



UNIVERSITAT POLITÈCNICA  
DE CATALUNYA

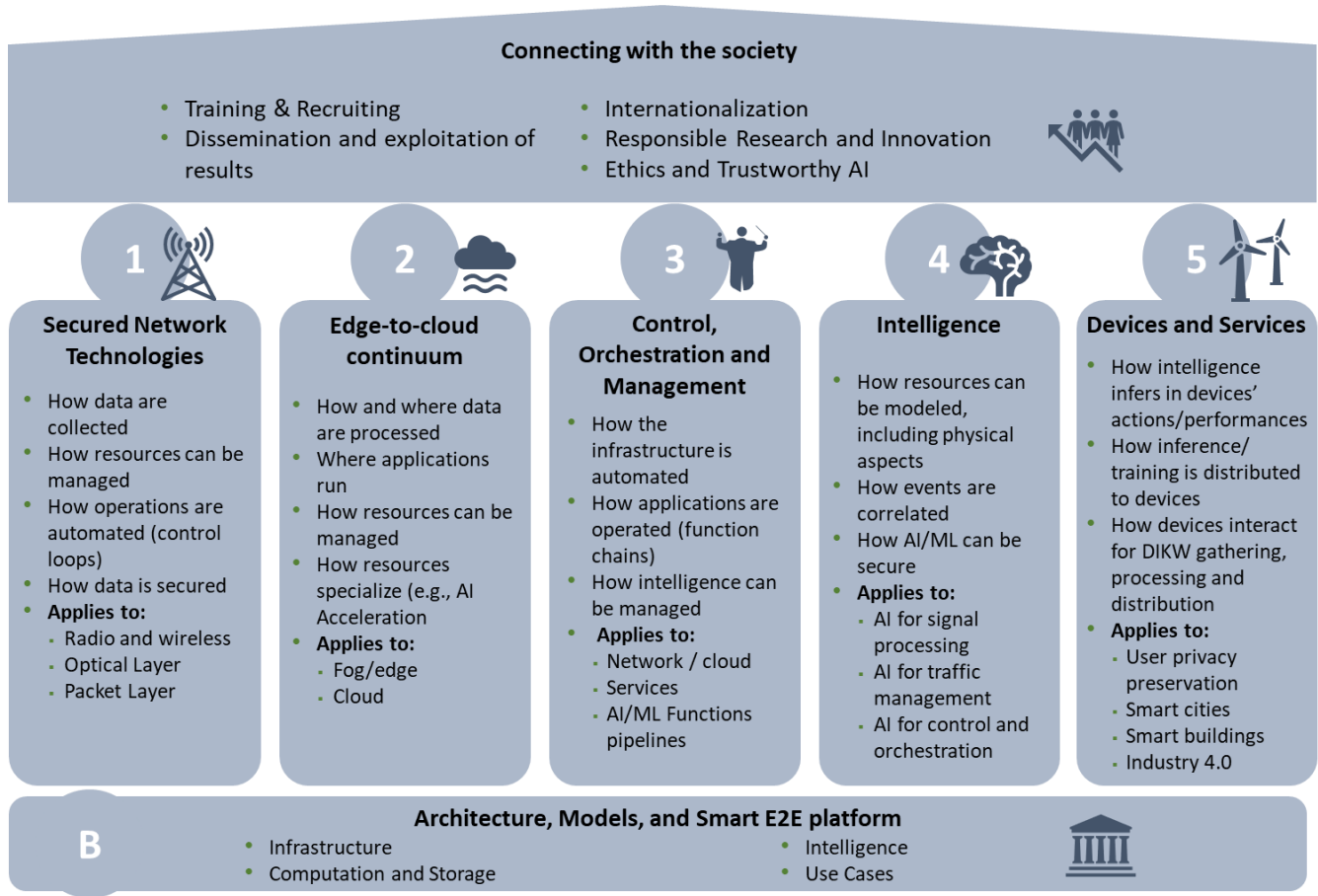
# Synthetic data for AI applications: a case for optical network modelling

Marc Ruiz, Ph. D.

