LINKSFOUNDATION.COM

FONDAZIONE LINKS

PASSION FOR INNOVATION

IL VALORE DEL DATO PER L'INTELLIGENZA ARTIFICIALE

Data Science for Industrial and Societal Applications Research Area

FABRIZIO DOMINICI – HEAD OF DATA SCIENCE AREA

ANFOV, 15 APRIL 2021

The Data Science Area

A **multidisciplinary** group of researchers focused on getting value from the whole data value chain

The research area of **Data Science for Industrial and Societal Application** focuses on **Artificial Intelligence** services for innovative solutions across several domains based on a **holistic data-driven** approach that span over the whole data chain.

Our final goal is to **bridge the gap between research and innovation** of product/services supporting the growth of companies and territories

KEY COMPETENCES

- ARTIFICIAL INTELLIGENCE
- DATA SPACES
- DATA DRIVEN SOLUTIONS
- UX & SERVICE DESIGN

KEY DOMAINS

- AGRICULTURE
- SECURITY
- INDUSTRY
- ENVIRONMENT
- MOBILITY
- AEROSPACE
- WELLBEING
- CULTURAL HERITAGE
- FINANCE
- INSURANCES
- ENERGY

20+ PEOPLE
32 average age
14 engineers

1.3 €M YEARLY BUDGET
78% EU
7% Commercial
15% Other
+50% Over 2014

213 PARTNERS
24 countries
56% Companies
24% Research
8% Public
12% Other

81 PROJECTS
20 EU Funded
44 Commercial
17 Other
7 Coordinated

STRATEGIC PARTNERS

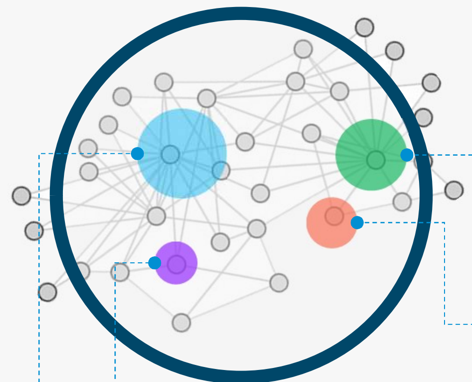
Microsoft, ENGINEERING, OSPEDALE SAN RAFFAELE, rcs, JRC EUROPEAN COMMISSION, Guala Closures Group, BDV BIG DATA VALUE ASSOCIATION, SIE, ERTICO ITS EUROPE

Technologies and Domains

A wide spectrum of knowledge and skills

ARTIFICIAL INTELLIGENCE

- Machine Learning
- Deep Learning
- Incremental Learning
- Federated Learning
- Transfer Learning & Domain
- Reinforcement Learning
- Embedded AI
- Natural Language Processing
- Ethics and Explainable AI
- Behavioral Analysis



DATA ENGINERING

- Hybrid & Federated Arch.
- Big Data Architectures
- Business Intelligence
- Full stack prototyping
- IoT & Wearable
- Decision Support Systems
- Crowdsourcing

DATA SPACES

- Geospatial Data
- Open & Linked Data
- Ontologies & Reasoning
- Interoperability & Data Provenance
- Heterogeneous data fusion
- Synthetic Dataset for AI

UX AND SERVICE DESIGN

- Human Centered Design
- Co-design
- Gamification
- User-data interaction

ENVIRONMENT
Sustainability – Circular
Disaster Resilience
Air Quality
Climate Change

WELLBEING
Health, Quality of Life,
Quality of Urban Spaces and Workspace

CULTURAL HERITAGE
Digital Technologies 4 Art
AI for Creative Industries

SECURITY
of public spaces
of digital spaces

AGRICULTURE
Smart Agriculture
Sustainable production

MOBILITY
AI to profile and change people behaviour

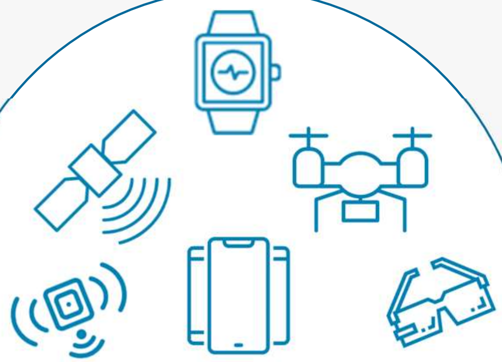
PEOPLE
AI meat Social Sciences
AI to support the migration flows

INDUSTRY & ENERGY
Factory of the Future
Sustainable production

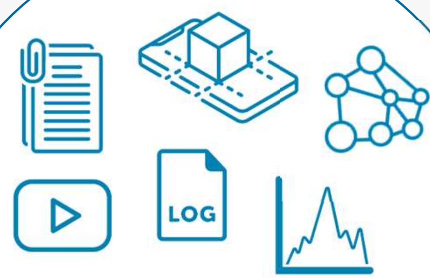
FINANCIAL SERVICES
Market prediction
AI on Blockchain & DLTs

Data are the fuel of the AI

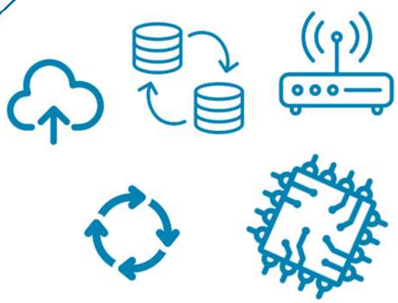
Technologies convergence enables the pervasiveness of data generation and processing



Convergence of **data generation** technologies



Pervasiveness of **data generation** capabilities



Advancement of **data processing** technologies

"It seems probable that once the machine thinking method had started, it would not take long to outstrip our feeble powers… Alan Turing

Technologies convergence enabled the modern AI

1950

1956 — John McCarthy coins the term AI

1986 — Backpropagation algorithm

1997 — First Convolutional Network (LeNet)

2012 — AlexNet, first deep learning model trained on GPU, beats human performance

2021

AI needs Data...

“No MARTINI No party”

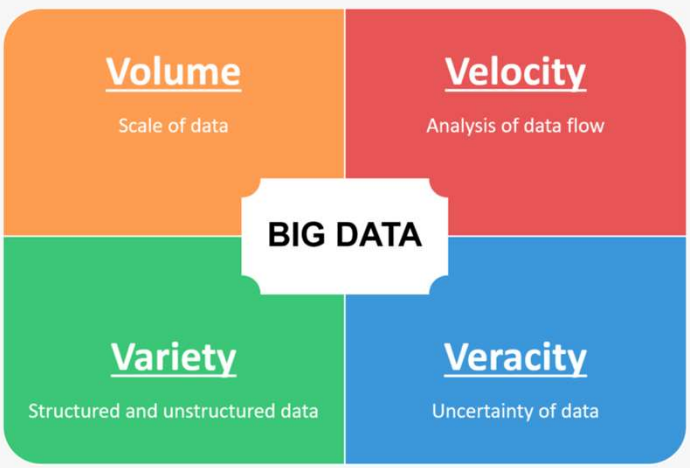
“No data, No party”

ARTIFICIAL INTELLIGENCE

Artificial intelligence is the **ability of a computer** or a robot controlled by a computer **to do tasks that are usually done by humans** because they require human intelligence and discernment.

MACHINE LEARNING

Machine learning is the study of **computer algorithms that improve automatically through experience and by the use of data**. It is seen as a part of artificial intelligence



AI needs for data, high-quality data….

