ANNUAL REPORT
2024

ellis
UNIT
ALICANTE

Contents

Letter by the President of the Board

CHRISTOF BEAUPOIL

President of the Board of
ELLIS Alicante

Dear friends,

It is a pleasure to share with you the 2024 Annual Report of the Fundación de la Comunitat Valenciana Unidad ELLIS Alicante.

This year has been full of important achievements, growth in our team, and the strengthening of the foundations we need to expand both the size and impact of our foundation in the years to come.

In 2024, ELLIS Alicante has reinforced its role as a key part of ELLIS, combining excellent scientific research with a strong ethical, social, and human-centered approach. We have welcomed nearly a dozen visiting researchers and two ELLIS pre-doctoral researchers during their international stays, expanded our group of associated researchers, and developed new research areas, while building strategic partnerships nationally and internationally.

This year, we have also increased our public presence by engaging citizens in the conversation about AI, organizing educational activities for young people, and actively promoting responsible and safe AI aligned with democratic values. It's no surprise that we rank as the top Google search result for several key terms related to our work, including "Artificial Intelligence for Social Good."

We have continued to be active participants in the European Horizon 2020 Project ELIAS, focused on sustainable AI research; in the RESUMAIS Center of Excellence, dedicated to Responsible AI; and in PERTE de la Economía de la Lengua in partnership with the University of Alicante.

Our progress depends greatly on the trust and collaboration of many individuals and organizations. I want to sincerely thank everyone who supports and works with us. Your commitment to ELLIS Alicante is an investment in the future, a support for European values, and a valuable contribution to advancing AI research centered on people. Thanks to your support, we continue shaping the future of Artificial Intelligence with the hope of creating a positive social impact.

I invite you to explore this Annual Report as a reflection of the collective effort made in 2024, and as a glimpse of the path we are building—step by step—with determination and conviction toward AI that truly benefits society.

NURIA OLIVER

ELLIS Alicante Director

It is an honor to present the 2024 Annual Report of ELLIS Alicante—a year marked by the consolidation of our team, the growth of our research lines, and increased social impact.

We have made significant progress in areas such as the detection and mitigation of algorithmic bias, human bias and AI, beauty filters, large language model safety, privacy, algorithmic censorship, and the potential of chatbots to foster critical thinking in education. Our scientific output has grown in both quantity and impact, and we have strengthened ties with leading institutions across Europe as well as at national and local levels.

Beyond technical achievements, 2024 has also been a year of open dialogue: we have launched initiatives to bring AI closer to society, to inspire young minds, and to contribute to the public debate on the ethical development of technology. Our new location at the privileged Pier 5 in the Digital District has undoubtedly played a role in attracting visiting researchers and collaborators.

I would like to express my deepest gratitude to the wonderful team at ELLIS Alicante, whose effort, dedication, and talent have been essential to achieving the milestones documented in this report. I also want to acknowledge the commitment and vision of our Board of Trustees, who guide and support us with steady confidence. Finally, we are deeply grateful to our sponsors and partners, especially the Government of the Generalitat Valenciana, whose financial and strategic support enables us to carry out excellent research with real social impact. Without the combined contributions of all these actors, our mission would not be possible.

invite you to read this report, which offers a detailed overview of our activities—from the evolution of our research projects and outreach efforts to our collaborations and alliances, awards and recognitions, and the future challenges we face with enthusiasm and a sense of responsibility.

We continue to move forward with the conviction that a different kind of Artificial Intelligence is possible: one that is more human, more just, and truly serves the common good.

Milestones **2024**

19 scientific publications

8 awards and recognitions

1 tutorial

1 dataset

1 European Horizon 2020 Project ELIAS

1 Intel Research Center for Responsible Human-AI Systems (RESUMAIS))

1 National Project (PERTE de la Economia de la Lengua)

Milestones 2024

16
new collaboration agreements

6
new institutional/corporate donors

Milestones 2024

68
guest presentations

7
HCML reading group sessions

4
conferences and workshops organized

8

ELLIS PhD students

1

ELLIS postdoctoral researcher

9

visiting researchers

16

associated researchers

Number 1 ranking on Google for keywords including:

Beauty Filters Bias

IA para el Bien Social

Art Censorship Social Media

Milestones 2024

3

articles written for El País and Euractiv

+ 90

media appearances

Awards and recognitions

DR OLIVER AWARDED AT THE 8TH EDITION OF THE TRIFERMED SOCIAL IMPACT IN HEALTH AWARDS

The director of ELLIS Alicante was awarded at the 2024 Trifermed Awards for her work in artificial intelligence applied to health and its impact during the COVID-19 pandemic.

DR OLIVER HAS BEEN RECOGNIZED ON STANFORD'S LIST OF THE TOP 2% OF SCIENTISTS WORLDWIDE SINCE 2020.

This prestigious ranking highlights the most influential researchers worldwide.

DR DERNER AWARDED BY THE CZECH TECHNICAL UNIVERSITY IN PRAGUE

He received the Best Doctoral Thesis Award from the Czech Technical University in Prague for his work on efficient learning methods in robotics with limited data, recognized for its quality and international impact.

2024 ICMI TECHNICAL IMPACT AWARD FOR AN INNOVATIVE STUDY ON CRIME PREDICTION USING MOBILE DATA

The 2014 article titled "Once upon a crime: towards crime prediction from demographics and mobile data" was awarded the 10-Year Technical Impact Award at the 26th ACM International Conference on Multimodal Interaction (ICMI 2024).

DR DERNER RECEIVES THE 2023 WERNER VON SIEMENS AWARD

He was recognized with the 2023 Werner von Siemens Award in the category of "Best Thesis in the Industry 4.0 Concept Area."

DR OLIVER IS RECOGNIZED AS A FELLOW OF THE ASIA-PACIFIC ARTIFICIAL INTELLIGENCE ASSOCIATION (AAIA).

In 2024, Dr Nuria Oliver was recognized as a Fellow of the Asia-Pacific Artificial Intelligence Association (AAIA), an honor that highlights her international leadership in the field of artificial intelligence.

DR OLIVER RECEIVES THE MINERVA AWARD FOR DIGITAL TRANSFORMATION FROM INECO OF THE MINISTRY OF TRANSPORT AND SUSTAINABLE MOBILITY.

She was awarded the Digital Transformation Award granted by Minerva, the women's network of Ineco, a public company of the Spanish Ministry of Transport and Sustainable Mobility.

DR OLIVER IS RECOGNIZED AS A LEADER IN COMPUTER SCIENCE RESEARCH IN SPAIN BY RESEARCH.COM.

According to Research.com, Nuria Oliver is the top-ranked female computer science researcher in Spain by D-index.

ELLIS Alicante

The ELLIS Alicante unit is the only Spanish non-profit research foundation within the ELLIS network that is dedicated exclusively to excellence in ethical, responsible, and human-centred artificial intelligence, with the aim of contributing to social good.

Located in Muelle 5 of Digital District in Alicante

# We are the only ELLIS unit dedicated exclusively to this topic.

# An AI revolution by and for the people

Purpose

VISION

ELLIS, and by extension ELLIS Alicante, is committed to conducting excellent, open, and freely accessible research, with the aim of ensuring that Europe plays a leading role in shaping the way modern AI is transforming the world. The vision is to establish an open and collaborative system that enables fundamental research to translate into practical applications, generating positive economic and social impact.
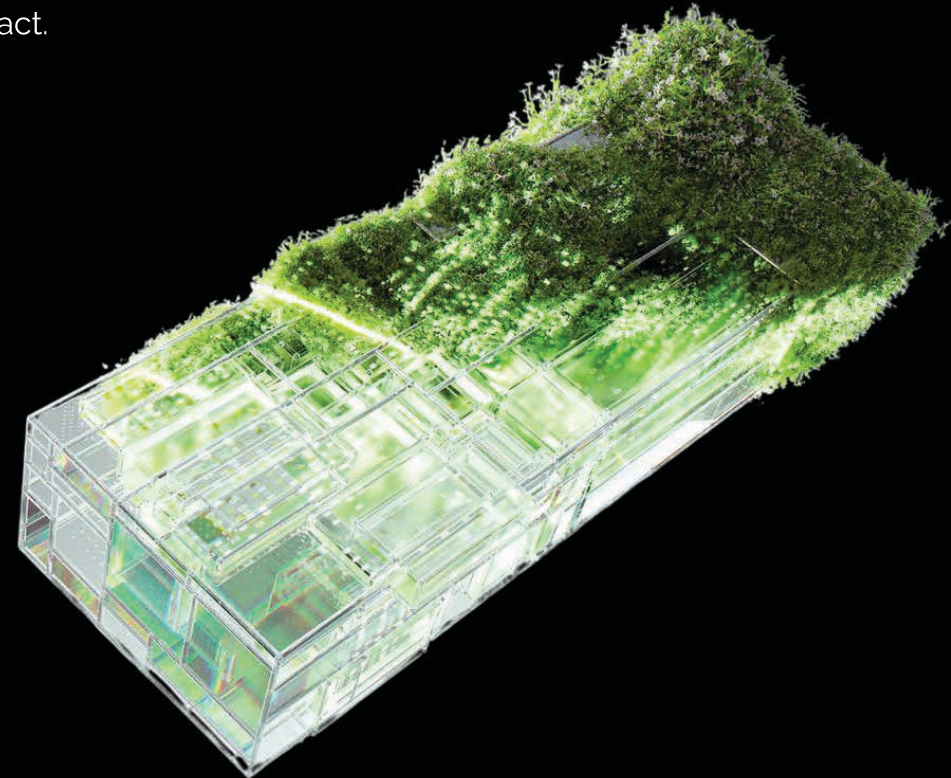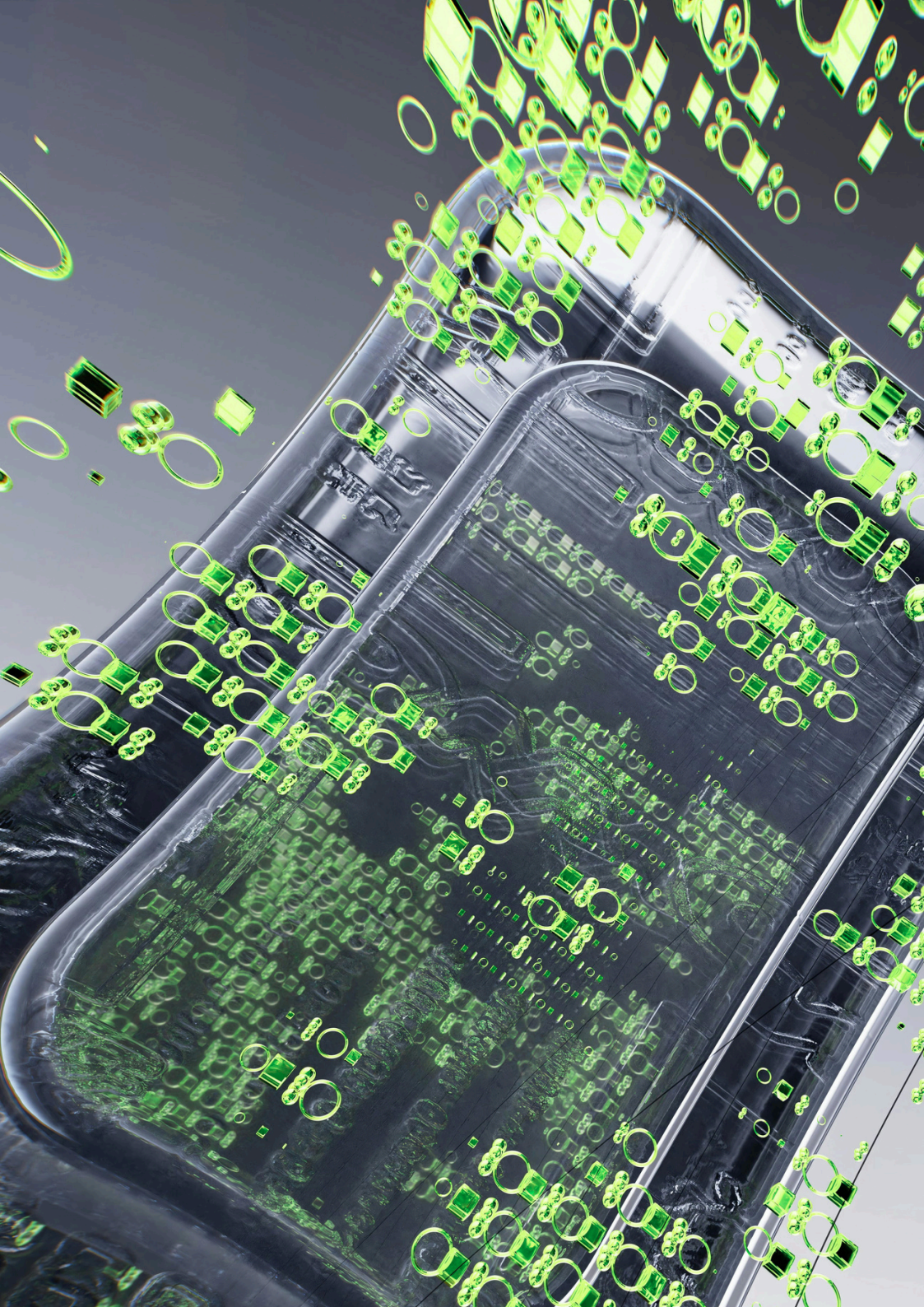
MISSION

ELLIS aims to become Europe's leading employer and source of outstanding research talent in modern Artificial Intelligence—namely, AI techniques based on learning from data.

In addition to supporting established researchers, ELLIS places a strong emphasis on attracting the brightest young minds from around the world to Europe through its ELLIS PhD programme.

The goal is to train the next generation of top researchers who will continue to advance our shared vision.

ELLIS Alicante fully supports the mission of ELLIS Europe, while adding a particular focus on human-centred AI and Artificial Intelligence for Social Good.

Our ambition at ELLIS Alicante is to attract excellent research talent to our area of influence: Alicante, the Valencian Community, Spain, and Europe as a whole.

Values

1 We strive for excellence and pursue our mission with passion.

2 We embrace calculated risks and new approaches. We have an entrepreneurial spirit that welcomes innovation, diversity of ideas and innovative thinking.

3 We are truthful, fair and trustworthy in all aspects of our work. We hold ourselves and our partners to the highest ethical standards. We expect others to do the same.