*Advanced Broadband Communications Center (CCABA)  
Universitat Politècnica de Catalunya (UPC), Barcelona, Spain*

# CCABA

- CCABA - Centre de Comunicacions Avançades de Banda Ampla
  - <https://ccaba.upc.edu/en>



# Research topics and ongoing projects

- Optical Communications Group (GCO) – DAC
- Beyond 5G (B5G) / 6G networks:
  - Multi-Agent Systems (MAS) for autonomic networking
  - Digital Twin (DT) including mobile and optical networks
  - AI-based Quantum Communications
  - Optical Security
  - Deterministic Networks



# Let's play!

- Access site : <http://sli.do>
- Enter event code: 9733538
- Answer three questions



Joining as a participant?  [→](#)

## Your go-to interaction app for hybrid meetings

Engage your participants with live polls, Q&A, quizzes and word clouds  
— whether you meet in the office, online or in-between.

# Q1

- Have you ever trained and validated ML models using synthetic data?

- 20 votes

Yes, I did!



Not yet



mmm...I do not know..."they" gave me the data and I simply used them



I never trained a ML model before...



No! Come on! What kind of person do you think I am??



## Q2

- Imagine that you are training a Neural Network to predict ocean surface temperature. Which dataset you prefer (if you can choose only one)?
  - 19 votes

100,000 samples generated by "Ocean Dynamics Simulator" (from UPC)



10,000 samples obtained from "Ocean Digital Twin" (from MIT).



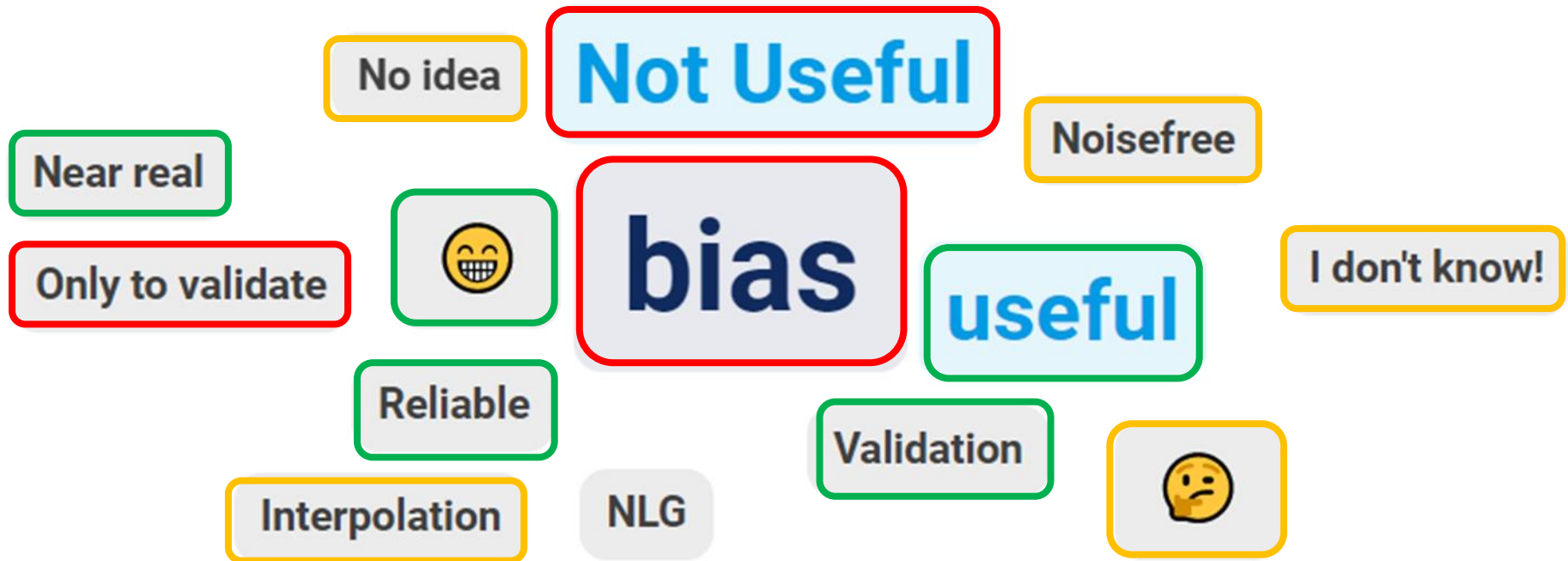
1,000 samples measured between November 2002 and May 2004 @ Indian Ocean