Who owns the data is in a strategic position

Volume: the size of data
Velocity: the speed which the data is generated
Variety: the different types of data
Veracity: the trustworthiness of the data

+ the **Valuable richness data**
given by the temporal and spatial dimensions

And much more…

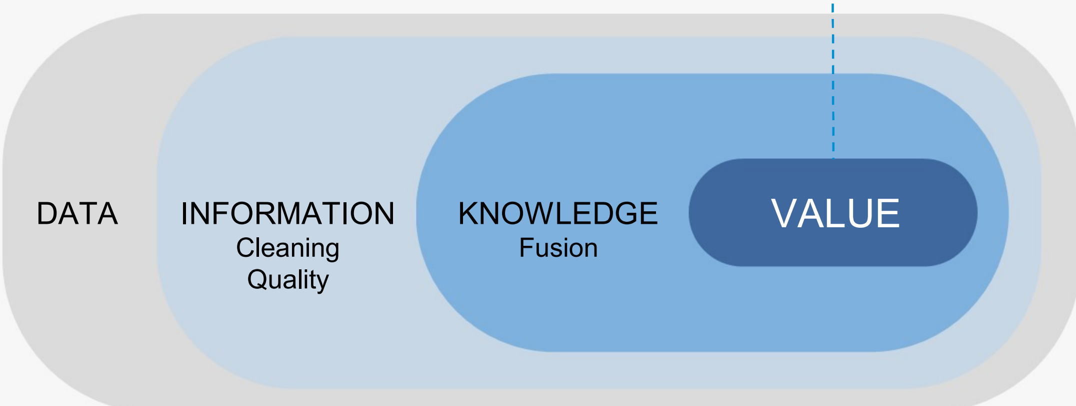
… and needs to move from Data to Value

ARTIFICIAL INTELLIGENCE

Artificial intelligence is the **ability of a computer** or a robot controlled by a computer **to do tasks that are usually done by humans** because they require human intelligence and discernment.

MACHINE LEARNING

Machine learning is the study of **computer algorithms that improve automatically through experience and by the use of data**. It is seen as a part of artificial intelligence



Adapted version of the DIKW diagram by N. Shedroff (1994). DIKW hierarchy originary devised by Ackoff and al. (1989)

MULTIDISCIPLINARITY

Partnerships between **domain** & **technology experts** plays a crucial role for the innovation process

DATA ALLIANCES

Alliances between **data owners** raise the data value

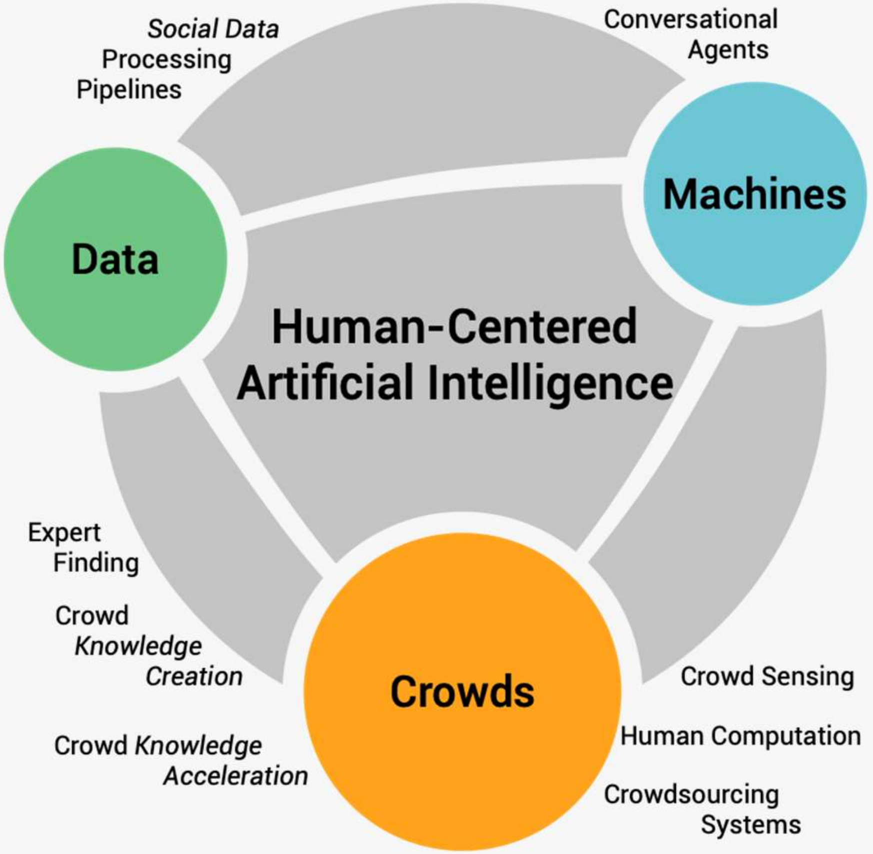
COMMITMENT

Domain experts' knowledge is crucial to obtain **proper labeling to feed AI**. Their commitment play a strategic role to set the path towards the **data valorization**

Human-Centered AI and Data Generation

The trend of convergence between technical and social sciences increases the availability of qualitative measures

- **Human-centered AI** learns from human inputs and collaboration. It is defined by systems that are:
  - continuously improving because of human input while
  - providing an effective experience between human and automated services
- People are **data generator** for AI algorithms and **final users** of the AI services

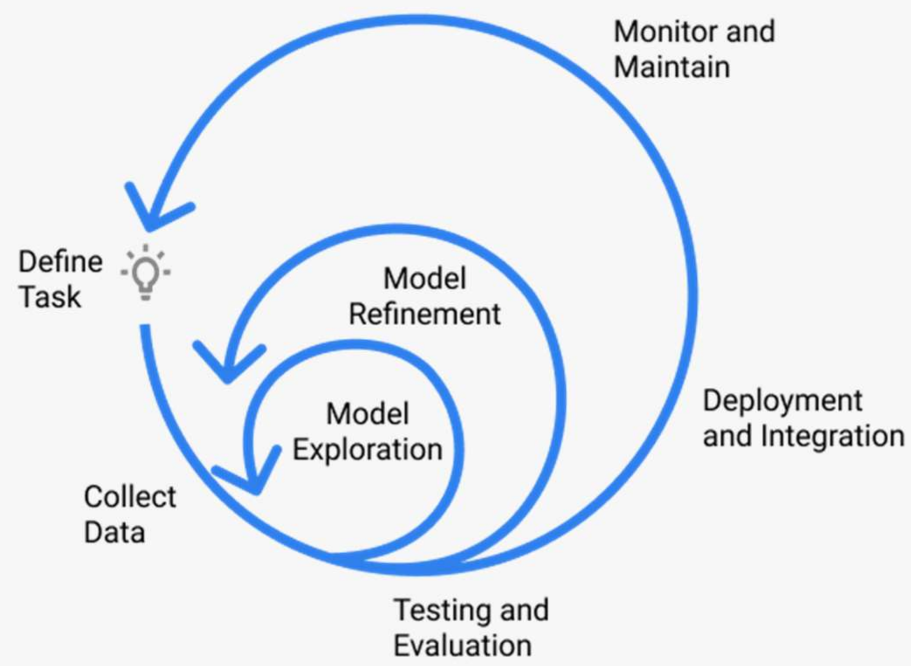


Persistent Improvement of AI Algorithms through new Data

Data are key for continuous learning processes

- Humans are the greatest producers of annotated data for incremental learning processes. **Humans continuously create new labeled** data through:
  - Spontaneous decision: manually labelling and classifying data
  - Unconscious decisions: selecting elements in preference algorithms, interacting on social media, etc...
- Newly created data are used to continuously fine-tune and improve AI algorithms in an **incremental learning process**. This is the foundation of a successful machine learning development lifecycle