4 We understand that our ultimate success depends on that of our partners and funding agencies. Our humility will be the key to building legitimate and trusting relationships.

5 We value each of our team members as individuals, but we believe that we achieve the best results by working together and being team players.

6 We encourage the enjoyment of work as a key elemnt to foster creativity and motivation.

Goals

1º

Conduct excellent scientific research in ethical, responsible, and socially beneficial Artificial Intelligence.

2º

Promote collaborative research activity in the field of Artificial Intelligence with public and private entities at both international and national levels through the joint development of research projects, scientific outreach activities, and active participation in the various programs of the ELLIS European network of excellence.

3º

Attract, promote, and retain excellent research talent, following the criteria of excellence defined by ELLIS Europe, with a multidisciplinary approach that enables collaboration among Foundation researchers both internally and with external entities (national and international) to enhance the development of excellent research in Artificial Intelligence focused on people.

Human-centred

For Social Good

Artificial Intelligence

Responsible Ethics
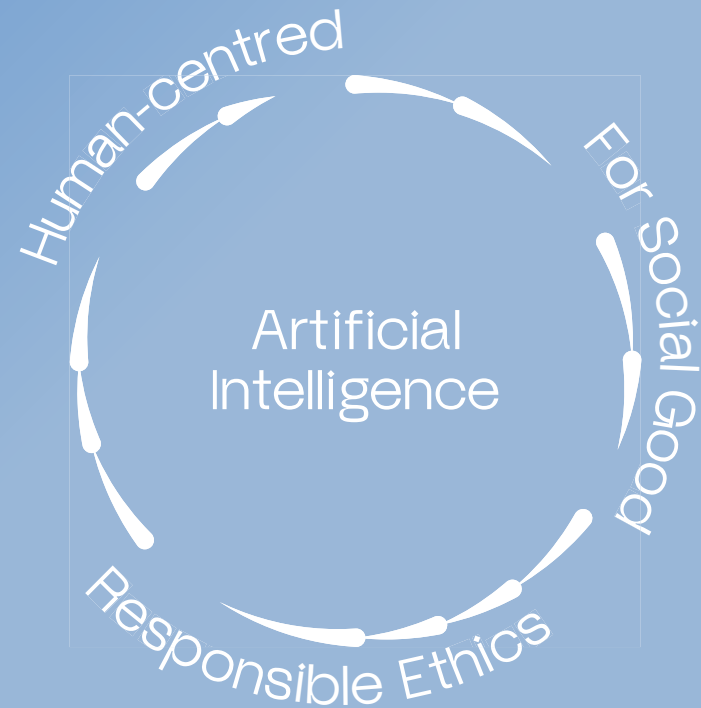
Beneficiaries of our activity

At ELLIS Alicante we firmly believe in the power of technology as a generator of progress and universal welfare.

The actions of the foundation must benefit society in its broadest sense.

Governance

Board of trustees

Responsible for the governance and representation of the Foundation, the fulfillment of its objectives and the administration and management of its assets.

President
Christof Beaupoil

Vice-president
Bernhard Schölkopf

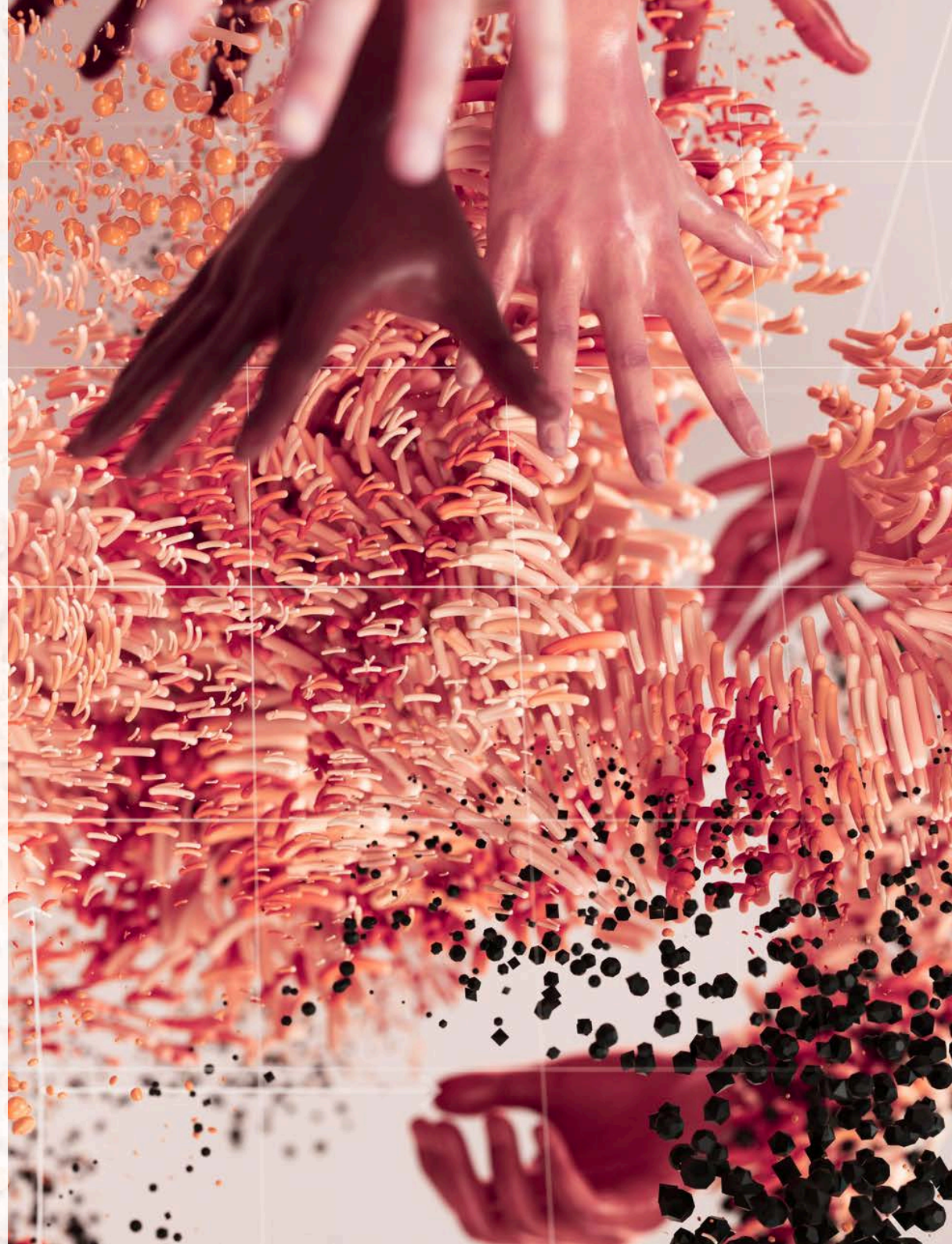Secretary (non-trustee)
Victor Riera Pastor

Trustee
Sara Gómez Martín

Trustee
Elías Fereres Castiel

Trustee
María Victoria Carrascosa Chaques

Scientific advisory board

Composed of scientists of recognised intenrational prestige in the research areas of ELLIS Alicante. Its function is to support the research agenda of ELLIS Alicante and to provide its expert opinion.

Max Planck Institute for Software Systems
Krishna Gummadi

The Alan Turing Institute / Univ Cambridge
Adrian Weller

European Commission Joint Research Center
Emilia Gómez

Universitat Pompeu Fabra
Carlos Castillo

Bruno Kessler Foundation
Bruno Lepri

Universidad de Cornell
Tanzeem Choudhury

Consejo Superior de Investigaciones Científicas
Ramón Lópex de Mántaras

Massachusetts Institute of Technology (MIT)
Sandy Pentland

ELLIS Europe

ELLIS (The European Laboratory for Learning and Intelligent Systems) is a pan-European network of excellence in AI that focuses on basic research in modern AI with positive social and economic impact.

Founded in 2018, ELLIS focuses on machine learning as the driver of modern AI and aims to ensure European sovereignty in this competitive field by creating a multicenter AI research laboratory.

One of the cornerstones of ELLIS is the ELLIS units.

As a result of several selection processes through competitive bidding, there are currently 41 ELLIS units in 16 European countries and Israel.

hey include some of the best research centers in modern Artificial Intelligence, not only in Europe but worldwide, such as the universities of Oxford and Cambridge, ETH Zurich, EPFL, several Max Planck institutes, or the Technion in Israel.

ELLIS Alicante is the only ELLIS unit that has been created from scratch as a startup unit

41
Units located in Europe & Israel

17
different countries

600
Millions of euros in budget over the past five years

+50M€
European funding through ELISE, ELSA, ELIAS, and ELLIOT

Team

Researchers

ELLIS Alicante brings together senior, junior, and pre-doctoral researchers with internationally recognised track records.

The Centre's Director is a Fellow of ELLIS, as well as of the ACM, IEEE, and EurAI—three of the most prestigious organisations in the field of computing and artificial intelligence.

Our pre-doctoral and postdoctoral researchers are part of the highly competitive ELLIS PhD and postdoctoral programmes.

At ELLIS Alicante, we conduct open research in collaboration with leading researchers from universities and other research institutions.

NURIA OLIVER

ELLIS Alicante Director

Scientific Director and Co-Founder of ELLIS Alicante
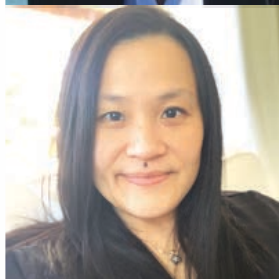
ERIK DERNER

Postdoctoral Researcher

Associated Researchers

BRUNO LEPRI

FBK, Italy
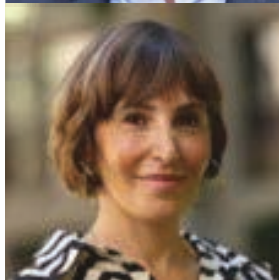
Member of the Scientific Advisory Committee

FLORA SALIM

UNSW Sydney, Australia
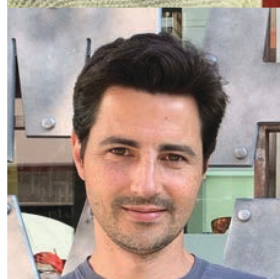
EMMANUEL LETOUZÉ

Data-Pop Alliance & UPF, Spain

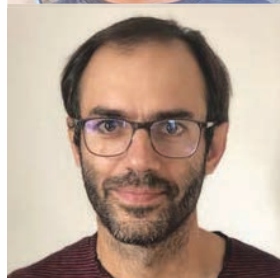**GEORGINA CURTO**

Notre Dame University, US

TANJA KÄSER

EPFL, Switerland

MIGUEL ÁNGEL LOZANO

University of Alicante, Spain

MANUEL GOMEZ RODRIGUEZ

University of Alicante, Spain

ELLIS Fellow

NOVI QUADRIANTO

University of Sussex, UK

ELLIS Scholar

SHIRLEY OGOLLA

HIIG, Germany

THOMAS HOFFMANN

ETH Zürich, Switzerland

ELLIS Fellow

DALIBOR KUČERA

University of South Bohemia, Czech Republic

KAISA VÄÄNÄNEN

Tampere University Finland
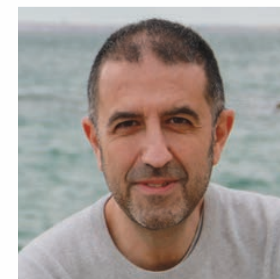
JUAN ANTONIO PEREZ

UdAlicante, Spain
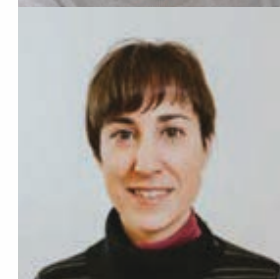
MARINA MARTINEZ-GARCIA

Jaume I University, Spain

THOMAS SERRE

Brown University, US

ELLIS Fellow
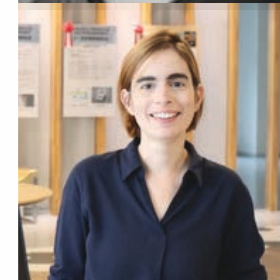
NOA GARCIA

Osaka University, Japan

Pre-Doctoral Researchers

ADITYA GULATI

GERGELY D. NÉMETH

LUCILE FAVERO

ADRIÁN ARNAIZ-RODRÍGUEZ

JULIEN COLIN

PIERA RICCIO

ELLIS ALICANTE UNIT AS A SECONDARY INSTITUTION

BENEDIKT HÖLTGEN

University of Tübingen, Germany

KAJETAN SCHWEIGHOFER

Johannes Kepler University, Austria

Management and Administration Team

Composed of professionals with recognised experience, expertise, and a solid track record in managing foundations and non-profit organisations, as well as in the administration of science and research.