# Q3

- Write the word that better represents what do you think about synthetic data?
  - 18 votes



slido



# Synthetic Data 101

- Any production data applicable to a given situation that are not obtained by direct measurement
  - artificially manufactured (e.g. using models, simulators) rather than generated by real-world events
- Main use cases:
  - stand-in for test datasets of production or operational data
  - validate mathematical models
  - train ML models





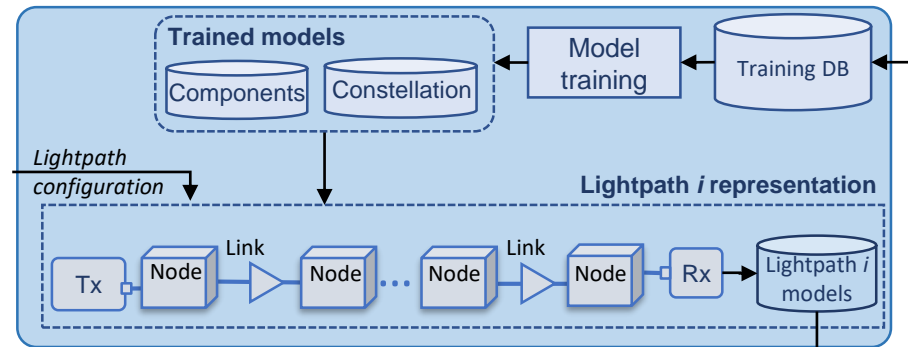
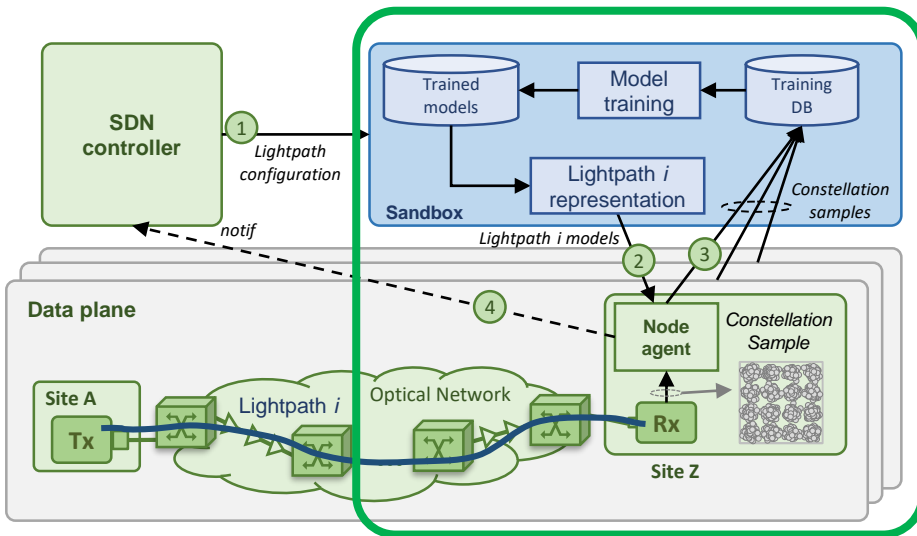
# Synthetic Data 101

- **Pros**
  - reducing constraints when using sensitive or regulated data
  - tailoring the data needs to certain conditions that cannot be obtained with authentic data
  - generating (large) datasets for software testing and quality assurance (including ML)
- **Cons**
  - inconsistencies or inaccuracies when trying to replicate complex scenarios.
  - inability to replace authentic data outright.

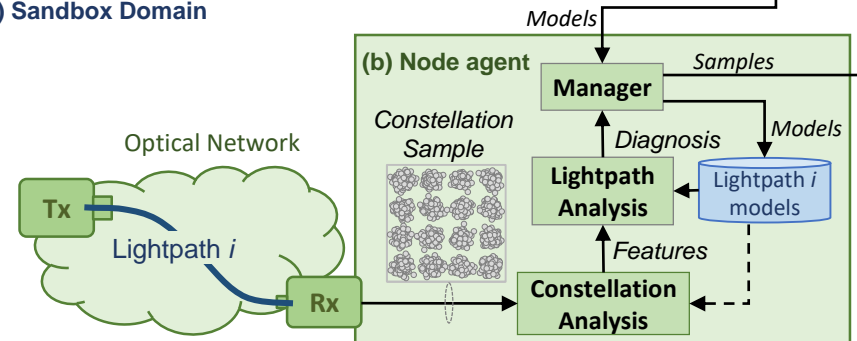
# Deep Learning -based Real-Time Analysis of Lightpath Optical Constellations

M. Ruiz, D. Sequeira, and L. Velasco, "Deep Learning -based Real-Time Analysis of Lightpath Optical Constellations [Invited]," IEEE/OPTICA Journal of Optical Communications and Networking (JOCN), 2022.