Machine Learning Development Lifecycle



Persistent Improvement of AI Algorithms through new Data

Data are key for continuous learning processes

- Humans are the greatest producers of annotated data for incremental learning processes. **Humans continuously create new labeled** data through:
  - Spontaneous decision: manually labelling and classifying data
  - Unconscious decisions: selecting elements in preference algorithms, interacting on social media, etc...
- Newly created data are used to continuously fine-tune and improve AI algorithms in an **incremental learning process**. This is the foundation of a successful machine learning development lifecycle

Evolution Not Revolution

Innovation through different data sources

Selected research activities and outputs

Our innovation projects usually **combine heterogenous data** types and sources and aim to create positive economic and social impacts through **technology transfer** activity towards **enterprises** and **administrations**

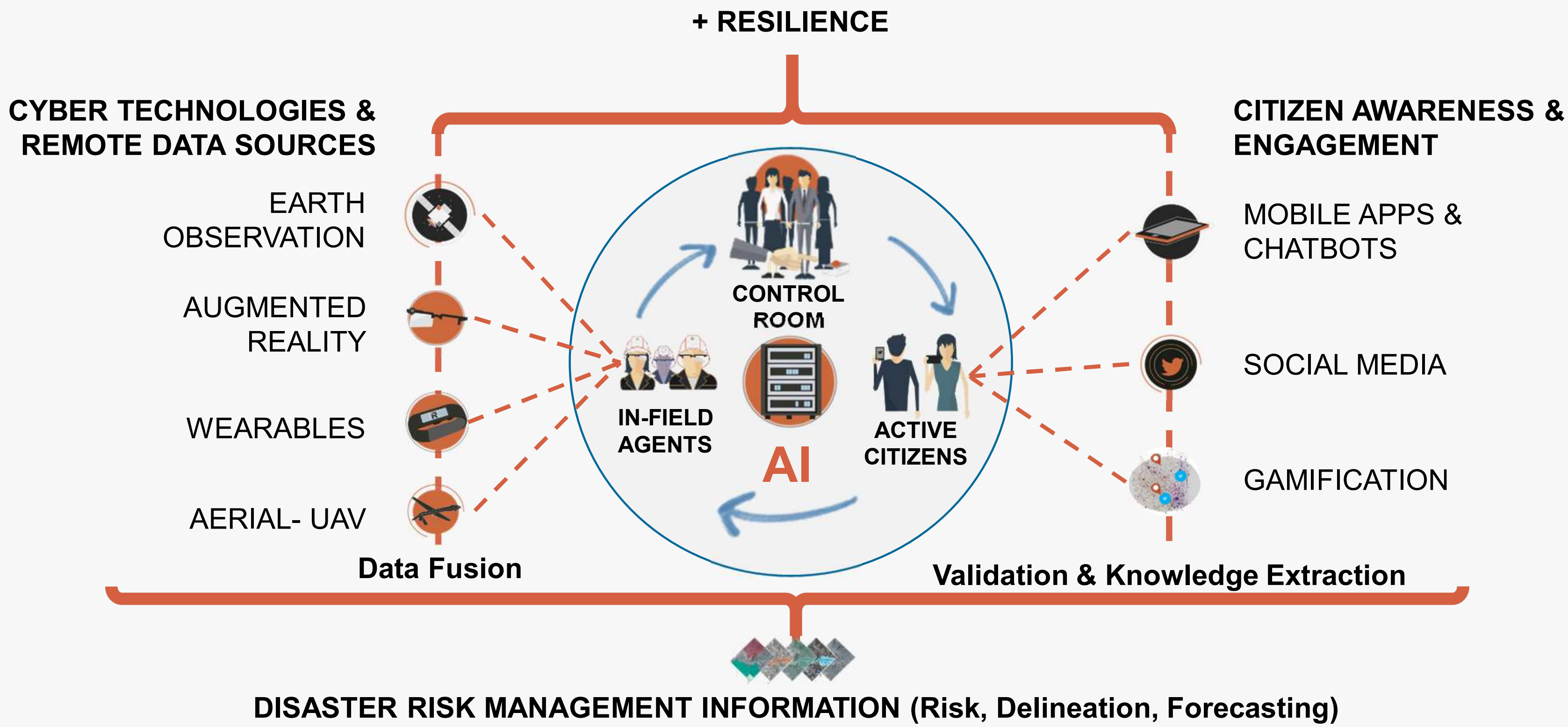
ENVIRONMENT AND DISASTER RESILIENCE

SECURITY AND INFRASTRUCTURE MONITORING

INDUSTRY

Environmental Resilience to Natural Hazards

Innovate the process by leveraging on heterogeneous data and top-notch AI to enhance situational awarness and forecast



DOMAINS

- Natural hazards
- Risk analysis
- Emergency Management

COMPETENCES

Unsupervised Machine Learning – Deep Learning, Decision Trees, SVM, clustering
NLP, word embeddings

FRAMEWORK/PROJECTS

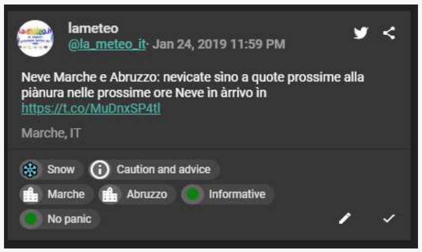
- FP7: Snowball, FLOODIS
- H2020: I-REACT, SHELTER, FASTER, SAFERS

FLOODIS
I-REACT
FASTER

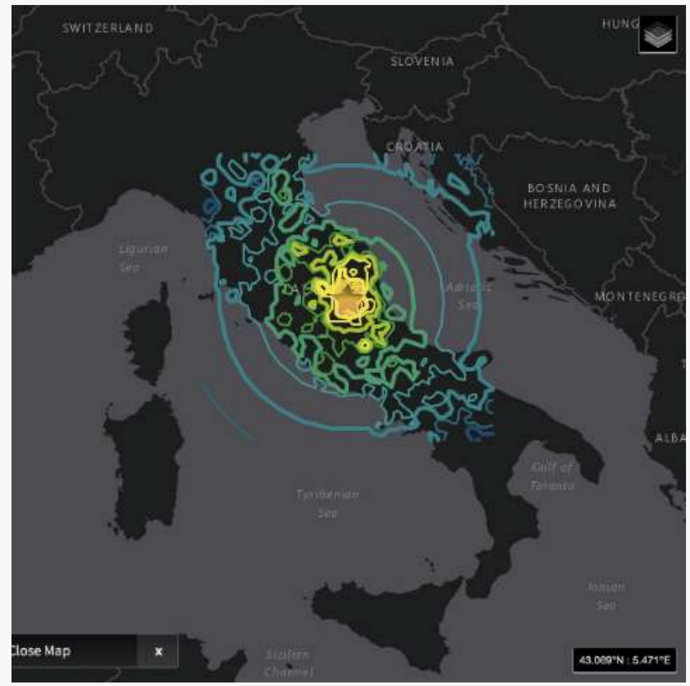
Human Generated Data

Text data from twitter used for events detection and combined with…

CLASSIFICATION: NLP PLUS MACHINE LEARNING



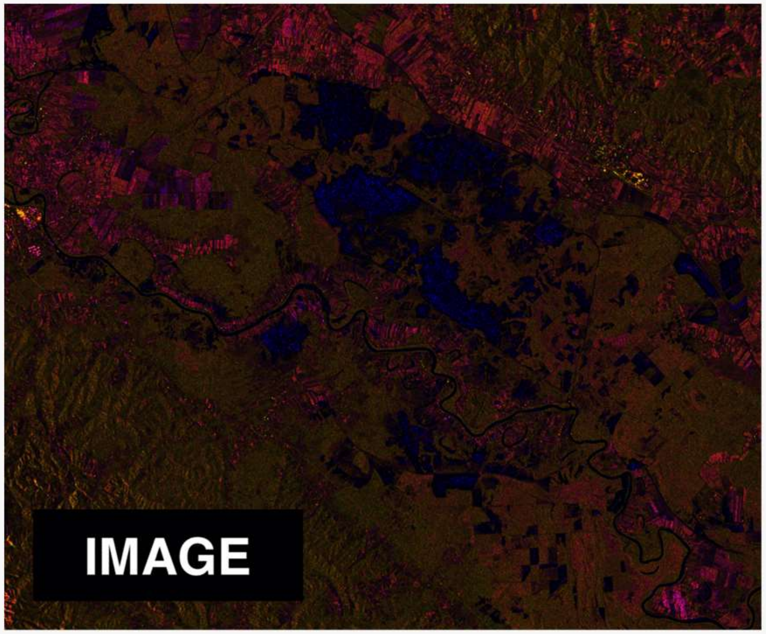
Topic detection
machine learning - clustering
Informativeness
binary classifier
Panic
sentiment analysis
Information type
multiclass classifier
Named Entity Recognition
Rule-based