Rebeca de Miguel
Head of Operations

Cristina González
Office Manager

Research areas

Research at ELLIS Alicante is driven by a commitment to advancing progress through the study of the intersection between artificial intelligence and people. With this focus, its scientific mission centres on addressing three key areas of fundamental research in modern AI.
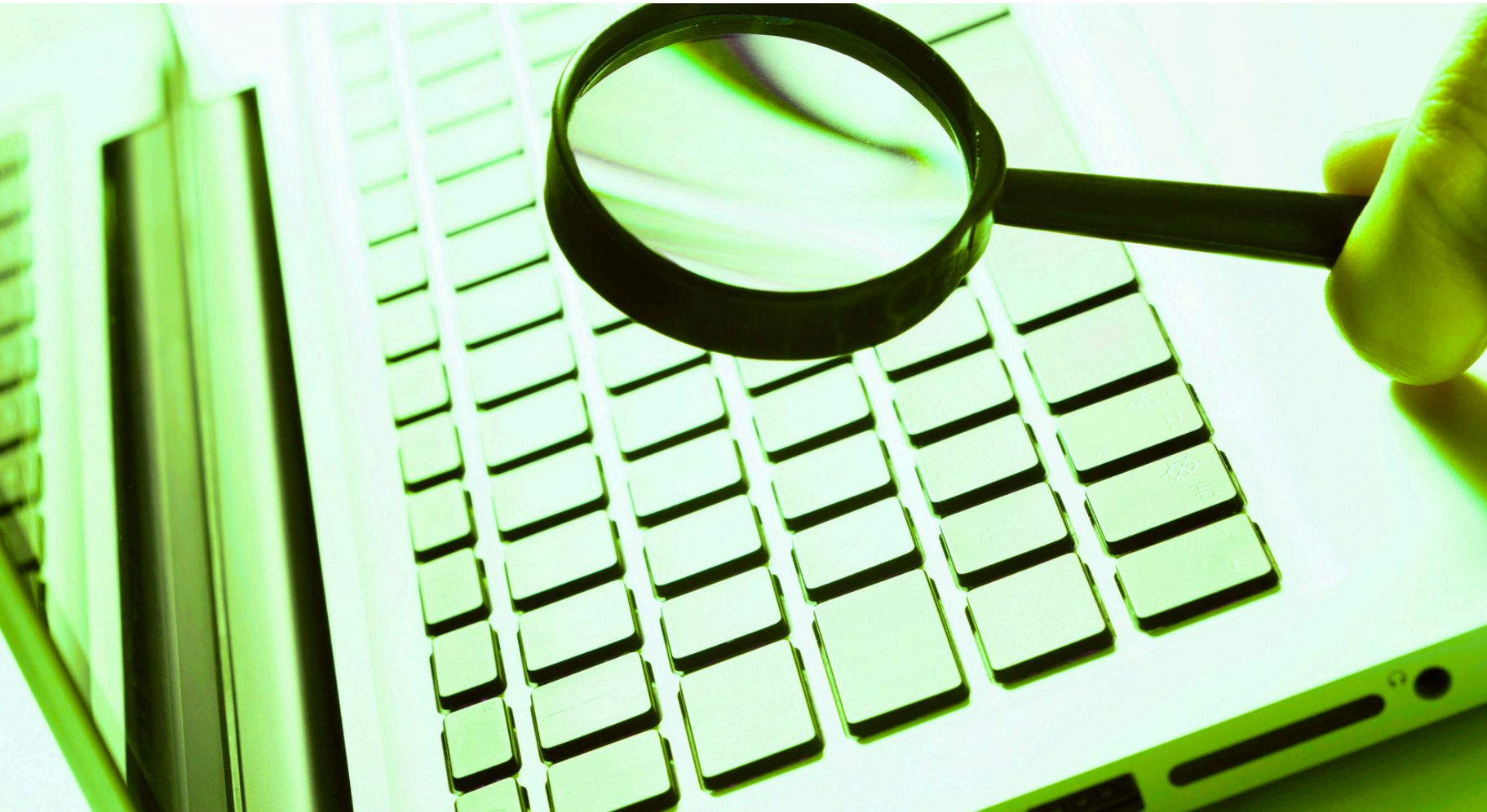
Program 1

ARTIFICIAL INTELLIGENCE TO UNDERSTAND US

The focus of this research area is modeling human behavior using AI techniques at both individual and aggregate levels. The work carried out in this area centers on the use and/or development of new Artificial Intelligence models that improve people's lives through better recognition and prediction of their behavior.

The practical applications are diverse, ranging from the development of algorithms that generate recommendations for users to accurate and fair credit models aimed at promoting financial inclusion. At the aggregate level, the goal is to model and predict human behavior on a large scale—such as in a country or region—which enables addressing challenges like pandemics, detecting potential economic crises, or responding to natural disasters.

Project - "BIASED"

COGNITIVE BIASES AND ARTIFICIAL INTELLIGENCE

This research project corresponds to the PhD thesis of ELLIS PhD student Aditya Gulati, a graduate with a Master's degree in Comuter Science from IIIT Bangalore in India with a strong interest in computational modelling of human behaviour. In addition to his work in India, he has collaborated with Prof Neumann at Ulm, Germany and Prof Gonzalez at CMU. His thesis supervisors are Dr Oliver (primary) and Dr Bruno Lepri (secondary) at FBK, who is an ELLIS Fellow in the ELLIS Human-Centric Machine Learning research program.

The research of Aditya's thesis focuses on studying cognitive biases and their relation with AI systems, modeling these biases using Artificial Intelligence techniques, and incorporating such biases into AI systems that interact with people. The goal is to design AI systems that help us learn about and address our limitations, as well as to design reliable intelligent interfaces.

During 2024, the main focus of the work has been research on the so-called beauty bias ("attractiveness halo effect") and its relation to AI-generated beautification filters. In the initial phase

of the project, we conducted a large user study on this cognitive bias, involving over 2,700 participants, to gather data on the presence of this bias when humans judged faces with and without beautification filters. Throughout 2024, we have carried out a pioneering, rigorous, and complex analysis of the data obtained from this study, which has provided valuable insights not only for the scientific community but also for society.

As part of this project, Aditya completed the second and final phase of his international stay in the ELLIS doctoral program at the Fondazione Bruno Kessler (FBK) in Trento, Italy, during the period from April to July 2024, in collaboration with Prof Bruno Lepri.

Scientific Publications

Gulati, A., Martínez-García, M., Fernández, D., Lozano, MA., Lepri, B., & **Oliver, N.** (2024). <u>What is Beautiful is Still Good: The Attractiveness Halo Effect in the era of Beauty Filters.</u> International Conference on Computational Social Science; International Conference on Thinking, Milán, Italia, 18 de junio de 2024 y en la revista *Royal Society Open Science, 11(11), November 27 2024.* **Top 5% of scientific publications according to Altmetric.**

Gulati, A., Lepri, B., & **Oliver, N.** (2024). <u>Lookism: The overlooked bias in computer vision. Fairness and ethics towards transparent AI: facing the chalLEnge through model Debiasing (FAILED).</u> *Workshop en el European Conference on Computer Vision (ECCV 2024). Milán, Italy, September 29, 2024.*

Additionally, Maria Hartikainen, a predoctoral researcher from the Human Technology research group at the University of Tampere (Finland), conducted a research stay from October 1 to November 30, 2024, investigating cognitive biases, AI biases, and their impact on human-AI interaction.

Furthermore, within the context of this project, we have begun studying biases in generative Artificial Intelligence systems, focusing particularly on a bias that has not yet been studied: appearance bias. Preliminary results of this research have been published in the following scientific article: Lookism: The overlooked bias in computer vision. Fairness and ethics towards transparent AI: facing the chalLEnge through model Debiasing (FAILED).

Program 2

ARTIFICIAL INTELLIGENCE THAT WE TRUST

This area focuses on the study of new intelligent interactive systems and personal assistants, including chatbots. Projects in this area encompass both the development of research prototypes exploring new intelligent interactive systems and the study of human–Artificial Intelligence interaction.

Proyect - "AIDEAS"

AI FOR EDUCATION BASED ON DIALOGUE AND SOCRATIC LEARNING

This research project corresponds to the doctoral thesis of ELLIS PhD student Lucile Favero, who joined the ELLIS Alicante predoctoral research team in November 2023. Lucile holds a Bachelor's degree in Mathematics from the University of Geneva and a Master's degree in Mathematics and Neuroscience. In her Master's thesis in Mathematics, supervised by Dr Sylvain Sardy, she developed a machine learning model to optimize heat pump systems. Concurrently, during her neuroscience studies, Lucile completed full-time internships for three years under the supervision of Dr Giulio Matteucci at the El-Boustani laboratory. There, she applied a decision-making model that promises to provide invaluable insights to experimenters when designing, analyzing, and refining behavioral tasks.

Currently, Lucile is pursuing her doctoral studies under the supervision of Dr Nuria Oliver (ELLIS Alicante), Dr Tanja Kaser (EPFL), and Dr Juan Antonio Pérez Ortiz (University of Alicante), focusing on integrating the Socratic method into chatbot development to improve educational outcomes and foster critical thinking skills.

Lucile's thesis focuses on studying the effectiveness of employing a Socratic approach in the development of educational chatbots, with a specific emphasis on promoting active learning, critical thinking, and knowledge retention. Based on existing literature on the Socratic method in education and the potential of chatbots for learning, this project aims to develop a pioneering Socratic educational chatbot called Maike. The ultimate goal is to create a learning environment that is intellectually more stimulating and emotionally more supportive.

The findings derived from this project aim to offer valuable insights and implications for the design of AI-based conversational agents for learning.

During 2024, the initial tasks necessary to achieve the objectives of this project have been carried out. Specifically, Lucile has conducted research on the capacity of open-source, small-sized language models to apply some aspects of the Socratic method through prompt-tuning and fine-tuning techniques.

Se ha desarrollado un primer prototipo de investigación de Maike que se ha evaluado y comparado con otros chatbots y cuyos resultados han sido presentados en dos artículos científicos.

Scientific Publications

Favero, L. A., Pérez-Ortiz, J. A., Käser, T., & **Oliver, N.** (2024). Towards Student-Centric AI-Supported Learning: Teaching Chatbots to Ask the Right Questions. *Collaborative AI and Modeling of Humans, AAAI Bridge Program, 2024. February 21, 2024*

Favero, L. A., Pérez-Ortiz, J. A., Käser, T., & **Oliver, N.** (2024). Enhancing Critical Thinking in Education by means of a Socratic Chatbot. *International Workshop on AI in Education and Educational Research, European Conference on Artificial Intelligence 2024. Santiago de Compostela, Spain, October 19, 2024*
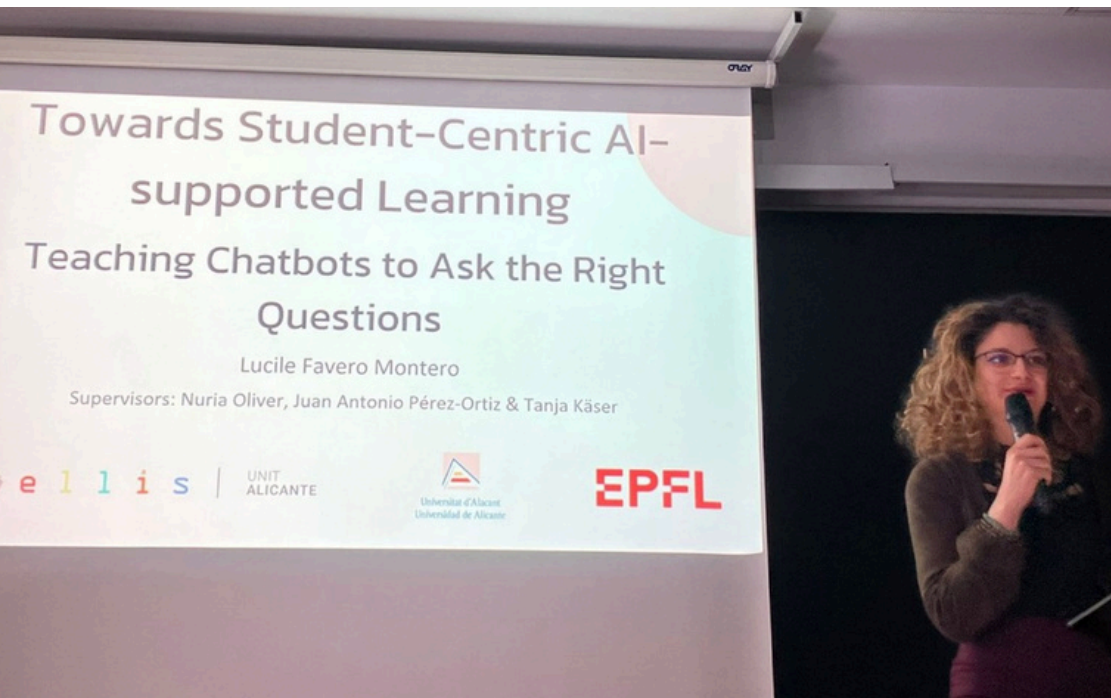
Program 3

ARTIFICIAL INTELLIGENCE THAT WE TRUST

The focus of this research area is to tackle the ethical challenges of today's AI, such as algorithmic discrimination, privacy violations, opacity, misinformation, and the subliminal manipulation of human behavior. Current AI algorithms are not invulnerable and suffer from limitations that must be identified and addressed to ensure their use is inclusive and human-centered. In this program, we also study the social impact of Artificial Intelligence, particularly the AI algorithms deployed on social networks and platforms.

Project - "FL"

FEDERATED LEARNING (FL) AND PRIVACY

This research project corresponds to the doctoral thesis of Gergely Németh, a PhD student in the ELLIS doctoral program. Gergely holds a master's degree in Computer Science from Hungary and professional experience at a Hungarian AI startup. He has extensive teaching experience and a strong interest in sharing his knowledge with students.

His primary thesis supervisor is Dr Oliver, and his secondary supervisor is Prof Novi Quadrianto from the University of Sussex (United Kingdom), who is an ELLIS Scholar in the ELLIS Human-Centric Machine Learning research program. The local PhD advisor at the University of Alicante is Professor Miguel Ángel Lozano.

Gergely's research focuses on federated learning. Federated learning (FL) is an approach within machine learning that has recently emerged to address privacy issues related to machine learning applications. Fairness in Artificial Intelligence is also a field that has gained awareness in recent years.

At the intersection of these two fields, new solutions are emerging for problems in centralized machine learning, but new challenges also arise that are unique to federated learning training architectures.

In 2024, this project has conducted research on the impact of federated learning on client data privacy, particularly the effect of different techniques for integrating client models into the server model, especially when client models are smaller than the server model (heterogeneous federated learning). As part of this project, a taxonomy of heterogeneous federated learning systems has been proposed, where different clients learn models of varying complexity to adapt to their computing

capabilities. Additionally, research has begun on the impact of data heterogeneity in federated learning. This work has resulted in the following scientific publication currently under review at the journal IEEE Access.

Finally, Gergely completed a research visit from May 20 to 22 at the University of Sussex with his international thesis supervisor, Prof Novi Quadrianto.

