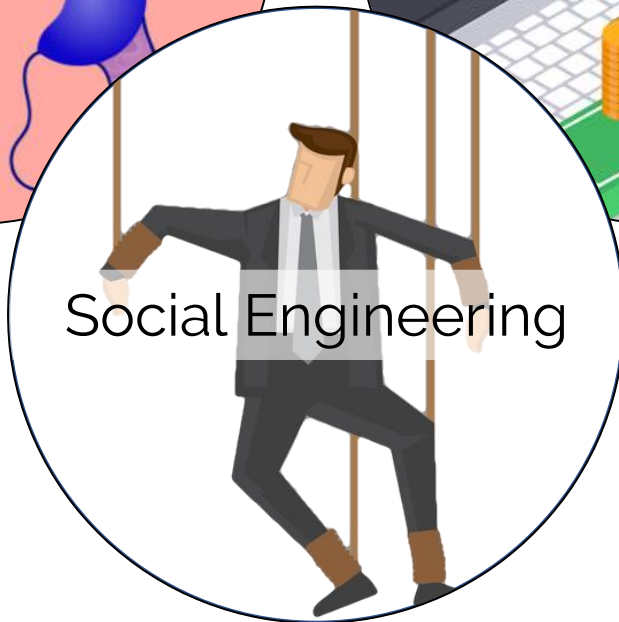
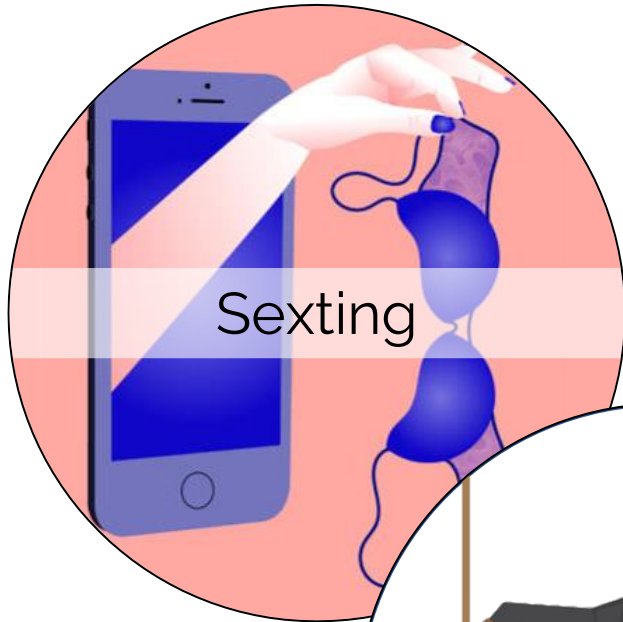
Cybersecurity challenges