AMATRICE EARTHQUAKE 24 AUGUST 2016

Delineating Flooded Areas from Satellite with AI

Using Sentinel-1 SAR data and Sentinel 2 images to delineate flooded areas through different AI model iterations



- **Data sources**
  - Copernicus EMS Sentinel-1: 2 bands, 30m/pixel, 5-6 days revisit time
  - Sen1Floods11 dataset composed of Sentinel-1 images of flooded areas

1. **Supervised delineation model – v1.0**
   - U-Net model trained on 48 images subset, with k-fold technique. Each fold consisted of 40 training images and 8 test images.
   - precision: **0.68** - recall: **0.83** - F1 score: **0.74**
2. **Supervised delineation model – v1.1**
   - U-Net model trained on an enlarged set of 196 images with a 0.75/0.25 split. Stronger data augmentation.
   - precision: **0.67** - recall: **0.83** - F1 score: **0.75**
3. **Supervised delineation model – v2.0**
   - U-Net model trained on Sen1Floods11 (available since Nov. 2020), 446 hand labeled images with a 0.75/0.25 split.
   - precision: **0.70** - recall: **0.50** - F1 score: **0.58**

DOMAINS
- Climate Change
- Natural Hazards
- Emergency Management

COMPETENCES

Supervised Machine Learning, clustering, **Deep Learning**, Computer Vision

RELETED PROJECTS
- FASTER
- SHELTER
- SAFERS

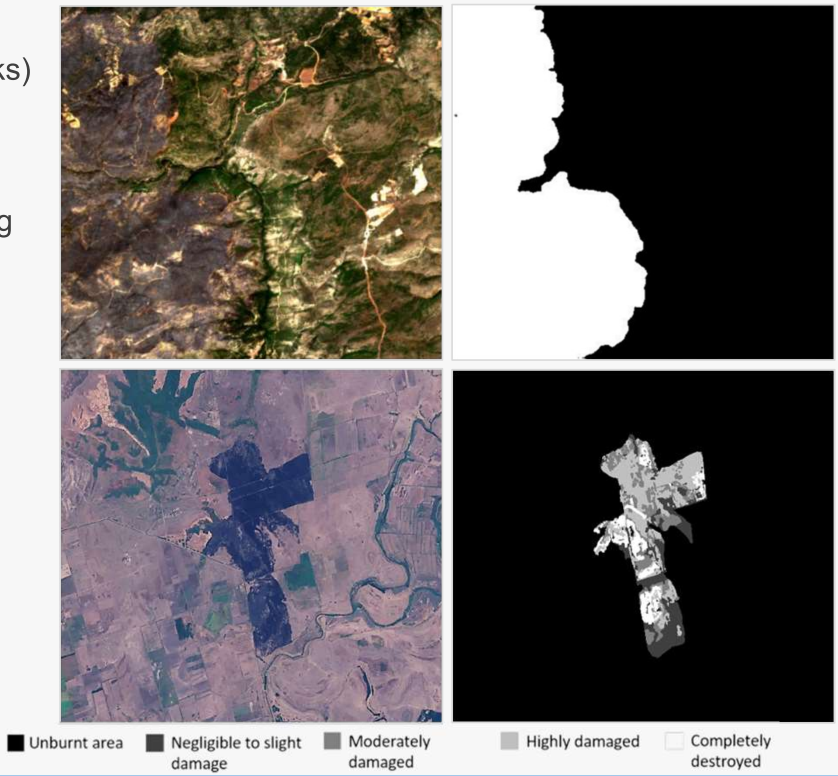
2020

Detecting Burned Areas from Satellite with AI

Using Sentinel-2 images to assess the damage after wildfire events through different AI techniques

- **Data sources**
  - Copernicus Emergency Management System (manually validated masks)
  - Sentinel-2: 12 bands, 10m/pixel, 5-6 days revisit time

1. **Unsupervised ML**
   - Before *and* after-wildfire pictures, Image processing and Self-Organising Maps (SOM)
   - precision: **0.81** - recall: **0.66** - F1 score: **0.70**
2. **Supervised delineation**
   - After-wildfire only pictures, Deep Learning approach (U-Net)
   - precision: **0.89** - recall: **0.78** - F1 score: **0.81**
3. **Supervised severity estimation**
   - After-wildfire only pictures, DL approach (Double-Step U-Net)
   - precision: **0.80** - recall: **0.97** - F1 score: **0.88**



Burned Area Delineation - ISCRAM (2020)
Severity Estimation - Journal, MDPI (2020)

DOMAINS
- Climate Change
- Natural Hazards
- Emergency Management

COMPETENCES
Unsupervised Machine Learning, clustering, **Deep Learning**, Computer Vision

RELETED PROJECTS
- FASTER
- SHELTER
- SAFERS

2020

Text and Video for Law Enforcement Agencies (LEA)

Leverage big data using deep learning models for LEAs threats detection tasks

Goals

- **Data Intelligence Platform** for real-time threat detection
- **Soft targets risk assessment** using web content, social media and on-site sensory data
- Increase **collaboration** between **private** and **public LEAs**

Why?

- Increasing number of criminal attacks on soft targets, showing lack of real-time situational awareness tools
- Lot of collected heterogenous data (CCTV, sensors, web, social media, UAV) not fully exploited **yet**

SOFT TARGETS RISK ASSESSMENT TASK

- **Terrorist groups** are known to use weakly controlled social media platforms to organize themselves and find new members across the globe. (**Telegram** has 500 million active users/month)

PERSON RE-IDENTIFICATION TASK

- **Person Re-Identification** aims at matching people across **non-overlapping** camera views distributed at distinct locations. This is a key tool for LEAs surveillance tasks
- NVIDIA labs at CVPR2019 shown efforts into generating synthetic images of pedestrians wearing different clothes, helping models be appearance-independent
- DG-Net presented by NVIDIA has **Rank@1** score of **77.2** (**13% better** than other models on MSMT17 dataset)

DOMAINS

- Surveillance of public/private spaces
- Threat intelligence
- Cybersecurity

AI TOPICS

Deep Learning, Computer Vision, Natural Language Understanding, Domain Adaptation, Information Retrieval