Scientific Publications

Németh, G. D., Lozano, M. A., Quadrianto, N., & **Oliver, N.** (2024). Addressing Membership Inference Attack in Federated Learning with Model Compression. Sent a IEEE Access.
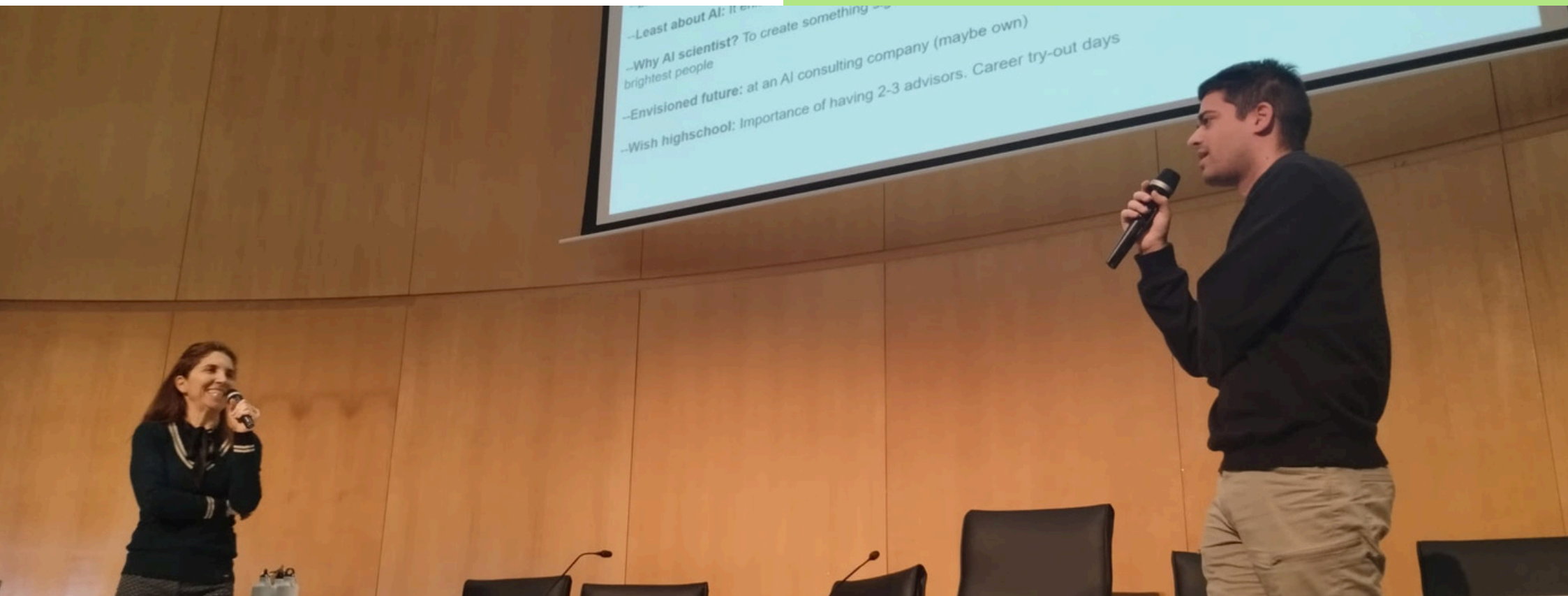
Project - "SOCIALAI"

SOCIAL IMPLICATIONS OF ARTIFICIAL INTELLIGENCE ON SOCIAL PLATFORMS

This research project forms part of the doctoral thesis of Piera Riccio, a PhD student within the ELLIS Doctoral Programme. Piera has been affiliated with Harvard's Metalab and holds a dual Master's degree in ICT for Smart Societies(Politecnico di Torino, Italy) and Data Science and Engineering (Telecom Paris – EURECOM, France). Her main research interest lies in exploring the limitations and challenges of Artificial Intelligence systems used in artistic decision-making processes.

Her primary supervisor is Dr Nuria Oliver, with Professor Thomas Hofmann from ETH Zurich (Switzerland) – an ELLIS Fellow in the ELLIS Human-Centric Machine Learning research programme – serving as co-supervisor. Her local doctoral advisor at the University of Alicante is Professor Miguel Ángel Lozano.

The core aim of the project is to examine the social implications of AI algorithms embedded in social media platforms used by billions of people worldwide. These platforms shape the flow of information in our society and are redefining paradigms of mass communication. Due to their ubiquitous nature, they continually influence our daily choices and have a profound cultural impact. Within these platforms, AI algorithms play a pivotal role — selecting and moderating content, offering automated editing tools, and promoting (or censoring) what is shown.

In 2024, the project focused on the issue of algorithmic censorship of art. Over the past decade, the use of AI techniques to generate creative content or support human creativity has attracted growing academic interest. Generative models proposed in the literature have transformed agency

and dynamics within artistic practices. However, a lesser-explored area at the intersection of AI and creativity concerns the indirect impact of AI on creativity via content moderation algorithms on social media. These algorithms often censor artistic works depicting nudity, thus acting as barriers to human creative expression.

Using both qualitative and quantitative methods — and in collaboration with the US-based initiative Do Not Delete Art — we have examined the impact of algorithmic censorship on artists, while also evaluating current AI algorithms' ability to distinguish human art from other content.

The project also investigated potential discrepancies between safety guidelines published by generative AI platforms for text-to-image generation and the actual moderation algorithms applied by these same platforms. Finally, the study concluded its work on biases in beautification filters, which have become widespread on social media platforms.

As part of the project, we collaborated with visiting PhD student Mirian Doh, who holds a joint affiliation with the ISIA Lab at the University of Mons (UMONS) and the IRIDIA Lab at the Université Libre de Bruxelles (ULB). During her three-month research stay at ELLIS Alicante (November 2024 – February 2025), she

explored bias in generative AI systems for human face generation, as well as approaches to mitigate racial bias in AI systems — both areas of high social relevance.
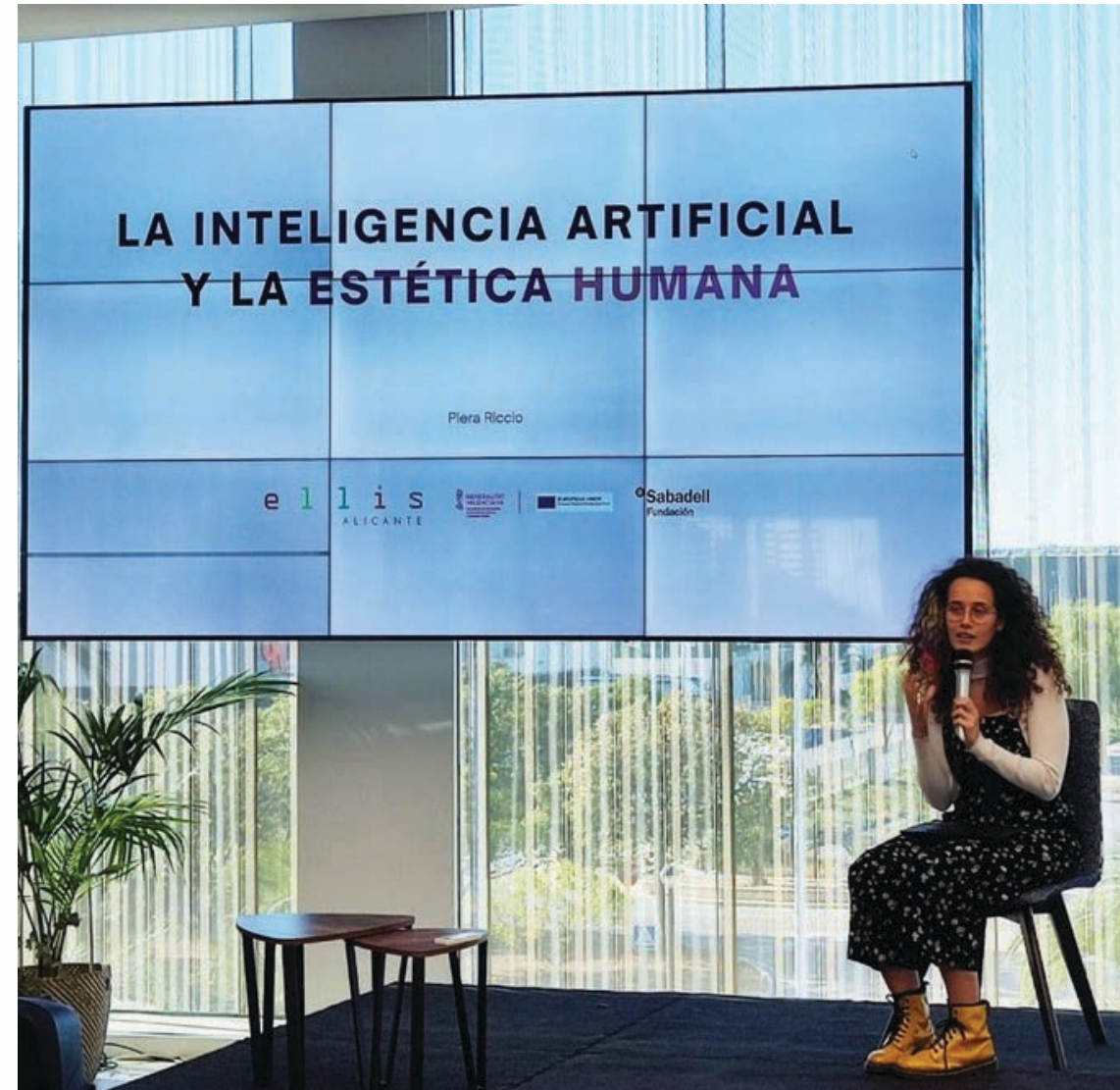
Scientific Publications

Riccio, P., Colin, J., Ogolla, S., & **Oliver, N.** (2024). Mirror, Mirror on the wall, who is the whitest of all? Racial biases in social media beauty filters. *Social Media and Society, 10(2)*, 20563051241239295. April 19 2024

Riccio, P., Hofmann, T., & **Oliver, N.** (2024). Exposed or Erased: Algorithmic Censorship of Nudity in Art. Proceedings of the ACM Conference on Human Factors in Computing Systems *(CHI)*, 1-17. 2024. May 11 2024.

Riccio, P., Curto, G., Hofmann, T., & **Oliver, N.** (2024). An Art-centric perspective on AI-based content moderation of nudity. *Artificial Intelligence for Visual Arts (AI4VA) workshop at the European Conference on Computer Vision (ECCV) 2024.* Milán, Italia, September 29 2024

Riccio, P., Curto, G., & **Oliver, N.** (2024). Exploring the Boundaries of Content Moderation in Text-to-Image Generation. *Critical evaluation of generative models and their impact on society (CEGIS) workshop at the European Conference on Computer Vision (ECCV) 2024.* Milán, Italia, September 29 2024

Project - "FAIRNESS"

FAIR AND TRANSPARENT MACHINE LEARNING MODELS

This research project forms part of the doctoral thesis of Adrián Arnáiz, a PhD student within the ELLIS Doctoral Programme. Adrián holds a Master's degree in Data Science from the Open University of Catalonia (UOC) and has conducted research at the University of Burgos in collaboration with Professor Kanber of University College London (UK). His previous work includes the development of new autoencoder architectures for reconstructing brain MRI scans

His primary supervisor is Dr Nuria Oliver, with Professor Manuel Gómez Rodríguez of the Max Planck Institute for Software Systems in Saarbrücken (Germany) – an ELLIS Fellow in the Human-Centric Machine Learning research programme – serving as co-supervisor. His local advisor at the University of Alicante is Professor Miguel Ángel Lozano.

Adrián's thesis focuses on algorithmic fairness and transparency. Machine learning models are increasingly being deployed to tackle complex societal problems and are also being used to support or even make decisions that affect individuals' lives in key domains such as justice and healthcare. As such, the ethical implications of these systems must be carefully considered – including issues of privacy, transparency, accountability, trustworthiness, autonomy, and fairness. Many of these concepts, however, lack universally agreed definitions.

In 2024, Adrián's research focused on fairness-enhancing methods through data pre-processing. In particular, he explored the possibility of improving algorithmic fairness by automatically assigning weights to individual data points. Additionally, he investigated the use of spectral graph theory to measure both social capital within a social network and the isolation of distinct social groups defined by protected attributes of societal importance, such as gender, religion, or age. Once a methodology for measuring social isolation was developed, research was conducted to design a novel algorithm to mitigate discrimination caused by higher isolation of specific groups – for example, women.

Two other PhD students from the ELLIS Programme, for whom ELLIS Alicante is a secondary institution, contributed research within the context of this project. Kajetan Schweighofer from JKU Linz (Austria) carried out a secondary institution research stay at ELLIS Alicante from 1 October 2023 to 31 March 2024, focusing on the implications of Deep Ensembles for algorithmic fairness. Ben Höltgen from the University of Tübingen (Germany) began his stay on 1 October 2024, which will run until 31 March 2025, and is investigating fairness in relation to the use of protected attributes.

Between March and August 2024, Adrián completed a research placement at his secondary institution, the prestigious Max Planck Institute for Software Systems, under the supervision of Professor Manuel Gómez Rodríguez.

In the summer of 2024, Adrián was selected to attend the renowned LOGML (London Geometry and Machine Learning) Summer School, which brings together mathematicians and computer scientists to work on a wide range of problems at the intersection of geometry and machine learning.

As part of this project, Adrián also co-led a highly successful tutorial on graph learning alongside researcher Ameya Velingker at the International Conference on Machine Learning (ICML 2024) – the world's largest AI conference, with over 7,000 attendees. The tutorial was followed by more than 1,000 participants.
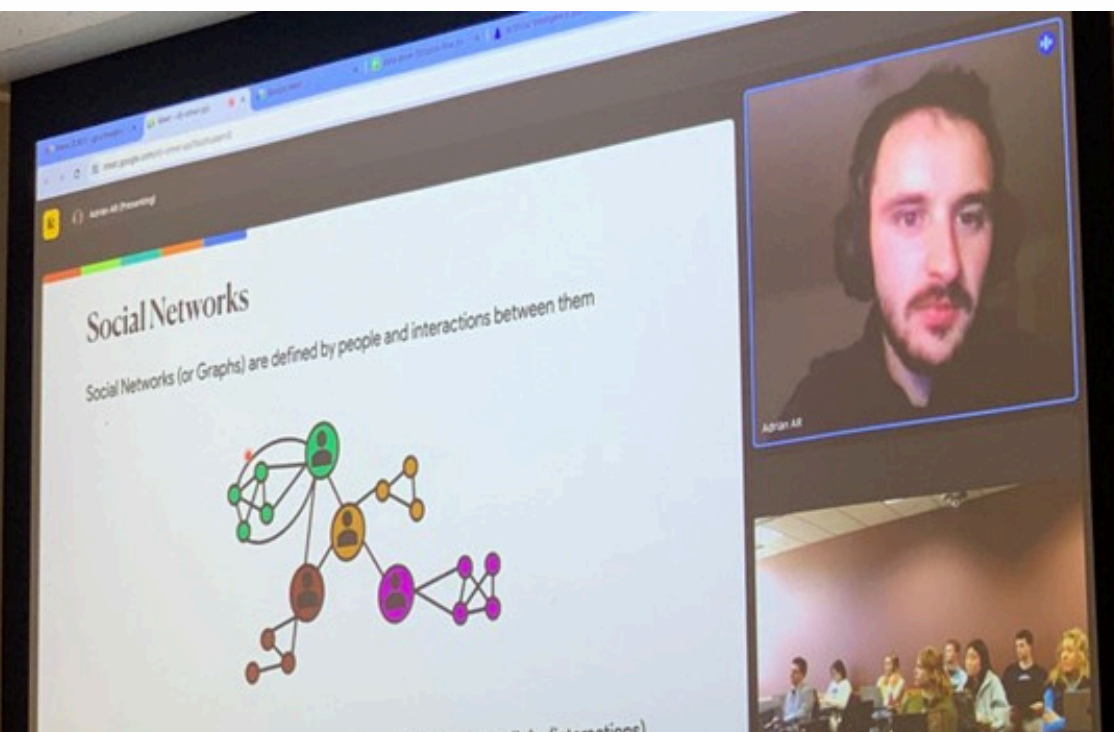
Scientific Publications

Arnaiz-Rodríguez, A., Curto, G., & **Oliver, N.** (2024). Structural Group Unfairness: Measurement and Mitigation by means of the Effective Resistance. *Workshop on Trustworthy Learning on Graphs (TrustLOG). International Conference on WWW 2024.* Singapore. June 3 2024.