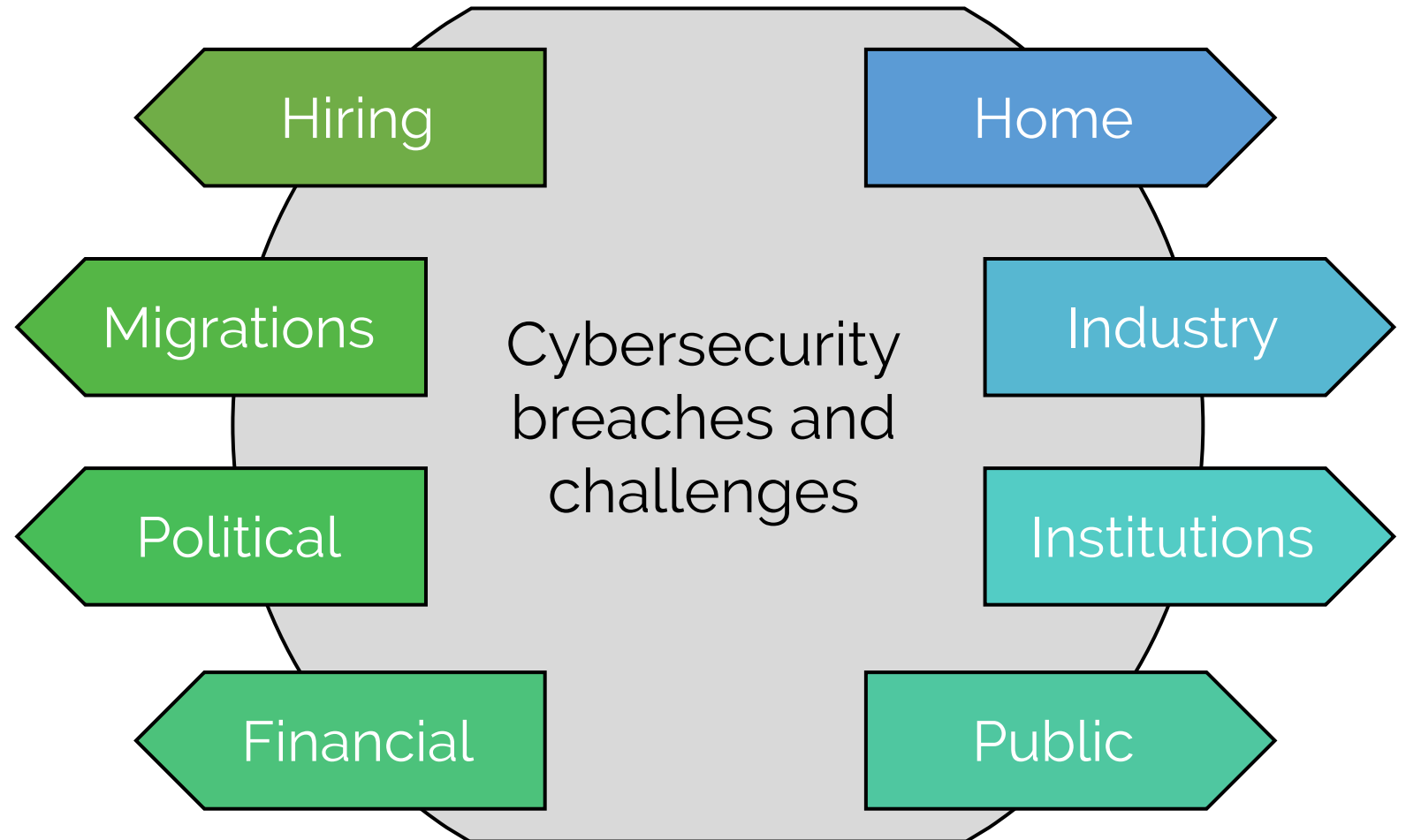
Introduction

Cybersecurity challenges



Introduction

Cybersecurity challenges





Introduction

Cybersecurity challenges



Cybersecurity/Cybercrime/Cyber...