FRAMEWORK

H2020 APPRAISE project

H2020 STARLIGHT project

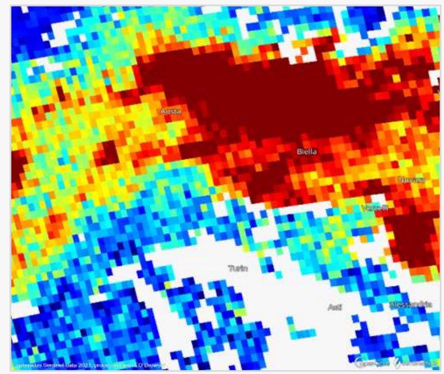

ATP
FINALS
TORINO
2021-2025

Air Quality Predictions and Forecasts

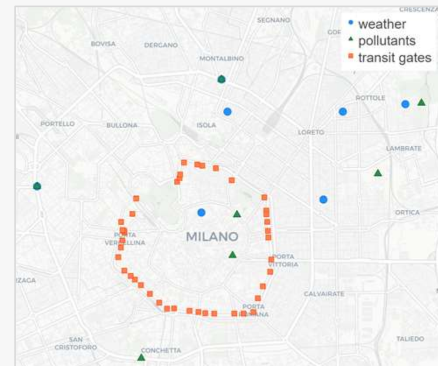
Monitor our environment from ground sensors to satellite feed

- Air pollution remains crucial for its implications on **health** and **environment.**
- High quality ground sensors are **expensive**: only a handful of systems are installed, even in big cities (fig. 2)
- Especially in urban areas, pollutant levels are directly linked to traffic trends and weather conditions → **pollutants can be estimated** from contextual information.
- Air quality varies both **temporally** and **spatially**: satellite feed becomes an invaluable resource for short and long-term estimates.
- **Sentinel 5P** (ESA's air quality monitoring satellite):
  - **Cons**: lower resolution, lower precision
  - **Pros**: high availability, lower costs, high coverage, daily revisits
- **Next steps**: combining **AirQuality** with people **wellbeing** and **perception**, **urban planning** and **GHG emission assesment**

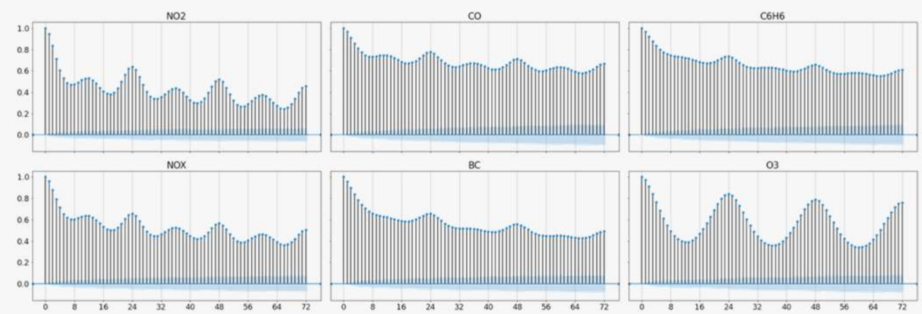
[uAQE – Urban Air Quality Estimator](), Springer (2019)
[Air quality estimation - comparative analysis](), MDPI (2020)



1. NO2 concentration over North-west Italy (S-5P)



2. Pollutant sensors in Area C, Milan



3. Autocorrelation for pollutants in the Milan area over a 72-hours window

DOMAINS
- Climate Change
- Natural Hazards
- Emergency Management

COMPETENCES
Deep Learning, Decision Trees, SVM, Time Series, Forecasting, Earth Observation

FRAMEWORK/PROJECTS
- Internal research

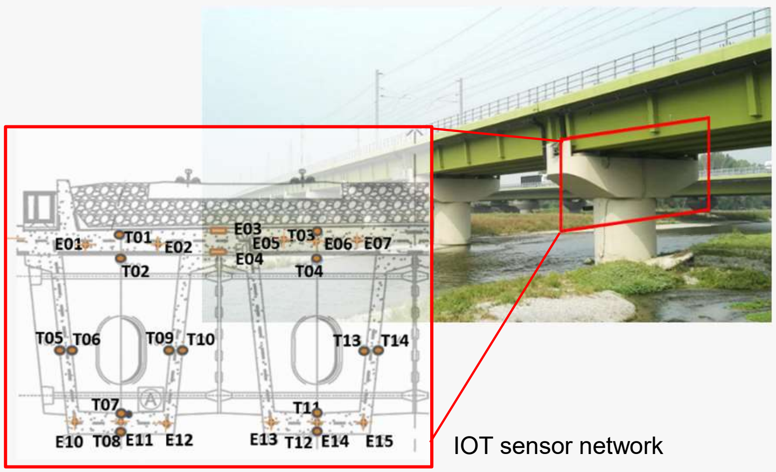
Monitoring of Critical Infrastructures

Machine learning approach to the safety assessment of a railway bridge

Machine Learning regression models to detect the damage of railway bridge using a network of IOT sensors.

Input

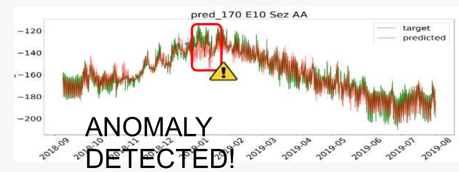
IOT sensors data – Oglio railway bridge



IOT sensor network

PIPELINE

- IOT sensors data: Temperatures Deformations
- ML models use Temperature to estimate the Deformations: ML ESTIMATED
- Comparison between Estimated Deformations and Deformations ESTIMATED
- If the real signal deviates from the estimated one:



DEFORMATION	MINIMUM DEVIATION FROM THE REAL SIGNAL DETECTED
Assessment of static performance of the deck	10%
Oversight of piers movements	5%
Piers rotation & joint expansion	10%

DOMAINS

- Infrastructure monitoring
- IOT

AI TOPICS

Machine Learning – Regression techniques, Neural Networks, Random Forest

FRAMEWORK

Joint Research with PoliTo

Closer look: Aerial Imagery and EO

Detecting solar panels from images with higher resolution

Huge growth of photovoltaic installations over the last decades: **maintaining a census** is tough.

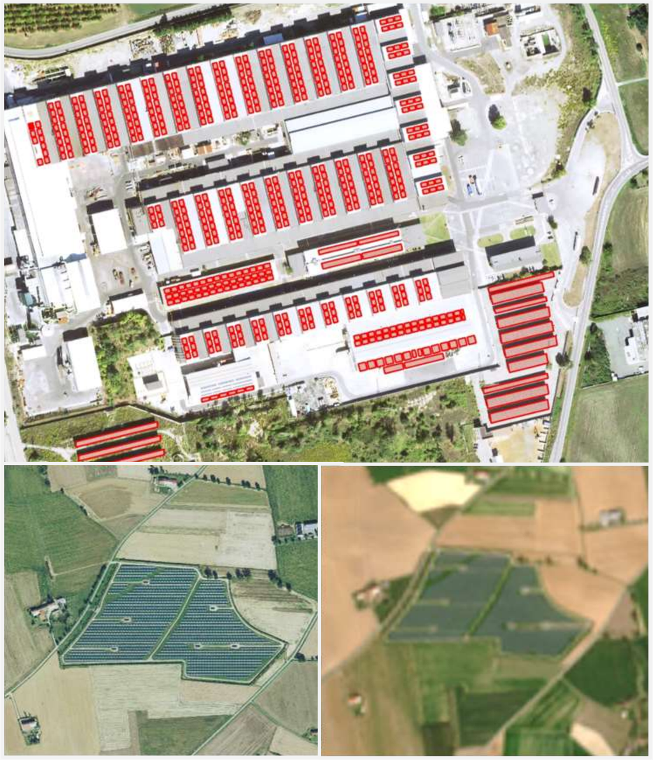
Machine Learning can automate such tasks, exploiting aerial photos and estimate the production by considering exposition, solar radiation etc.