Arnaiz-Rodríguez, A., & **Oliver, N.** (2024). Towards Algorithmic Fairness by means of Instance-level Data Re-weighting based on Shapley Values. *Workshop on Data-centric Machine Learning Research (DMLR). International Conference on Machine Learning (ICML). 2024. Vienna, Austria. Online.*

Schweighofer, K., **Arnaiz-Rodríguez, A.**, Hochreiter, S., & **Oliver, N.** (2024). The Disparate Benefits of Deep Ensembles. *arXiv:2410.13831.*

Project - "XAI"

TRANSPARENCY, UNCERTAINTY AND ALGORITHMIC EXPLAINABILITY

This research project is part of the doctoral thesis of Julien Colin. Julien holds a Bachelor's degree in Physics and Chemistry (2019, University of Lorraine) and a Master's degree in Cognitive Science: Natural and Artificial Cognition (2021, INP Grenoble). Prior to starting his PhD, he worked as a research assistant – first at ANITI for six months (2021, Toulouse), and then at Brown University for five months (2022, Providence).

Julien joined ELLIS Alicante in November 2022. His doctoral research lies at the intersection of explainable AI and neuroscience, with a focus on developing methods that help humans better understand intelligent systems. His thesis supervisors are Dr Nuria Oliver (ELLIS Alicante) and Professor Thomas Serre (ANITI and Brown University). His local doctoral advisor at the University of Alicante is Professor Miguel Ángel Lozano.
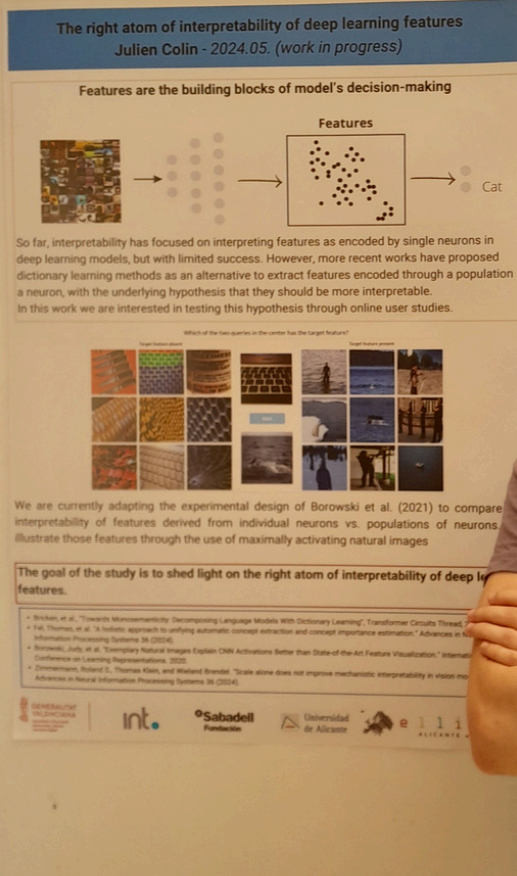
The project focuses on algorithmic transparency and explainability. Explainable AI aims to develop methods that allow humans to interpret and understand how models behave, with attribution methods being the most widely used. Recent work on human evaluation of attribution methods in computer vision has shown that (1) these methods are generally effective at indicating where the model is looking, but the research community must now turn its attention to developing complementary explainability techniques that also show what the model is seeing; and (2) current benchmarks used to evaluate explainability methods are poorly aligned with human understanding and therefore need improvement.

This project seeks inspiration from neuroscience to develop better benchmarks that can provide a more accurate assessment of the usefulness of explainability methods.

Scientific Publications

Colin, J., Goetschalckx, L., Fel, T., Boutin, V., Gopal, J., Serre, T, & **Oliver, N.** (2024). <u>**Local vs distributed representations: What is the right basis for interpretability?**</u> *arXiv:2411.03993*. November 6 2024

In 2024, the project explored feature visualisations as tools for explainability and conducted an online user study to examine whether representations based on the stimulation of individual neurons in a deep neural network are harder for humans to interpret than distributed representations. The aim was to shed light on the potential advantages of distributed representations for understanding the inner workings of complex deep neural networks. The results of this study provided empirical evidence supporting the superiority of distributed representations.

Project - "VLMBE"

BIASES AND ETHICS IN LANGUAGE MODELS IN VALENCIAN

Dr Erik Derner earned his PhD in Robotics and Artificial Intelligence from the Czech Technical University (CTU) in Prague, Czech Republic, in 2022. His doctoral thesis was awarded the 2023 Werner Von Siemens Prize in the Industry 4.0 category, as well as the Dean's Award from the Faculty of Electrical Engineering at CTU. His research focuses on the social and ethical implications of large language models (LLMs). He is a postdoctoral researcher within the ELLIS postdoctoral programme, with Professor Robert Babuška from CTU in Prague serving as his secondary postdoctoral advisor.

This project focuses on evaluating bias in the datasets used to train large language models, proposing mitigation methods, and studying the interaction between humans and LLMs from ethical and privacy perspectives. The aim of his work is to contribute to the development of safe, robust, and user-friendly language models, with particular attention to low-resource languages such as Valencian, which holds great cultural and social relevance in the Valencian Community.

Scientific Publications

Derner, E., Sansalvador de la Fuente, S., Gutiérrez, Y., Moreda, P., & **Oliver, N.** (2024). Leveraging Large Language Models to Measure Gender Bias in Gendered Languages. *arXiv:2406.13677*. June 19 2024.

Project - "HCLM"

HUMAN-CENTRED
LANGUAGE MODELS

Dr Erik Derner's second project focuses on the human and ethical aspects arising from the widespread use of large language models, such as ChatGPT.

In 2024, research was conducted on the security risks associated with large language models, which can potentially be manipulated into revealing harmful or privileged information. Additionally, pioneering work was carried out on the ability of a large language model like ChatGPT to automatically infer users' personalities based on their text input.

Dr Derner has completed two research stays at the CIIRC CTU, the Czech Institute of Informatics, Robotics and Cybernetics at the Czech Technical University in Prague, as part of his ELLIS postdoctoral training programme. The first stay took place between December 2023 and January 2024, while the second was carried out during March 2024.

Scientific Publications

Derner, E., Kučera, D., **Oliver, N.,** & Zahálka, J. (2024). Can ChatGPT Read Who You Are?. *Computers in Human Behavior: Artificial Humans, 2(2),* 100088. July 23 2024

Derner, E., Batistič, K., Zahálka, J., & Babuška, R. (2024). A Security Risk Taxonomy for Prompt-Based Interaction with Large Language Models. *IEEE Access, 12*, 126176-126187. August 26 2024.

Publications

Favero, L. A., Pérez-Ortiz, J. A., Käser, T., & **Oliver, N.** (2024). Towards Student-Centric AI-Supported Learning: Teaching Chatbots to Ask the Right Questions. *Collaborative AI and Modeling of Humans, AAAI Bridge Program,* 2024.

Arnaiz-Rodríguez, A., & **Oliver, N.** (2024). Towards Algorithmic Fairness by means of Instance-level Data Re-weighting based on Shapley Values. *Workshop on Data-centric Machine Learning Research (DMLR).* International Conference on Learning Representations (ICLR) 2024.

Arnaiz-Rodríguez, A., Curto, G., & **Oliver, N.** (2024). Structural Group Unfairness: Measurement and Mitigation by means of the Effective Resistance. *WWW 2024 Workshop on Trustworthy Learning on Graphs (TrustLOG).*

Riccio, P., **Colin, J**., Ogolla, S., & **Oliver, N.** (2024). Mirror, Mirror on the wall, who is the whitest of all? Racial biases in social media beauty filters. *Social Media and Society*, *10(2),* 20563051241239295, 2024.

Bolt, K., Gil-González, D., & **Oliver, N**. (2024). Unconventional data, unprecedented insights: leveraging non-traditional data during a pandemic. Frontiers of Public health, 12, 1350743, 2024

Riccio, P., & **Oliver, N.** (2024). A Techno-Feminist Perspective on the Algorithmic Censorship of Artistic Nudity. Bibliotheca Hertziana–Max Planck Institute for Art History, 3, 2024

Riccio, P., Hofmann, T., & **Oliver, N.** (2024). Exposed or Erased: Algorithmic Censorship of Nudity in Art. *Proceedings of the ACM Conference on Human Factors in Computing Systems, 1-17. CHI 2024.*

Derner, E., Sansalvador de la Fuente, S., Gutiérrez, Y., Moreda, P., & **Oliver, N.** (2024). Leveraging Large Language Models to Measure Gender Bias in Gendered Languages. arXiv:2406.13677., 2024

Arnaiz-Rodríguez, A., & Losada-Carreño, J. (2024). Intersection of reliable AI with Labor Law. A legal and technical study from a tripartite taxonomy. Revista General de Derecho del Trabajo y de la Seguridad Social, 2024.

Gulati, A., Martinez-Garcia, M., Fernandez, D., Lozano, MA., Lepri, B., & **Oliver, N.** (2024). What is Beautiful is Still Good: The Attractiveness Halo Effect in the era of Beauty Filters. *Int. Conf. on Computational Social Science; Int. Conf. on Thinking de Milan y Royal Society Open Science, 11(11), 2024.* ***Top 5% de las publicaciones científicas según Altmetric.***

Derner, E., Kučera, D., **Oliver, N.,** & Zahálka, J. (2024). Can ChatGPT Read Who You Are?. *Computers in Human Behavior: Artificial Humans, 2(2), 100088, 2024.*

Derner, E., Batistič, K., Zahálka, J., & Babuška, R. (2024). A Security Risk Taxonomy for Prompt-Based Interaction with Large Language Models. I*EEE Access, 12,* 126176-126187, 2024.

Gulati, A., Lepri, B., & **Oliver, N.** (2024). Lookism: The overlooked bias in computer vision. *Fairness and ethics towards transparent AI: facing the chalLEnge through model Debiasing (FAILED) - Workshop at the European Conference on Computer Vision (ECCV) 2024.*

Riccio, P., Curto, G., Hofmann, T., & **Oliver, N.** (2024). An Art-centric perspective on AI-based content moderation of nudity. *Artificial Intelligence for Visual Arts (AI4VA) workshop at the European Conference on Computer Vision (ECCV) 2024.*

Riccio, P., Curto, G., & **Oliver, N.** (2024). Exploring the Boundaries of Content Moderation in Text-to-Image Generation. *Critical evaluation of generative models and their impact on society (CEGIS) workshop at the European Conference on Computer Vision (ECCV) 2024.*

Schweighofer, K., **Arnaiz-Rodríguez, A.**, Hochreiter, S., & **Oliver, N.** (2024). The Disparate Benefits of Deep Ensembles. *arXiv:2410.13831, 2024.*

Favero, L. A., Pérez-Ortiz, J. A., Käser, T., & **Oliver, N.** (2024). Enhancing Critical Thinking in Education by means of a Socratic Chatbot. *International Workshop on AI in Education and Educational Research, European Conference on Artificial Intelligence,* ECAI 2024 workshop, 2024.

Colin , J., Goetschalckx, L., Fel, T., Boutin, V., Gopal, J., Serre, T, & **Oliver, N.** (2024). Local vs distributed representations: What is the right basis for interpretability?. *arXiv:2411.03993.2c, 2024*

Arnaiz-Rodríguez, A., *& Losada-Carreño, J. (2024). Studying Causality in Algorithmic Decision Making: the Impact of IA in the Business Environment. Revista Internacional y Comparada de RELACIONES LABORALES Y DERECHO DEL EMPLEO, 2024*
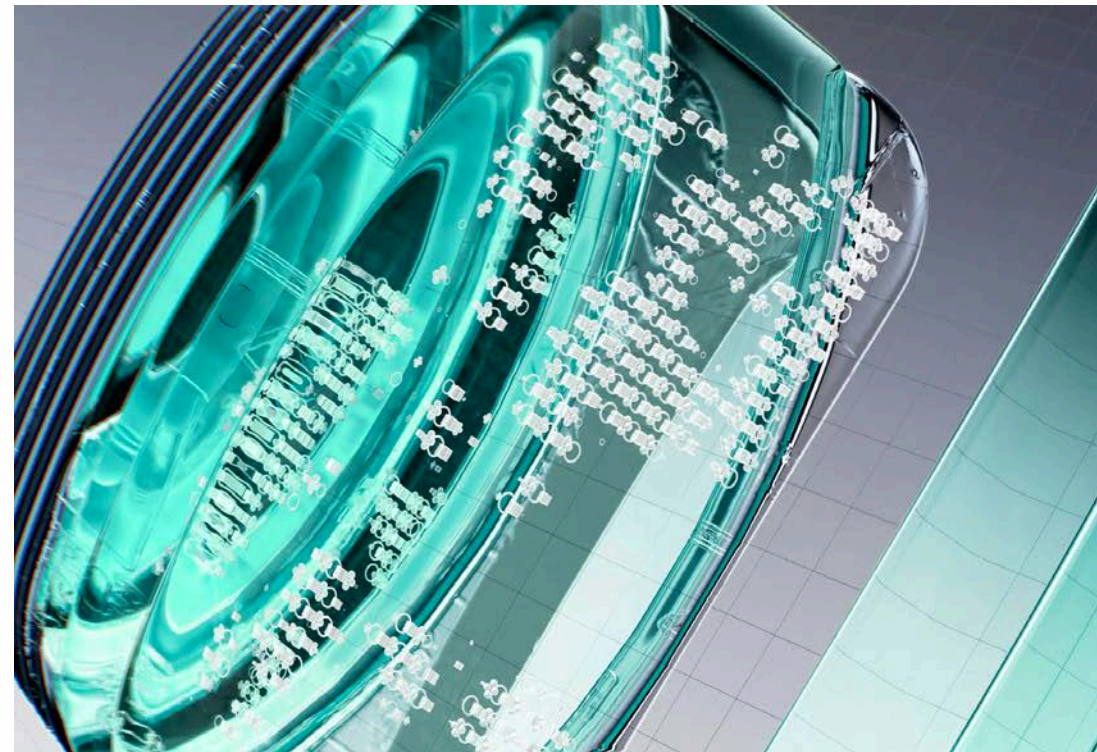
Communication Activities and Scientific-Technical Outreach Actions

ELLIS / ELIAS HCML Workshop 2024

Helsinki, Finland
27th of June 2024

ELLIS ALICANTE co-organises the ELLIS programme workshop **"Human-Centric Machine Learning (HCML)"** in collaboration with the ELISE and ELIAS projects.

