VOLUME 11
ARTIFICIAL INTELLIGENCE, ROBOTICS & DATA SCIENCE

Topic Coordinators
Sara Degli Esposti & Carles Sierra

CSIC SCIENTIFIC CHALLENGES: TOWARDS 2030
Challenges coordinated by:
Jesús Marco de Lucas & M. Victoria Moreno-Arribas

VOLUME 11 SUMMARY

CHALLENGE 3

MACHINE LEARNING AND DATA SCIENCE

Coordinators

J. J. Ramasco Sukia (IFISC)

L. Lloret Iglesias (IFCA, CSIC)

Participant researchers and centers

D. de Hertog (Avantopy)

A. Pizarro (Avantopy)

M. D. del Castillo (CAR, CSIC-UPM)

J. Villagra (CAR, CSIC-UPM)

N. Campillo (CIB, CSIC)

H. Mueller (IAE, CSIC)

L. Menéndez de la Prida (IC, CSIC)

A. Navas-Olive (IC, CSIC)

H. Domínguez (ICE, CSIC)

J. Aurentz (ICMAT, CSIC-UAM-UC3M-UCM)

F. Borondo (ICMAT, CSIC-UAM-UC3M-UCM)

M. Fontelos (ICMAT, CSIC-UAM-UC3M-UCM)

V. Gallego (ICMAT, CSIC-UAM-UC3M-UCM)

A. Kosgodagan (ICMAT, CSIC-UAM-UC3M-UCM)

R. Naveiro (ICMAT, CSIC-UAM-UC3M-UCM)

D. Ríos Insua (ICMAT, CSIC-UAM-UC3M-UCM)

I. Villanueva (ICMAT, CSIC-UAM-UC3M-UCM)

A. Villaseñor (ICTJA, CSIC)

J. Fernández Recio (ICVV, CSIC)

M. Romero Durana (ICVV, CSIC)

D. Ramiro (IEGD, CSIC)

F. Aguilar Gómez (IFCA, CSIC – UC)

J. Baño (IFCA, CSIC – UC)

J. M. Diego (IFCA, CSIC – UC)

J. M. Gutiérrez (IFCA, CSIC – UC)

I. Heredia (IFCA, CSIC – UC)

D. Herranz (IFCA, CSIC – UC)

P. Martínez (IFCA, CSIC – UC)

D. Rodríguez (IFCA, CSIC – UC)

F. Albiol (IFIC, CSIC – UV)

L. Fiorini (IFIC, CSIC – UV)

J. E. García (IFIC, CSIC – UV)

A. Oyanguren (IFIC, CSIC – UV)

R. Ruiz (IFIC, CSIC – UV)

A. Argyris (IFISC, CSIC – UIB)

P. Colet Rafecas (IFISC, CSIC – UIB)

M. Cornelles Soriano (IFISC, CSIC – UIB)

I. Fischer (IFISC, CSIC – UIB)

D. Gomila (IFISC, CSIC – UIB)

V. Martínez Eguíluz (IFISC, CSIC – UIB)

S. Meloni (IFISC, CSIC – UIB)

C. Mirasso (IFISC, CSIC – UIB)

D. García-González (IG, CSIC)

J. L. Arcos (IIIA, CSIC)

J. Cerquides (IIIA, CSIC)

J. Hernández (IIIA, CSIC)

J. Antonio (IIIA, CSIC)

M. R. García (IIM, CSIC)

L. Campillos Llanos (ILLA, CSIC)

1. EXECUTIVE SUMMARY

There is no doubt that the progressive digitalization of the world has a ground-breaking impact on every sphere of people's lives. Since the beginning of the XXI century, digital technology has permeated every aspect of modern society, becoming an integral part of our everyday lives. This brings both thrilling opportunities and new challenges for the research communities in this ever-changing and somewhat revolutionary context, as it implies shifts in established paradigms and application of completely new study approaches. However, it is not only scientists who are facing the challenge on

Copia gratuita / Personal free copy http://libros.csic.es

sectoral applications (climate, agriculture, hydrology, health, energy, etc.). There are a number of statistical ad hoc approaches, but ML could provide ***an efficient general approach, making the best of all the available data for a particular problem*** (postprocessing model outputs). Additionally, climate models numerically solve the physical equations governing the climate system, but they also include parameterizations (empirical subcomponents) to balance and model subgrid processes. ML has been already applied to emulate those sub-components, allowing to speed up computations. Therefore, there is evidence of the potential of integrated models benefiting from both approaches.

- **Providing new information for Earth science.** In geoscience, data are acquired using a variety of methods (satellite sensors, field measurements, computer simulations) with varying spatial and temporal resolution. The most common approach to handle multi-resolution data is to upsample the datasets (e.g. using interpolation methods). New approaches are needed ***to identify patterns at different resolutions while using the uninterpolated datasets.***
- **Triggering and real time analysis of particle physics data.** The future particle physics experiments will require AI to cope with an ***unprecedented level of data processing in real time to select the events of interest among a massive data rate*** (several times larger than the bandwidth of internet giants like Facebook and Google) and ***reduce the event size by performing real time event reconstruction and store only a reduced set of key elements of the event.*** Additionally, the reconstruction of physics objects will need AI techniques different from the current ones based on pattern recognition, analytical filtering and iterative clustering, to achieve the level of performance required by the future experiments.
- **Estimation of distances in astrophysics and cosmology.** The accurate distance estimation based on photometric measurements is one of the biggest challenges. ML techniques have ***the potential to tackle this complex problem and play a central role in these surveys.*** Additionally, they can contribute as well in accelerating expensive numerical simulations like hydrodynamical simulations involved in galaxy formation and evolution and in the simulation of the dark matter structure of the Universe.

© CSIC © del autor o autores / Todos los derechos reservados

CSIC white paper on Artificial Intelligence, Robotics and Data Science sketches a preliminary roadmap for addressing current R&D challenges associated with automated and autonomous machines. More than 50 research challenges investigated all over Spain by more than 150 experts within CSIC are presented in eight chapters. Chapter One introduces key concepts and tackles the issue of the integration of knowledge (representation), reasoning and learning in the design of artificial entities. Chapter Two analyses challenges associated with the development of theories –and supporting technologies– for modelling the behaviour of autonomous agents. Specifically, it pays attention to the interplay between elements at micro level (individual autonomous agent interactions) with the macro world (the properties we seek in large and complex societies). While Chapter Three discusses the variety of data science applications currently used in all fields of science, paying particular attention to Machine Learning (ML) techniques, Chapter Four presents current development in various areas of robotics. Chapter Five explores the challenges associated with computational cognitive models. Chapter Six pays attention to the ethical, legal, economic and social challenges coming alongside the development of smart systems. Chapter Seven engages with the problem of the environmental sustainability of deploying intelligent systems at large scale. Finally, Chapter Eight deals with the complexity of ensuring the security, safety, resilience and privacy-protection of smart systems against cyber threats.

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
CSIC
EDITORIAL CSIC