Solar panel delineation from aerial images:

- 105 orthophotos, 30 cm/pixel (Asti and Alessandria, CGR S.P.A.)
- 300+ industrial photovoltaic plants, manual annotations (ITHACA)
- Thousands of non-recorded commercial and domestic installations

Objectives:

- locate and delineate individual panels, then estimate energy production over a given period
- Exploit aerial data to enhance satellite feed (super-resolution)



- Solar panel annotations (top)
- Aerial image (bottom left) versus Sentinel-2 imagery (right)

DOMAINS
- Climate Change
- Renewable Energy

COMPETENCES
Deep Learning, Computer Vision, Earth Observation

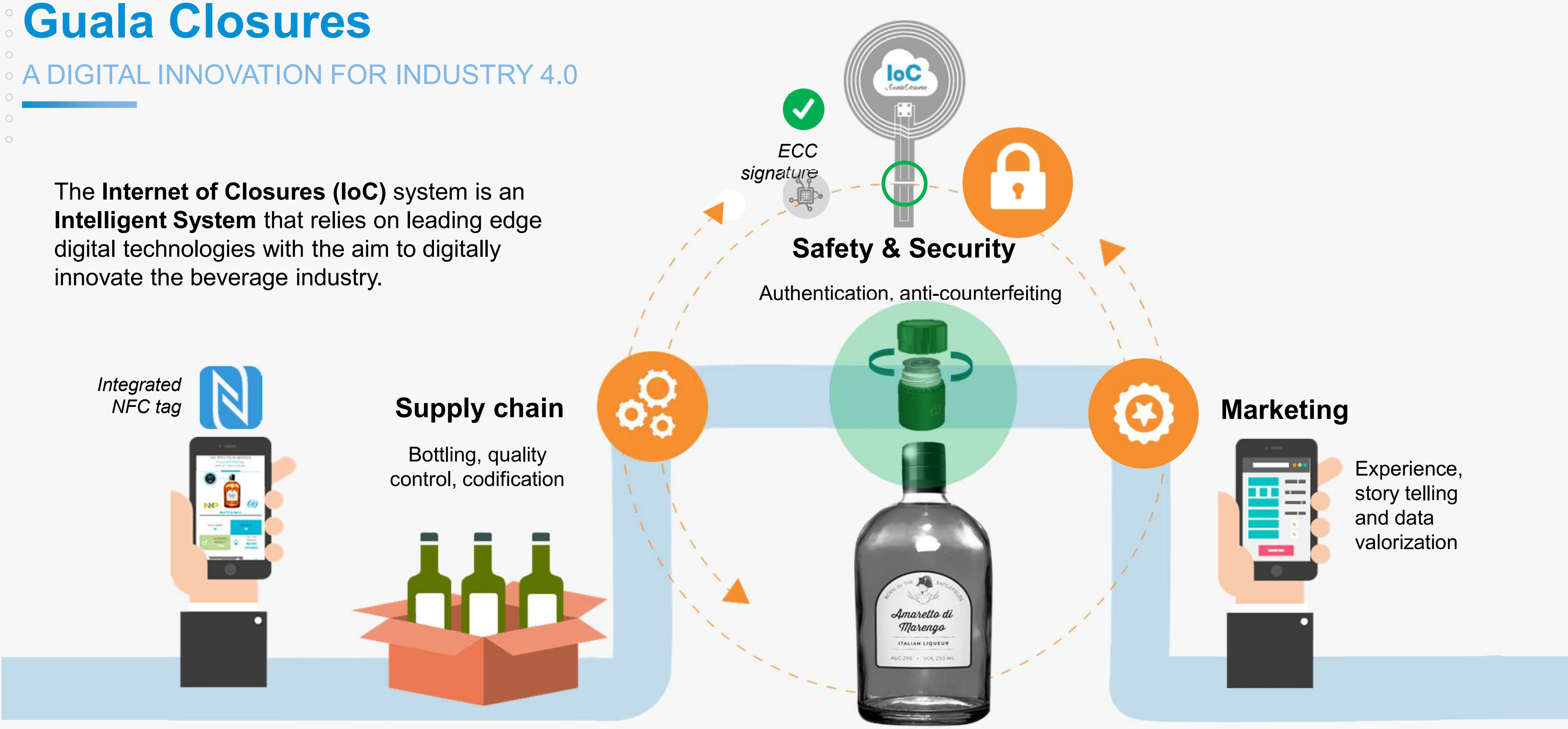
FRAMEWORK/PROJECTS
DYDAS

#Industry4.0 #NFC #Azure #BigData

Guala Closures

A DIGITAL INNOVATION FOR INDUSTRY 4.0

The **Internet of Closures (IoC)** system is an **Intelligent System** that relies on leading edge digital technologies with the aim to digitally innovate the beverage industry.



DOMAINS

- Food, Data Management, Supply chain

NUMBERS

4 Partners, 36 months

Budget 1 M€

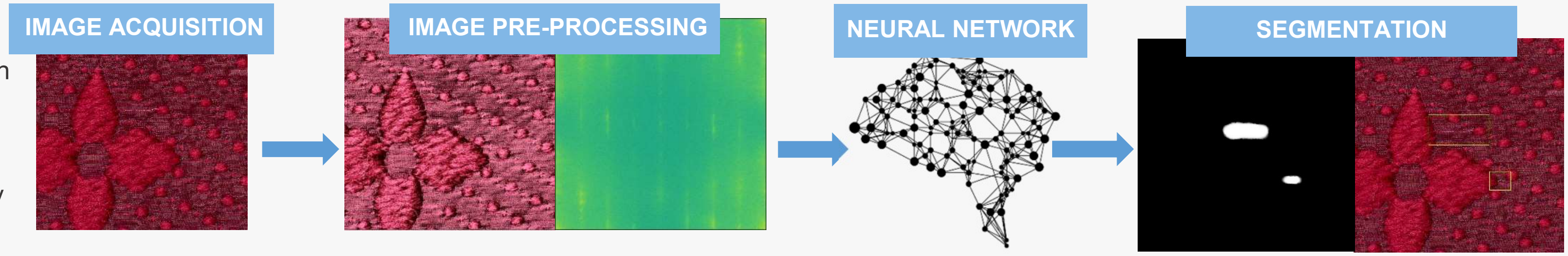
LINKS 153 k€ (Data Science)

LINKS ROLE

WP Leader – IoT platform, Big data analysis and data fusion, Blockchain PoC

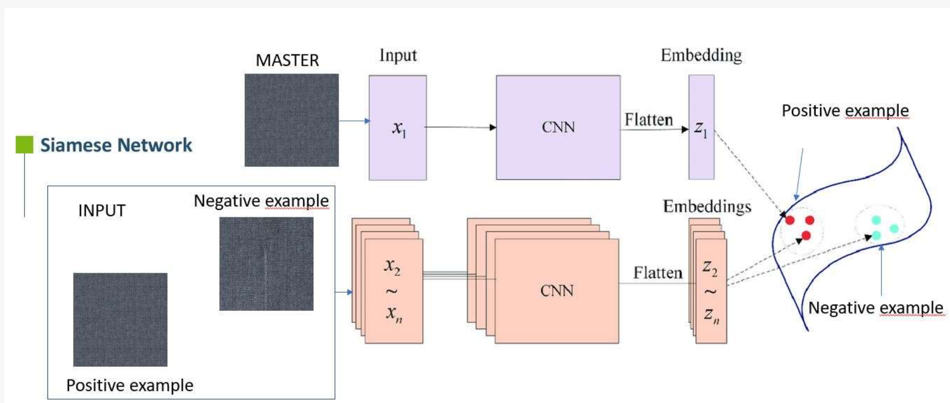
Textile Quality Evaluation

Image Vision and Artificial Intelligence applied to the high fashion textile quality assurance



Textile quality assurance for high fashion companies today is completely performed manually by extremely specialized operators, due to the complexity and the challenges of task. Quality check is a slow process that implies a careful analysis of fabrics.