Big Data, AI, ML, DL, RL, FL & BI



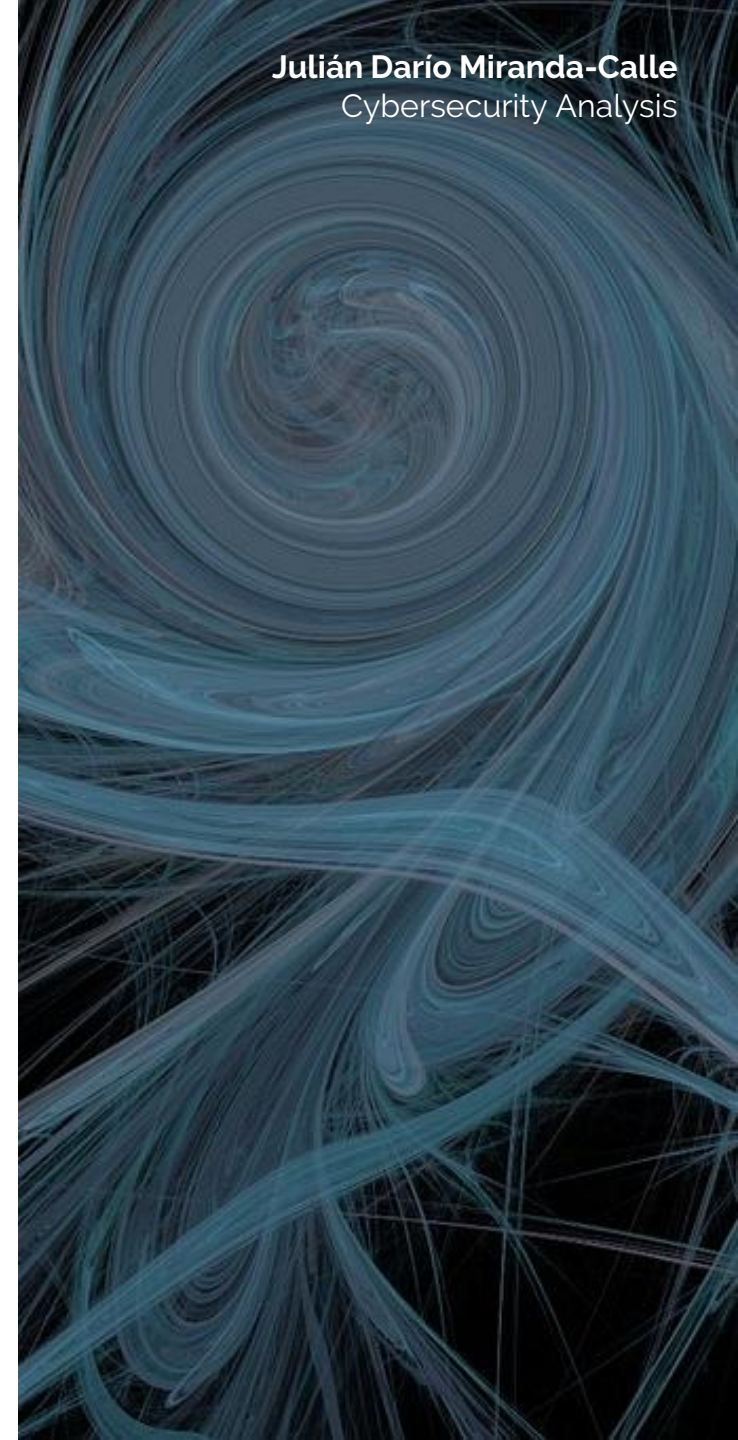
Quantum Computing



Lifi/Zigbee/MQTT/AMQP/CoAP



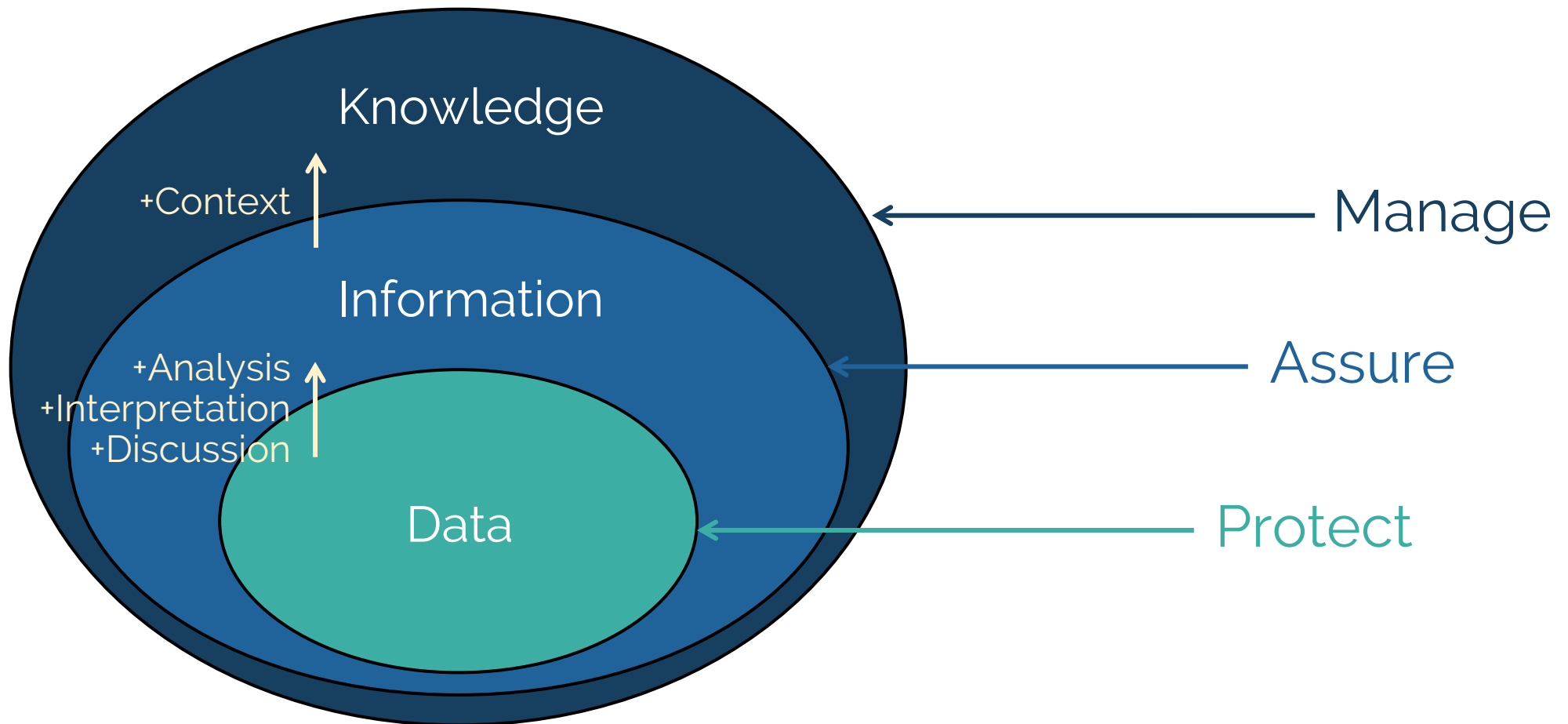
Cloud Computing, IoT, IIoT & IoE





Fundamentals

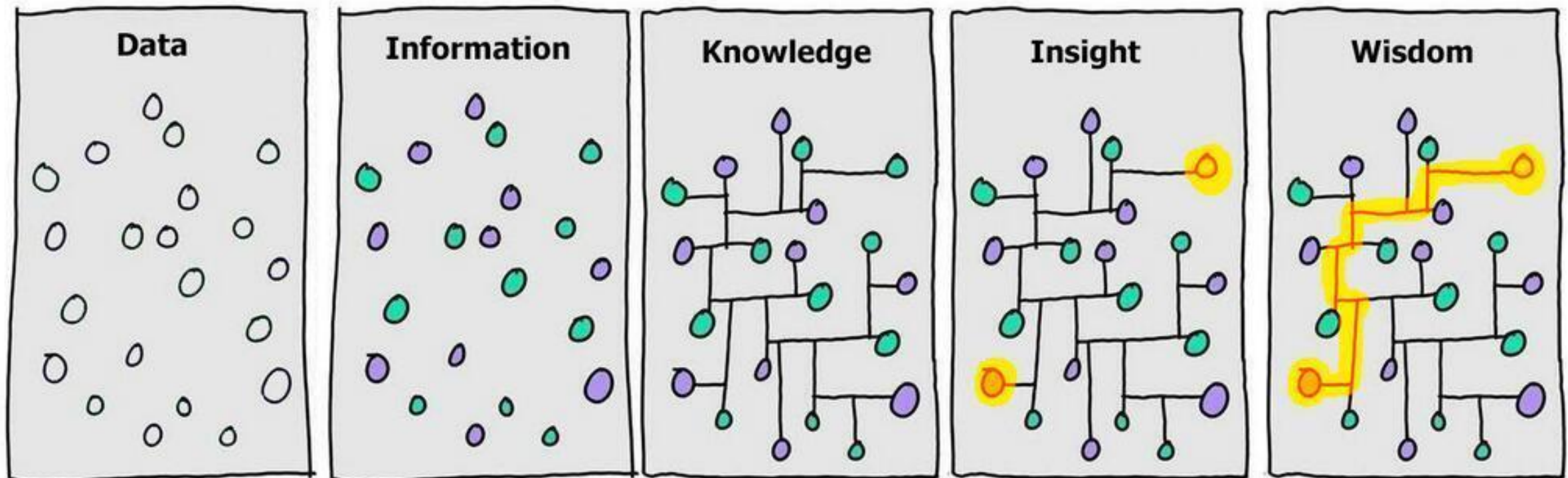
Key verbs in cybersecurity



Fundamentals

DIKW Pyramid

DIKW (Data, Information, Knowledge, Wisdom) Pyramid



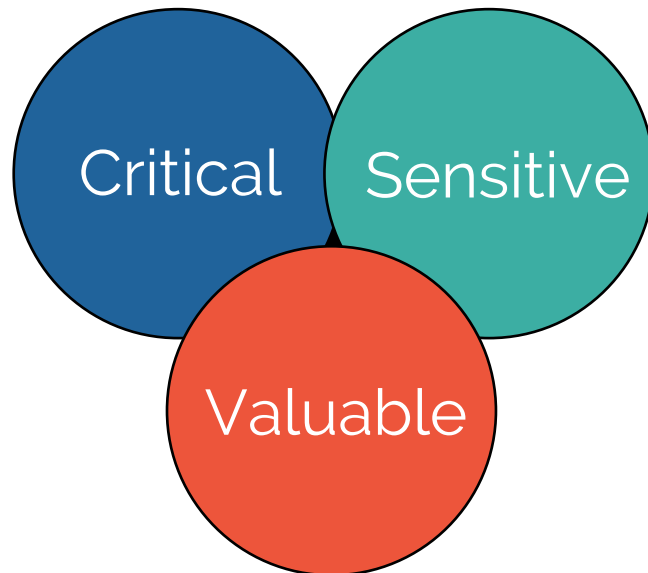
Fundamentals

Information properties

What makes the information so desirable to be accessed without permission, kidnapped and/or deleted?



Information is ...



Information = Competitiveness

Information = Advantage

Information = Power

Fundamentals

CIA Triad



Fundamentals

Anatomy of a cybersecurity attack

**01**

Reconnaissance

Steps taken to gather evidence on the targets to be attacked.

**02**

Scanning

Gather more technical in-depth information on the targets.

**03**

Gaining Access

Gain access to the targets through the vulnerabilities detected.

**04**

Keeping Access

Ensure the way back to the target machines for further procedures.

**05**

Covering Tracks

Conceal the success to avoid detection by security professionals.



Fundamentals

Key concepts in cybersecurity

Cyber attack

Any attempt to expose, alter, destroy, steal or gain unauthorized access of an asset.

Source port

Destination port