The event also featured talks from prominent experts such as Dr Bernhard Schölkopf, Chair of ELLIS Europe, and Professor Neil Lawrence. Additionally, there was a panel discussion on the relationship between humans and AI, followed by a plenary session and a dinner cruise to close the day.

The ELLIS / ELIAS HCML Workshop 2024 was held as part of the ELISE Closing Conference and ELLIS Community Event.

This workshop was part of the ELIAS H2020 project, co-organised by ELISE, and was also broadcast virtually.

The workshop, within the ELLIS research programme and co-led by Dr Oliver, focused on presentations by distinguished researchers working to make AI systems transparent, fair, interpretable, and verifiable.

"Destinguished Speaker" Series

As part of the ELLIS Distinguished Speaker Series, ELLIS Alicante invites leading artificial intelligence researchers from around the world to present their work to the ELLIS Alicante team. In collaboration with the University of Alicante, the event is open to all students and researchers at the university.

AI FOR
SOCIAL IMPACT

Dr Miguel Luengo, Chief Data Scientist, Spotlab (Madrid, Spain)

On January 31, 2024, ELLIS Alicante invited Dr Miguel Luengo to deliver a talk as part of the ELLIS Alicante Distinguished Speaker Series.

Dr Luengo shared with attendees his projects and lessons learned throughout his career in academia, international organisations, and as an entrepreneur. His work has involved using data science tools, complex systems, and AI to tackle challenges across diverse fields, including basic biological research, personalised medicine, infodemics, human rights, and humanitarian response.

He also provided a detailed overview of Spotlab.ai, an organisation that develops and implements AI systems for diagnostics and clinical research, with applications ranging from neglected tropical diseases.

AI-DRIVEN PERSONALIZATION TO SUPPORT HUMAN-AI COLLABORATION

Professor Cristina Conati, Computer Science, University of British Columbia (Vancouver, Canada)

On July 11, 2024, ELLIS Alicante invited Dr Cristina Conati to give a talk as part of the ELLIS Alicante Distinguished Speaker Series.

At the Human-AI Interaction group at the University of British Columbia, she researches how to support human-AI collaboration through AI artefacts that can understand relevant user properties (such as states, skills, and needs) and personalise interactions accordingly, while preserving transparency, user control, and trust.

In this talk, Prof Conati illustrated examples of her research in AI-driven personalisation, covering areas such as Adaptive Visualisations for Users, Intelligent Tutoring Systems, and Personalised Explainable AI.

EXPLORING NLP ADVANCEMENTS: ADDRESSING BIASES AND HUMAN VALUES IN LARGE LANGUAGE MODELS

Associate Professor Dr Ayoub Bagheri, NLP and Data Science, Utrecht University (Utrecht, The Netherlands)

On July 11, 2024, ELLIS Alicante invited Professor Ayoub Bagheri to deliver a talk as part of the ELLIS Alicante Distinguished Speaker Series.

During his presentation, Dr Bagheri showcased the key research lines of NLTP, a lab focused on developing and applying advanced NLP and artificial intelligence methods to textual data. The group's projects range from bias and personality detection to explainable AI and practical applications of language models in fields such as computational social sciences and health research.

The second part of the talk centred on a recent lab study analysing the sensitivity of large language models to human values and cultural biases. Dr Bagheri explained how these models, both monolingual and multilingual, can reflect moral norms from different countries due to the training data they are exposed to. Using explainable NLP tools, the study interpreted how these systems encode moral knowledge and how it varies across cultures, providing a critical perspective on ethical behaviour and responsible AI system design.

The talk was a valuable opportunity to gain firsthand insights into advances in ethically and culturally aware NLP, as well as to encourage exchange among European academic communities interested in human-centred artificial intelligence.

THE SEARCH FOR PERSONALITY MARKERS IN TEXT: CHALLENGES FOR CONTEMPORARY PSYCHOLOGICAL SCIENCE

Associate Professor Dr Dalibor Kučera, General, Social and Educational Psychology, University of South Bohemia (České Budějovice, Czech Republic)

On July 11, 2024, ELLIS Alicante invited Dr Dalibor Kučera to give a talk as part of the ELLIS Alicante Distinguished Speaker Series.

During his talk at ELLIS Alicante, Dr Kučera introduced the field of language psychology and word usage, outlining its objectives, methods, and key findings. He focused particularly on the method of computational psycholinguistic text analysis, which has been predominant in the field over the past three decades.

He presented an overview of major projects that have explored the relationship between the use of specific linguistic categories in written or spoken texts and the personality traits of their authors.

The talk also addressed different levels of text analysis, such as linguistic morphology, syntax, and stylistics. Throughout the presentation, Dr Kučera offered comparisons across various studies, synthesised their findings, and interpreted them from both cross-cultural and cross-situational perspectives. He concluded by highlighting the most pressing challenges facing contemporary psychological science — and science more broadly — particularly in the context of language analysis and understanding human personality through text.

USING AI BIAS FOR GOOD: POVERTY AND INEQUALITY MITIGATION

Assistant Research Professor Georgina Curto Rex, Lucy Family Institute for Data & Society, University of Notre Dame (Notre Dame, United States)

On July 11, 2024, ELLIS Alicante invited Dr Georgina Curto Rex to give a talk as part of the ELLIS Alicante Distinguished Speaker Series.

This talk was framed within the interdisciplinary area of AI for Social Good, which aims to direct AI research towards advancing the United Nations Sustainable Development Goals. In particular, she presented lines of research involving AI-enabled tools designed to open new paths toward poverty reduction by addressing social discrimination. While bias mitigation in AI has generated a substantial body of literature, she argued that online biases can be useful for identifying and measuring shared beliefs that influence social policy-making.

In her presentation, she proposed research directions aimed at developing a global discrimination index against poor people (using NLP and large language models) and optimizing poverty mitigation policies through AI-based simulations (Agent-Based Modeling).
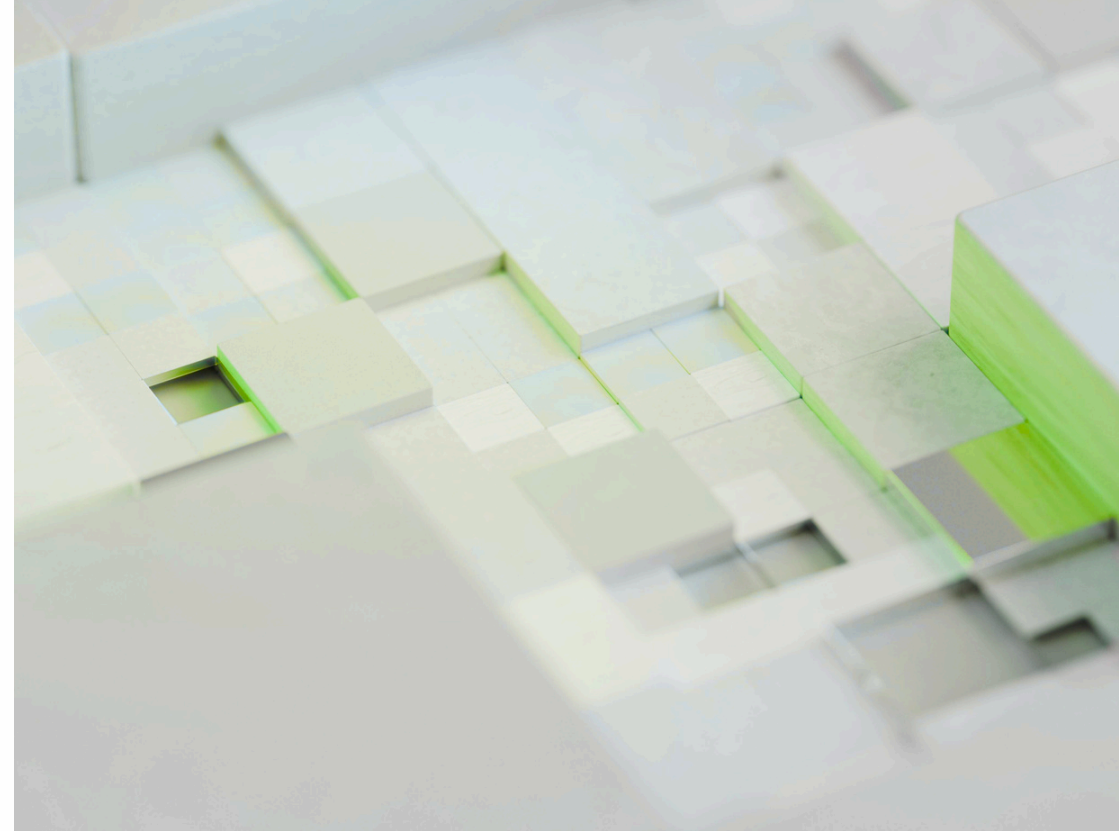
BREAKING BAD BIAS: GENDER STEREOTYPES IN GENERATIVE MODELS

Asst. Professor Noa García, Computer Vision, Osaka University (Osaka, Japan)

On September 27, 2024, ELLIS Alicante invited Prof. Noa Garcia to give a talk as part of the ELLIS Alicante Distinguished Speaker Series.

Her work addresses how generative models, when creating images from textual descriptions, make implicit assumptions that can reinforce unfair stereotypes related to gender, race, or socioeconomic status. The talk, held at the facilities of Distrito Digital 5, explored methods for measuring gender bias in text-to-image models such as Stable Diffusion, and examined how the images generated by such models can influence future computer vision systems, potentially affecting downstream tasks.

HUMAN-CENTERED AI: DESIGNING FOR URBAN AI INTERACTIONS AND LITERACY

Prof. Kaisa Väänänen, Human-Technology Interaction, Tampere University (Tampere, Finland)

On October 3, 2024, ELLIS Alicante invited Dr Kaisa Väänänen to give a talk as part of the ELLIS Alicante Distinguished Speaker Series.

Artificial Intelligence applications are permeating all areas of society, including urban environments. Although research and development in AI technologies have progressed significantly, understanding the human perspective on AI is still in its early stages.

In her talk, Dr Väänänen focused on how AI design qualities—such as proactivity, learning ability, and embodiment—can be used to enhance citizens' experiences and promote sustainable living. She also presented two case studies in urban AI interaction design: intelligent agents for residential community interactions in Nordic Superblocks, and advances in transparency and AI literacy in public urban spaces. The talk concluded with future directions for research and practice in human-centered urban AI.

"Student Speaker"Series

As part of the Student Speaker Series, ELLIS Alicante invites PhD students to present their work to the ELLIS Alicante team. In collaboration with the University of Alicante, the event is open to any student or researcher at the university.

PROBABILITY AND MACHINE LEARNING

Benedikt Höltgen
ELLIS Predoctoral Researcher, University of Tübingen (Tübingen, Germany)

On February 14, 2024, ELLIS predoctoral researcher Ben Höltgen—who holds a Bachelor's degree in Mathematics and Philosophy from LMU Munich, as well as a Master's in Philosophy of Science from LMU and a Master's in Computer Science from the University of Oxford—gave a talk as part of the ELLIS Alicante "Student Speaker Series."

Ben is co-supervised by Nuria Oliver as part of the ELLIS PhD program. Between October 2024 and March 2025, he is conducting research at ELLIS Alicante as his secondary host institution within the ELLIS PhD program. Ben is deeply interested in the intersection of machine learning, mathematics, and philosophy. As the starting point of his PhD, he aims to explore fundamental questions around algorithmic fairness.

In his talk, Ben challenged common assumptions about probability, particularly in the context of machine learning. His main argument was that there are no "correct" probabilities, as probabilities are always constructed rather than discovered.

Nevertheless, constructed probabilities can still be useful, and the assumptions under which, for instance, expected utility maximization is considered a reasonable policy (often taken for granted) deserve scrutiny.

Based on these general considerations, he outlined several perspectives on algorithmic fairness. Other key aspects of his talk included the tension between individual- and group-level fairness, and a broad notion of calibration. He also criticized how individual predictions are typically presented in machine learning. In the final part of his talk, Ben argued against the true distribution framework that underlies nearly all theoretical guarantee analyses in machine learning.

NAVIGATING PRIVACY IN HEALTHCARE TECH FOR SENIOR CARE, USER SIDE OF THE STORY

Tamara Mujirishvili
Marie Curie ITN VisuAAL, University of Alicante (Alicante, Spain)

On February 22, 2024, Tamara Mujirishvili, a predoctoral researcher from the Marie Curie ITN VisuAAL network based at the University of Alicante, gave a talk as part of the ELLIS Alicante "Student Speaker Series."