Machine Learning models applied to Image Vision can support these operations and **speed up the process** by a significant amount, while also increasing the accuracy of operators in detecting small defects.



- **Master – target** approaches
- **Segmentation**
- **Image Vision** pre-processing techniques
- **Continuous Learning**

DOMAINS
- Industrial
- Textile
- Quality Assurance

COMPETENCES

Unsupervised Machine Learning – Deep Learning, Image Vision

FRAMEWORK/PROJECTS
- Collaboration with Quality Biella and Pegaso Sistemi

2020

ABS – AI Based Credit Risk Scoring

Merge financial data with proxy variables to feed top notch Machine Learning algorithms

Goals

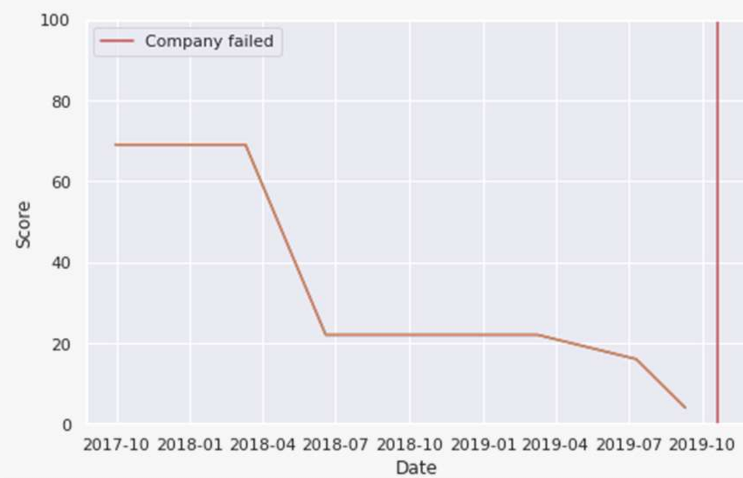
- The objective of the project is the development of a mix of regression and classification AI models to assess the credit reliability of the 8+ million Italian companies

Data

- The models leverage **financial and registry data**, together with **glocal economic indicators**, to predict ahead of time insolvencies, and give a measure of the risk

Why?

- Societal conditions rapidly change. Machine Learning can **extract the complex, latent patterns hidden in the data** itself, automate much of the analysis process, and **remove assumptions** from the scoring process



DOMAINS

- Business Intelligence

COMPETENCES

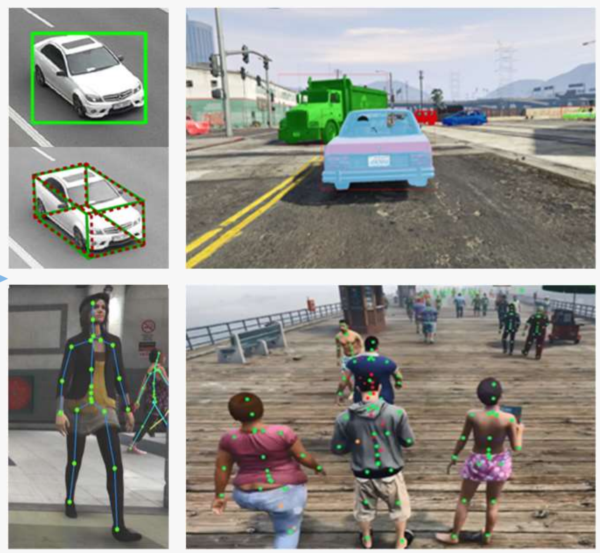
Regression and Classification, Decision Trees, Deep Learning, Finance and Business Data Analysis

Synthetic Data Generation for AI

Sometimes there are not enough real data to train complex models and there is the need of workaround…

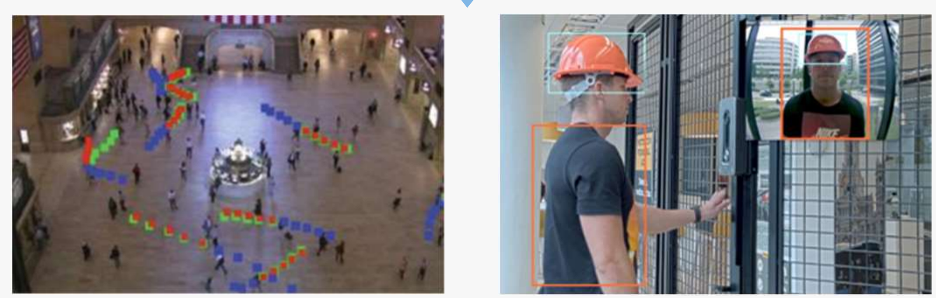
Using photorealistic synthetic information for training deep-learning approaches to be applied in real-world contexts, such as autonomous driving and safety.

Synthetic Data Environment



AUTONOMOUS DRIVING: SPATIAL SCENE UNDERSTANDING

SECURITY: PEDESTRIAN PATH PREDICTION AND SAFETY EQUIPMENT CHECK



DOMAINS
- Synthetic Data
- Autonomous driving
- Safety

AI TOPICS

Machine learning and Deep learning custom methodologies, derived and extended from the literature (i.e. Long-Short-Term-Memory (LSTM), Mask R-CNN, Inception-V3)

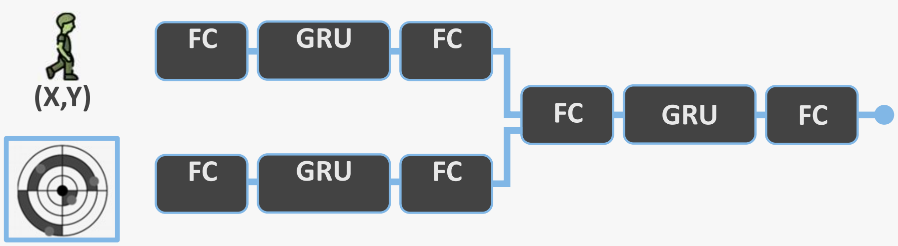
FRAMEWORK/PROJECT

Internal Research

Pedestrian Trajectory Estimation

Data augmented with Synthetic Data

TRAJECTORY PREDICTION ALGORITHM

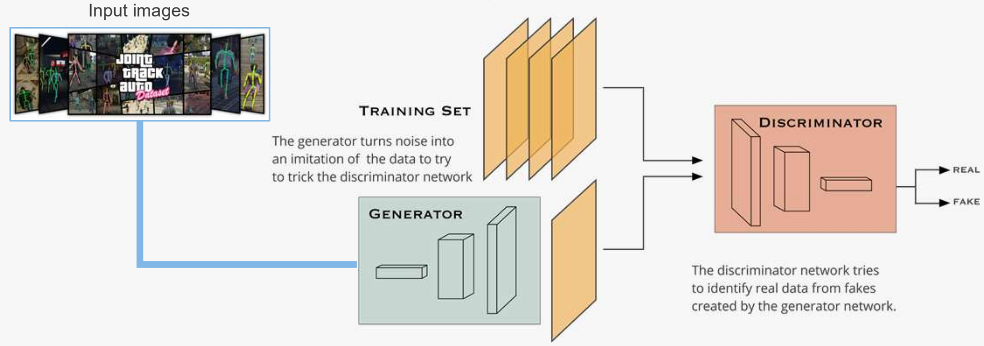


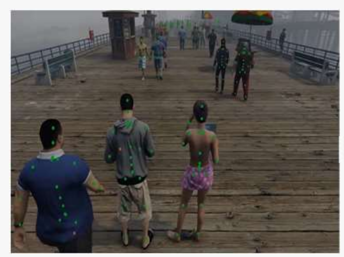
Deep learning - CNN (LSTM)

Deep Learning model to predict pedestrian trajectory. It can be used in autonomous driving algorithms or for urban security.