Endpoint (between [0, 65535]) to a logical connection that describes a service.

Logical port

Protocol

Set of rules to allow the transmission of information of two or more entities in a system.

IPv4 Address

IPv6 Address

Numerical label assigned to devices connected to a network that uses the Internet Protocol.

IP Address



Applied Data Science

How are data science practices applied in cybersecurity?

Risk assessment

Clustering of common risks to choose a treatment

Digital Forensics

To search for traces of file modification and information violation

Spam filtering

Using NLP techniques to detect language components in spam messages.

Phishing emails

Using NLP and Machine Learning techniques to unveil phishing attempts.

Event correlation

Relation of multiple events to identify common patterns

Data science in cybersecurity: applications

Most of the techniques are build on top of EDA, Regression, Classification, supervised and unsupervised learning tasks

Network monitoring

Identification of potential cyberattacks



Use case



Let's analyze a dataset of **real observations of cybersecurity attacks** to reveal some existing patterns



Source code and datasets available [here](#)

	Attack category	Attack subcategory	Protocol	Source IP	Source Port	Destination IP	Destination Port
0	Reconnaissance	HTTP	tcp	175.45.176.0	13284	149.171.126.16	80
1	Exploits	Unix 'r' Service	udp	175.45.176.3	21223	149.171.126.18	32780