# AI-based Constellation Analysis (OCATA)

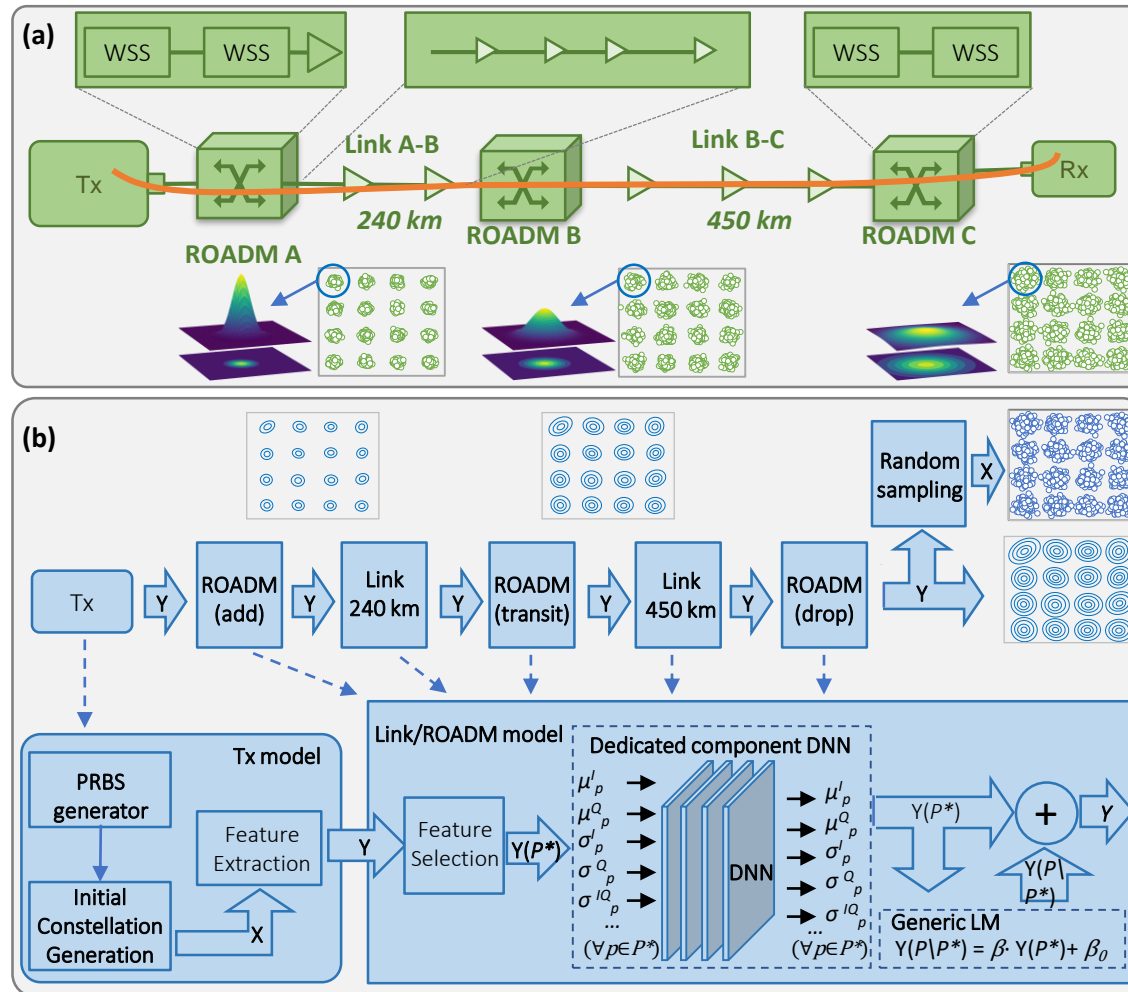


(a) Sandbox Domain

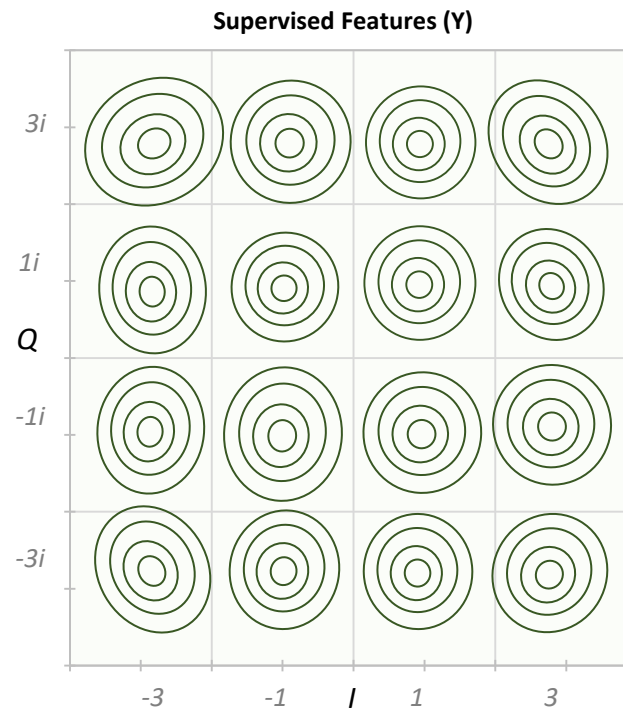
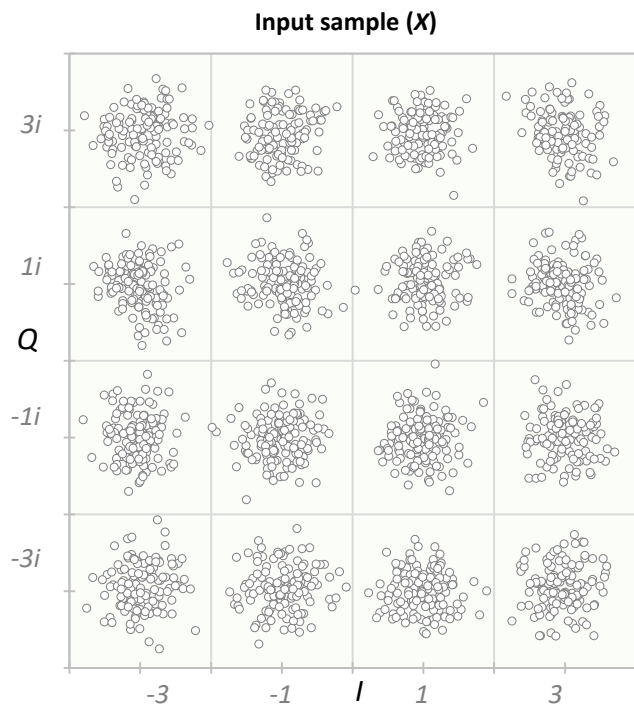


(b) Node agent

# OCATA



# Raw data and features



$P$	$\mu^I$	$\mu^Q$	$\sigma^I$	$\sigma^Q$	$\sigma^{IQ}$
$-3+3i$	-2.95	2.96	0.11	0.10	0.02
...	...	...	...	...	...
$3-3i$	2.99	-3.01	0.09	0.08	0.01

# Replication Data for OCATA

- **Openly available at CORA**
- Ruiz Ramírez, Marc; Velasco Esteban, Luis Domingo; Sequeira, Diogo Gonçalo, 2022, "Replication Data for: Optical Constellation Analysis (OCATA)", Repositori de Dades de Recerca, V1.
  - <https://doi.org/10.34810/data146>





UNIVERSITAT POLITÈCNICA  
DE CATALUNYA

# Thank you for your attention!

## Synthetic data for AI applications: a case for optical network modelling

Marc Ruiz, Ph. D.

*Advanced Broadband Communications Center (CCABA)  
Universitat Politècnica de Catalunya (UPC), Barcelona, Spain*