Input
Videos from fixed cameras public dataset and synthetic data. Synthetic Data from videogames is **domain adapted with generative networks** to look like real images.

GAN DATA AGUMENTATION SYNTHETIC TO REAL





JTA Synthetic Data video example

Deep Learning Generatative Adversarial Network (GAN) to transform synthetic data to real data.

Input
Synthetic scripted trajectories of pedestrian extracted from videogame engine.

Output
Synthetic2real trajectories transformed to mimic human behaviour before being used as training data

DOMAINS

- Environmental monitoring

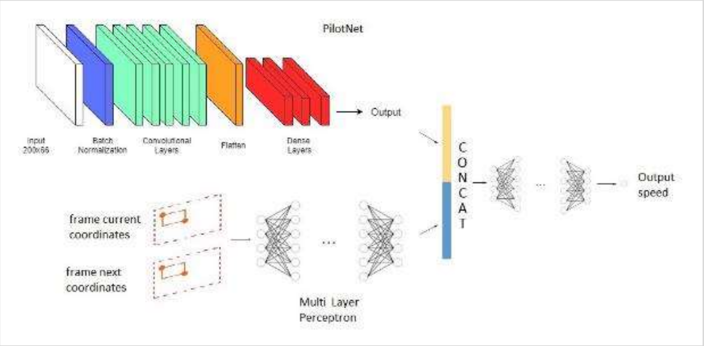
AI TOPICS
Supervised Machine Learning – Regression techniques, Neural Networks

FRAMEWORK/PROJECT
Internal research

DEEP LEARNING FOR ASSISTED DRIVING

Goal: distance and speed estimation

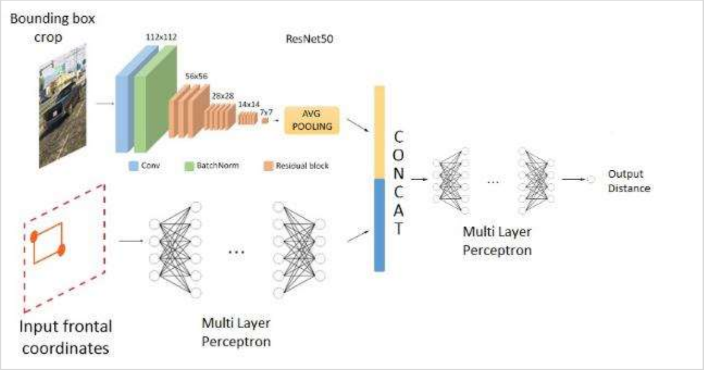
- Learning from synthetic traces from a game: GTA
- Deep Learning – evolution of 2 networks



Deep Neural Network used for **velocity estimation** adapted starting from **PilotNet** (NVIDIA)

In-game evaluation

Real evaluation



Deep Neural Network used for **distance estimation** adapted starting from **ResNet50** (Microsoft)

DOMAINS
- Industrial
- Automotive

AI TOPICS
Supervised Machine Learning – Deep Learning

FRAMEWORK/PROJECT
OptiTrucks (H2020-GV6-2015 IA - n.713788)

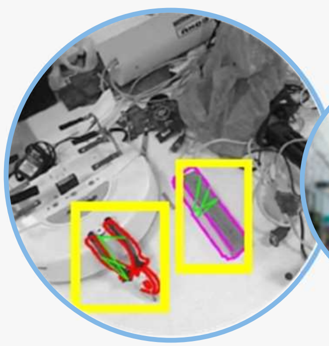
www.optitruck.eu

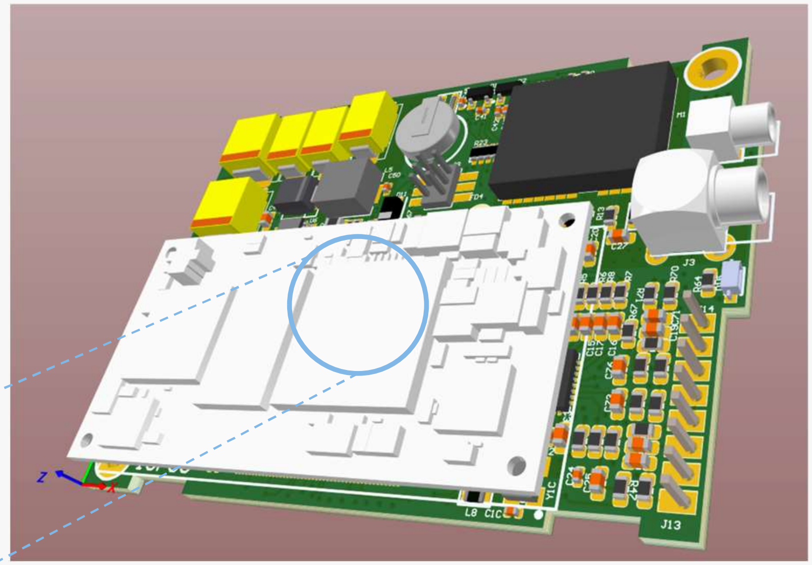
EMBEDDED AI

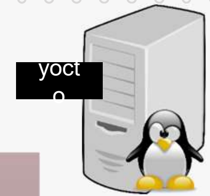
#yocto #audio #GNSS #multimedia #SoM
#edge-computing #embeddedAI #real-time
#face-recognition #object detection

A custom embedded board developed to generate data (e.g. audio and video streams, GNSS information).

The base project has been customized for several domains where we decentralized the intelligence **moving AI algorithms on the edge** by pruning state of the art models (e.g. video/audio-tagging, face-recognition, and object detection) or by designing new efficient models







YOCTO FRAMEWORK

Custom OS based on Linux Yocto

QT POWERED USER APP

Powerful C++ Software Core

- Embedded AI
- H.265 Video Capturing
- HQ Audio Recording
- GNSS Tracking
- 4G LTE Connectivity

DOMAIN
Multimedia, Embedded AI, Edge Computing

SIZE
2 Partners

AI TOPICS
Semantic segmentation with supervised machine learning
NLP – Understanding
Behavioral analysis

PARTNERS
S.A.E.T.

DURATION
12 months

LINKS ROLE
Platform Design
Coordinator
Development Leader

Lesson Learned

Leverage our strong local ecosystem to lead the innovation process taking local and international opportunities as a TEAM!

DATA-DRIVEN INNOVATION

Nowadays, those who own the data own a unique value that can also be unlocked thanks to the huge amount of open and publicly accessible data. **If we can say that AI is the "engine" of this innovation data is "fuel"**

EMBRACE THE FUTURE TRENDS

Be **aligned with the SDGs*** that are entering the innovation agendas of the EU, local governments, and enterprises. They drive us to **opportunities for tomorrow's business**.
New markets are being created; new questions are being asked...
Be ready for the next step of evolution: the **convergence between social sciences and technological sciences**. It will move the innovation paradigm in a **Human-Centric AI for a new data-humanism**.

BUILD DATA ALLIANCES

The **partnership between domain and technology experts** is the path towards innovation. Build alliances to create value chains based on data, so-called **data alliances**, is crucial to multiplying the value of single datasets and generate intelligent systems

DATA AND HUMAN-CENTERED AI

People at the center of AI processes means to have a continuous source of feedback and new data that will enable a **continuous and incremental learning approach**

Fabrizio Dominici

fabrizio.dominici@linksfoundation.com

Via Pier Carlo Boggio 61 | 10138 Torino (Italia)

+39 011 22 76 150

info@linksfoundation.com

linksfoundation.com