2	Exploits	Browser	tcp	175.45.176.2	23357	149.171.126.16	80



Conclusions

The procedure explained in the use case can be **extrapolated for all types of cybersecurity studies** in which cybersecurity attack data records are kept, which can provide great **information on what patterns occur** and how to effectively ensure services and servers

Data science allows to **analyze the large volume of data** that is generated daily by IoT devices, industry, servers, home, among other contexts, to **propose new information security policies and countermeasures** for cybersecurity attacks to come.

Future work

Current research

Artificial Intelligence applied to Digital Forensic Analysis:
A preliminary review.

J. Miranda, J. Cano. 2019

LSB Steganography Detection in Monochromatic Still
Images using ANN.

J. Miranda, D. Parada. 2019

A study of the state of the art of NLP techniques for
Digital Forensics Analysis.

Y. Reddy, J. Miranda. 2020

DL Techniques Applied against Steganalysis of Digital
Images: A preliminary review.

M. Nabawi, J. Miranda. 2020

Detection of LSB steganography in digital still images
using Convolutional Neural Networks.

A. García, J. Miranda. 2020



DPhi Data Science
Bootcamp



Thanks!

Julián Darío Miranda



juliandariomiranda@gmail.com



0000-0002-7580-2361



/juliandariomiranda



www.researchgate.net/profile/Julian_Miranda2