Tamara's research project focuses on perceptions of personal safety and privacy among older adults in the context of video-based lifelogging technologies. She obtained her Master's degree in Neuroscience from the University of Bordeaux (France) in 2020. Her master's thesis was conducted at the University of Cambridge, investigating how prior expectations influence perception

Active and Assisted Living (AAL) technologies aim to enhance the quality of life for older adults and support successful aging. Although video-based AAL solutions offer extensive capabilities to improve health management in old age, they also pose significant privacy risks. In response to these challenges, the AMI4AHA (Ambient Intelligence for Active and Healthy Aging) research team at the University of Alicante developed a video-based monitoring system for elderly care. This system incorporates multiple privacy-preserving filters to effectively address related risks.

In her talk, Tamara provided a user-centered perspective on this type of technology, based on a three-year research process focused on understanding the diverse viewpoints of stakeholders involved in caregiving and care reception.
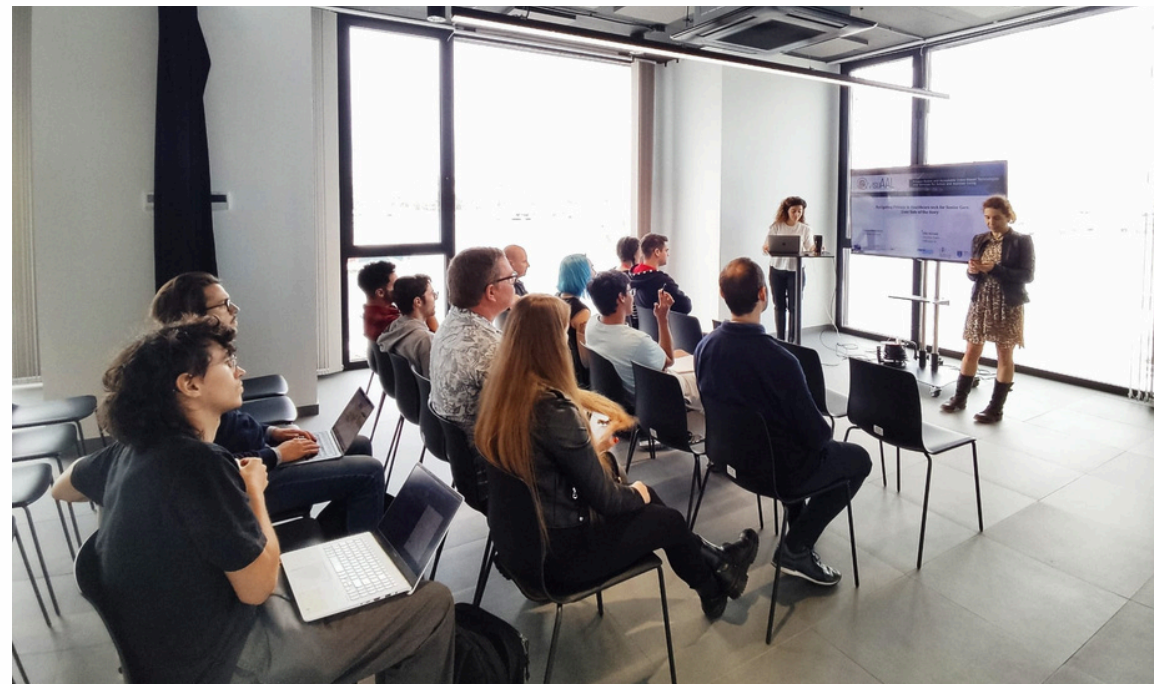
TOWARDS STUDENT-CENTRIC AI-SUPPORTED LEARNING: TEACHING CHATBOTS TO ASK THE RIGHT QUESTIONS

Lucile Favero
ELLIS Alicante Predoctoral Researcher
(Alicante, Spain)

On March 1, 2024, Lucile Favero, ELLIS predoctoral researcher at ELLIS Alicante, gave a talk as part of the ELLIS Alicante "Student Speaker Series" titled "Towards student-centric AI-supported learning: teaching chatbots to ask the right questions."

The rapid evolution of educational technology, including the integration of chatbots based on large language models (LLMs), offers a transformative opportunity to enhance learning experiences. During the talk, she presented her project "Maike," an innovative educational chatbot designed to revolutionize the traditional learning paradigm by fostering critical thinking, purposeful learning, and student self-efficacy. Unlike conventional chatbots that provide immediate answers, Maike employs Socratic dialogues to engage students, encouraging them to explore and reflect on their queries through guided questions.

During her presentation, she delved into a comprehensive literature review highlighting the importance of these technologies in education and detailed the theoretical framework underpinning Maike.
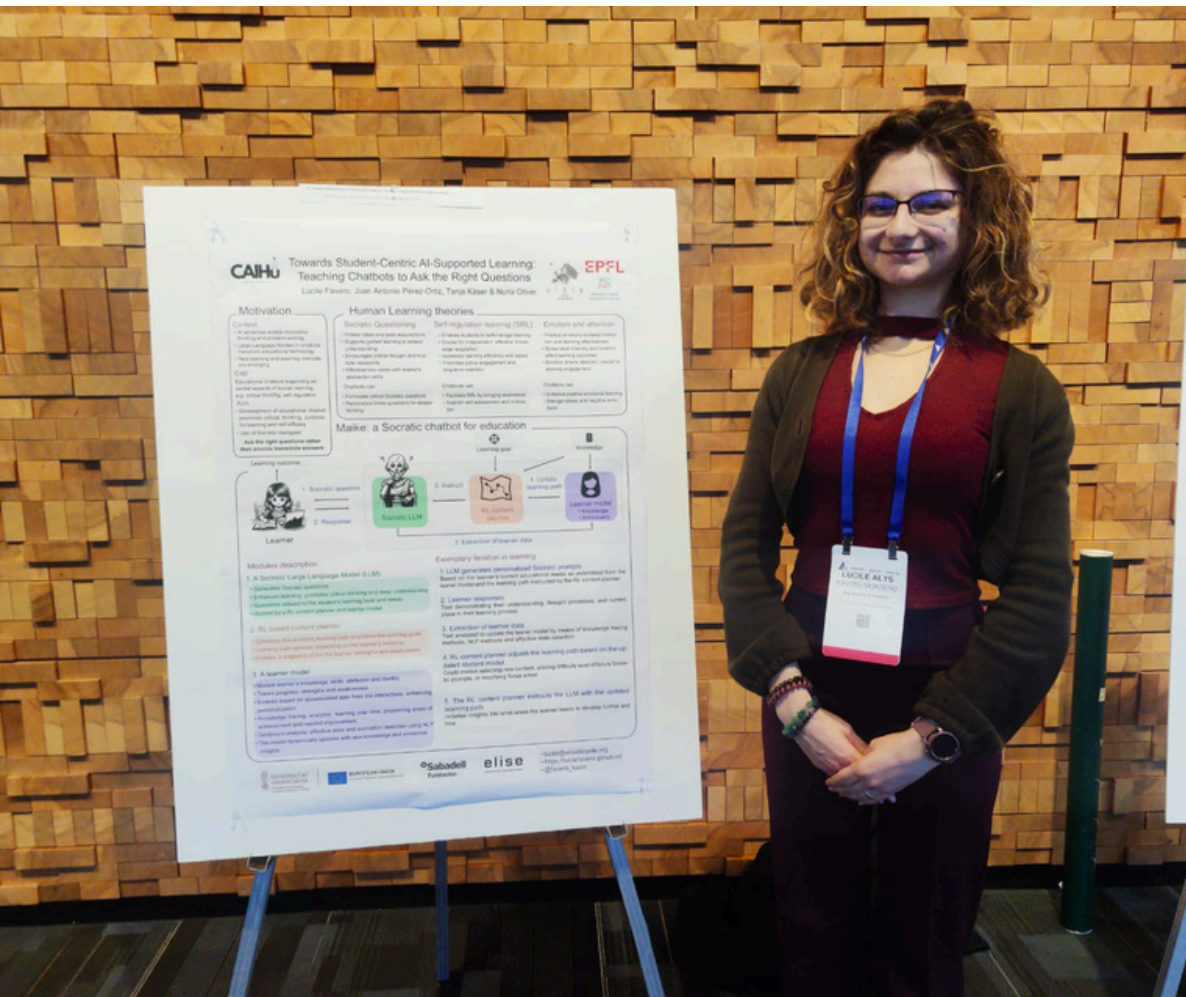
ON THE RELATIONSHIP OF FAIRNESS AND UNCERTAINTY

Kajetan Schweighofer
Predoctoral Researcher, ELLIS, University of
Tübingen (Tübingen, Germany)

On March 26, ELLIS predoctoral researcher Kajetan Schweighofer gave a talk as part of the ELLIS Alicante "Student Speaker Series."

Kajetan holds a bachelor's and a master's degree in physics, as well as a second master's degree in artificial intelligence from Johannes Kepler University Linz.

He is pursuing his PhD under the supervision of Dr Sepp Hochreiter at Johannes Kepler University Linz, with Dr Nuria Oliver as his secondary supervisor at ELLIS Alicante. His research focuses on improving uncertainty quantification in deep learning methods to make them reliable and safe when applied in critical environments. Additionally, during his stay at ELLIS Alicante in the winter semester 2023/24, Kajetan investigated the implications of uncertainty on fairness in deep learning methods and the impact on algorithmic justice of methods known as Deep Ensembles.

His presentation delved into the interaction between predictive uncertainty and algorithmic fairness, addressing questions such as: Do models exhibit the same levels of unfairness with high or low certainty? What type of uncertainty leads to unfairness? Is it possible to leverage predictive uncertainty to develop fairer models?

To ensure reliability and trust in machine learning models in real-world applications, it is crucial to assess uncertainty in their predictions (predictive uncertainty). In practical scenarios, uncertain decisions are delegated to human experts or a safe alternative is chosen. It has been found that machine learning models currently operating in our society to make inferences or predictions in important life domains—such as health, credit access, social programs, or promotions—show performance differences based on gender, race, or age, leading to unfair treatment of certain groups. Research on fairness seeks to identify, analyze, and address these systematic biases in machine learning models.

Reading Group: Human-Centric Machine Learning

ELLIS Alicante organizes the **Human-Centric Machine Learning Reading Group**, an international reading group led by them and composed of dozens of researchers from around the world. During 2024, this group has held 6 sessions dedicated to the following topics. The names of presenters who are researchers directly affiliated with ELLIS Alicante are highlighted in bold.

February 2024
Fairness and Inclusivity in Urban Transportation Design Using Reinforcement Learning
Presenter: Dimitris Michailidis
Article link

March 2024
Annotations from speech and heart rate: impact on multimodal emotion recognition (Kaushal Sharma and Guillaume Chanel), 2023
Presenter: Kaushal Sharma
Article link

June 2024
Pruning for feature preserving circuits in CNNs (Chris Hamblin, Talia Konkle, George Alvarez), 2022.
Presenter: **Julien Colin**
Article Link

July 2024
What is Beautiful is Still Good: The Attractiveness Halo Effect in the Era of Beauty Filters (Aditya Gulati, Marina Martinez-Garcia, Daniel Fernandez, Miguel Angel Lozano, Bruno Lepri, Nuria Oliver), 2024
Presenter: **Aditya Gulati**
Article link

September 2024
Describing Differences in Image Sets with Natural Language (Lisa Dunlap, Yuhui Zhang, Xiaohan Wang, Ruiqi Zhong, Trevor Darrell, Jacob Steinhardt, Joseph E. Gonzalez, Serena Yeung-Levy), 2024.
Presenter: **Piera Riccio**
Article link

October 2024
Creating Suspenseful Stories: Iterative Planning with Large Language Models (Kaige Xie, Mark Riedl), 2024.
Presenter: **Maria Hartikainen**
Article link

Participation in Conferences and Workshops

The ELLIS Alicante team has participated as organisers, speakers, or invited panellists in numerous national and international conferences and seminars, including the following.

January 2024

Dr Oliver was an invited panelist at the Hotusa Forum to discuss "Artificial Intelligence and Tourism," Madrid, Spain.

February 2024

Dr Oliver was an invited speaker at the 4th Annual Doctoral Students' Congress with a talk titled "Artificial Intelligence in the Scientific Process", held in Elche, Spain and online

.

Dr Oliver was an invited panellist at the "AI Regional Ecosystems" conference organised by ADRA, Amsterdam, the Netherlands, and online.

Dr Oliver was an invited speaker at the International Day of Women and Girls in Science event at the Neuroscience Institute, delivering a talk titled "Artificial Intelligence: My Personal Experience", Alicante, Spain.

Piera Riccio and Lucile Favero gave a talk on the occasion of the International Day of Women and Girls in Science at IES Mutxamel, Spain.

Dr Oliver was an invited speaker in the Orihuela-Alicante Diocese Faith-Culture series with a talk titled "Towards a Socially Sustainable Artificial Intelligence", Alicante, Spain.

Dr Oliver was an invited speaker at the seminar on "Strategic Aspects of Artificial Intelligence and its Regulatory Framework" with a talk titled "Towards Socially Sustainable Artificial Intelligence", CESEDEN, Madrid, Spain.

Dr Oliver delivered the opening lecture at the "Learning Schools" seminar organised by the Trilema Foundation, presenting "Artificial Intelligence: Do We Know What We're Talking About?", Madrid, Spain.

Adrian Arnaiz gave a guest lecture at the University of Notre Dame (Indiana, USA) for the course "Ethics of Data Analytics and AI", Notre Dame-IBM Tech Ethics Lab, Online.

Dr Oliver and Aditya Gulati co-organised the Bridge Program on Collaborative AI and Human Modelling at AAAI in Vancouver, Canada

.
Lucile Favero presented her work at the Collaborative AI and Human Modelling Bridge Program at AAAI in Vancouver, Canada.

March 2024

Lucile Favero gave a talk titled "Towards Student-Centred AI-Supported Learning: Teaching Chatbots to Ask the Right Questions" at the University of Alicante, Spain, and online.

Dr Oliver delivered a keynote lecture titled "Towards Socially Sustainable Artificial Intelligence" as part of the European Hipatia Award for Excellence in Science ceremony, Barcelona, Spain.

Piera Riccio was an invited panelist at the event "Artificial Intelligence and Gender Equality" organised by the Equality Unit of the Ministry of Territorial Policy and Democratic Memory, Madrid, Spain, and online.

Dr Oliver presented a talk on "Algorithmic Censorship of Art" at the International Congress on Intellectual Property and Artificial Intelligence, Madrid, Spain.

Dr Oliver participated in the consortium meeting of the H2020 ELIAS project, Amsterdam, Netherlands.

April 2024

Dr Oliver delivered a keynote lecture in the IESE MBA programme titled "Towards Socially Sustainable Artificial Intelligence", IESE, Barcelona, Spain, and online.

Dr Oliver was an invited panelist speaking on "Artificial Intelligence and Art" at the International Congress on Artificial Intelligence in Social Sciences and Humanities, Madrid, Spain, and online.

Dr Oliver moderated a panel on "Artificial Intelligence, Ethical and Legal Aspects" at the Annual Congress of COIT, Alicante, Spain.

Aditya Gulati presented his work on the attractiveness halo effect at the Behavioral Insights Bicocca Journal Club.

May 2024

Lucile Favero was invited to the event "Shaping the Future, Keeping us Human," organised by the Global Educational Forum at SEK International Schools, Madrid, Spain.

Adrian Arnaiz presented "Towards Algorithmic Fairness by Means of Instance-level Data Re-weighting Based on Shapley Values" at the DMLR Workshop of the International Conference on Learning Representations (ICLR) in Vienna, Austria.

Adrian Arnaiz presented "Structural Group Unfairness: Measurement and Mitigation by Means of the Effective Resistance" at the TrustLog Workshop of the ACM Web Conference 2024 in Singapore and online.

Piera Riccio presented "Exposed or Erased: Algorithmic Censorship of Nudity in Art" at the ACM CHI Conference on Human Factors in Honolulu, Hawaii, USA.

Dr Oliver was an invited speaker at the European Central Bank conference with a talk titled "Diversity and Artificial Intelligence," Frankfurt (online), Germany.

Dr Oliver delivered a masterclass on "Artificial Intelligence" at the Global MBA at IESE, Madrid, Spain.

Gergely Nemeth presented "Observations of Using Model Complexity Reduction as a Defence Against Membership Retrieval" at the ML-DAS Workshop, Lucca, Italy and online.

Dr Oliver was an invited panellist on "Artificial Intelligence" at the Annual Congress of the Association of Property Managers, Alicante, Spain.

June 2024

Dr Oliver was invited to moderate the debate on the film "ELLA" as part of the CNIO + Círculo de Bellas Artes cycle on cinema and science.

Dr Oliver was an invited speaker at the CNIO with a talk titled "Artificial Intelligence: My Personal Story."

Aditya Gulati presents "What is Beautiful is Still Good: The Attractiveness Halo Effect in the Era of Beauty Filters" at the International Conference on Thinking, Milan, Italy.

Dr Oliver is an invited speaker at the International Congress on TOWARDS A RESPONSIBLE DEVELOPMENT OF THE METAVERSE, with a talk titled "Biases in the Beautyverse," Alicante, Spain.

Dr Oliver, Aditya Gulati, Adrian Arnaiz and Gergely Nemeth participate in the 4th Plenary Meeting of the H2020 ELIAS project. Dr Oliver presents progress on WP3, online.

Piera Riccio co-organises the "Colors of AI" workshop at the International Conference on Computational Creativity, in Jönköping, Sweden and online.

Dr Oliver co-organises the ELLIS programme workshop "Human-centric Machine Learning" in collaboration with the ELISE and ELIAS projects, Helsinki, Finland.

Piera Riccio participates as an invited speaker in the "PhD highlights" roundtable at the ELISE wrap-up and ELLIS Community event in Helsinki, Finland.

Aditya Gulati participates as an invited speaker in the roundtable "AI PhD and Postdoc Education: ELLIS Experience" at the 4th Community Workshop 2024 & AIDA Symposium in Thessaloniki, Greece.

Dr Derner presents his work "Leveraging Large Language Models to Measure Gender Bias in Gendered Languages" at the ELLIS Summer School on Collaborative and Generative AI in Helsinki, Finland.

July 2024

Dr Derner presents his work "Leveraging Large Language Models to Measure Gender Bias in Gendered Languages" at the ELLIS Summer School on Collaborative and Generative AI in Helsinki, Finland.

Dr Oliver is an invited speaker at the Artificial Intelligence conference organised by El Economista with a talk titled "Towards Socially Sustainable Artificial Intelligence", Madrid, Spain.

Piera Riccio presents the progress made during the X School at xCoAx 2024 in Treviso, Italy.

ELLIS Alicante organises a multidisciplinary workshop of Distinguished Talks, Alicante, Spain.

Aditya Gulati presents his work "What is Beautiful is Still Good: The Attractiveness Halo Effect in the Era of Beauty Filters" at the 10th International Conference on Computational Social Science in Philadelphia, USA.

Adrian Arnaiz presents his work titled "Structural Group Unfairness: Measurement and Mitigation by Means of the Effective Resistance" at the 10th International Conference on Computational Social Science in Philadelphia, USA.

Adrian Arnaiz delivers a tutorial on "Graph Learning: Principles, Challenges, and Open Directions" at ICML 2024, Vienna, Austria.

August 2024

Lucile Favero, Julien Colin, Aditya Gulati, and Gergely Nemeth present their work at the ELLIS Doctoral Symposium 2024, Paris, France.

Dr Oliver is an invited speaker at the Responsible AI Day at KDD'24 with a talk titled "Biases, Humans, and Algorithms," Barcelona, Spain.

September 2024

Dr Oliver is an invited speaker at the Human-centric Machine Learning workshop at ECML-PKDD'24 with a talk titled "Biases, Humans, and Algorithms," Vilnius, Lithuania, and online.

Dr Oliver is an invited speaker at the RESCOM 2024 conference with a talk titled "Towards Socially Sustainable Artificial Intelligence," Alicante, Spain.

Dr Oliver, Aditya Gulati, and Gergely Nemeth participate in the H2020 ELIAS project consortium meeting, Bologna, Italy, and online.

Dr Oliver receives the Trifermed 2024 Award for Transformational Impact and delivers a lecture on "Artificial Intelligence," Barcelona, Spain.

Dr Oliver is an invited speaker at the EuGMS 2024 International Congress of the European Geriatric Medicine Society with a talk on "Artificial Intelligence and Ageing," Valencia, Spain.

Dr Oliver is an invited speaker at the 2nd National Congress on Sterility and Infertility of SEGO with a talk on "Artificial Intelligence and Reproductive Medicine," Alicante, Spain.

Adrian Arnaiz delivers a guest lecture in the Delve Deep Learning Education course, online.

Aditya Gulati presents his work "Lookism: The Overlooked Bias in Computer Vision" at the ECCV 2024 conference workshop "Fairness and Ethics Towards Transparent AI: Facing the Challenge through Model Debiasing," Milan, Italy.

Piera Riccio presents her work "An Art-Centric Perspective on AI-Based Content Moderation of Nudity" at the ECCV 2024 conference workshop "Artificial Intelligence for Visual Arts," Milan, Italy.

Piera Riccio presents her work "Exploring the Boundaries of Content Moderation in Text-to-Image Generation" at the ECCV 2024 conference workshop "Critical Evaluation of Generative Models and Their Impact on Society," Milan, Italy.
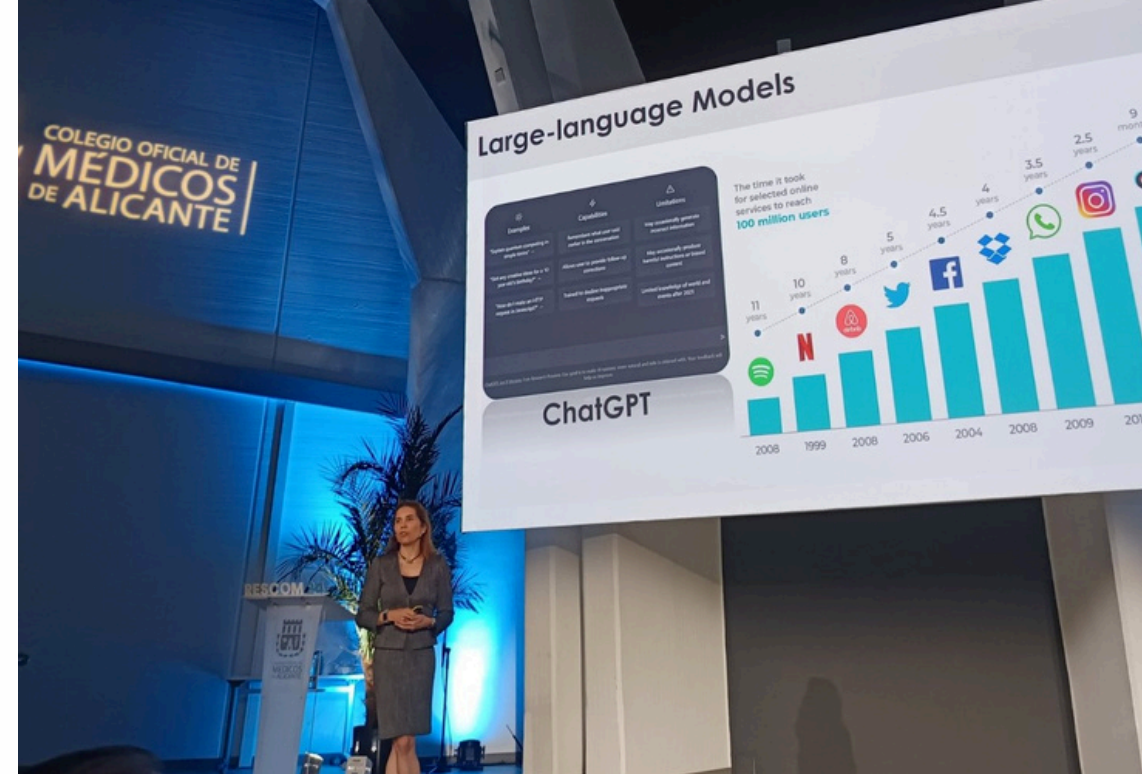
October 2024

Dr Oliver is the moderator of a panel on the Social Impact of Artificial Intelligence at the 10th edition of the Cremades & Calvo Sotelo and Prodware Dialogues, El Escorial, Spain.

ELLIS Alicante is co-organiser of the 3rd European Artificial Intelligence Forum, Alicante, Spain.

Piera Riccio presents her work "Mirror, Mirror on the Wall, Who Is the Whitest of All? Racial Biases in Social Media Beauty Filters" at the Centre for Gender Studies, Lancaster University, United Kingdom.

November 2024

Adrián Arnaiz and Erik Derner are invited speakers at the 5th Conference on Intellectual and Developmental Disability with a talk titled "Artificial Intelligence and Human Wellbeing," Alicante, Spain.

Piera Riccio participates as an invited speaker in a roundtable on "Artificial Architectures (AA)" at the School of Architecture, Technical University of Madrid, Spain.

Aditya Gulati delivers an invited lecture in the Managerial Marketing Analytics course at the Aalto University School of Business, Helsinki, Finland, and online.

Dr Oliver is an invited speaker at the BNAIC 2024 conference with a presentation titled "Towards a Fairer World – Uncovering and Addressing Human and Algorithmic Biases," Utrecht, Netherlands.

Dr Oliver participates in a "Conversation on AI," Alicante, Spain.

Julien Colin presents his work on human-centred interpretability at the Artificial and Natural Intelligence Toulouse Institute (ANITI) during ANITI Days, Toulouse, France.

December 2024

Dr Oliver is an invited speaker at the "Embodied Intelligence" Forum organised by the Bankinter Innovation Foundation, delivering a talk titled "Human Biases and Artificial Intelligence," Madrid, Spain.

Other scientific outreach activities

Media Coverage

The work of ELLIS Alicante has been featured in **over a hundred publications across print and digital media**, as well as receiving extensive coverage on television and radio, both nationally and internationally. Below is a brief selection showcasing this impact in 2024.

elEconomista.es
Oliver (ELLIS): "La IA ha dejado de ser una disciplina tecnológica para pasar al ámbito político"

- Esta tecnología está involucrada en la toma de decisiones como sentencias o tratamientos médicos

Foro Empresarial 'La reinvención de la Inteligencia Artificial' - Cierre
elEconomista

ver en DAILYMOTION

EL PAÍS
ARTIFICIAL INTELLIGENCE > | TRIBUNE ⓘ
Digital natives aren't born, they're made: towards a society of digital scholars

The skills and knowledge necessary to be able to utilize technology to solve problems, come up with solutions and create new services must be taught. We can't assume that people will learn these skills as mere users

Two girls use their phones to take a photo of the flowering trees in Piazza Tommaseo on March 14, 2024 in Milan, Italy.
EMANUELE CREMASCHI (GETTY IMAGES)

EL PAÍS
INTELIGENCIA ARTIFICIAL > | TRIBUNA ⓘ
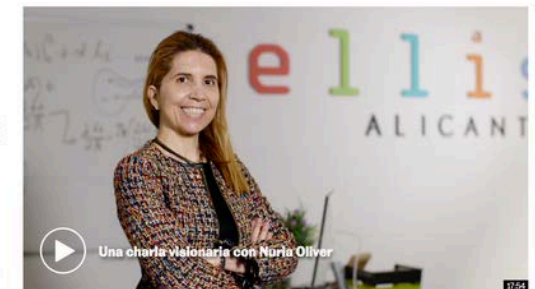### Ⓔ Humanos robotizados, algoritmos humanizados

A medida que la IA se vuelve más sofisticada, también lo hacen nuestras expectativas sobre su comportamiento. No nos conformamos con meros ejecutores de órdenes, buscamos capacidad de comprender, empatizar y adaptarse a nuestras necesidades

EL PAÍS
INTELIGENCIA ARTIFICIAL
Nuria Oliver, ingeniera: "Debemos combatir la cultura tremendamente misógina y sexista del sector tecnológico"

La experta en inteligencia artificial habla en el Proyecto Tendencias sobre el oligopolio de Estados Unidos y China en materia de IA, los retos de Europa y la necesidad de generar, atraer y retener talento

Una charla visionaria con Nuria Oliver
La ingeniera y experta en inteligencia artificial Nuria Oliver.
Vídeo: EPV

elPeriódico
Percepción digital
Los filtros de belleza hacen que las mujeres sean percibidas como menos inteligentes que los hombres, según un estudio

- Una investigación realizada en España explora el impacto de los sesgos de género en los embellecedores digitales

Television Coverage

La aventura del saber - 09/04/24
00:57:45   09/04/2024

rtve Igualdad

Objetivo Igualdad
Sesgos y violencias machistas de la IA: "No me da miedo la tecnología, me da miedo la maldad humana"

▸ Los algoritmos contienen sesgos de género que perjudican a mujeres en procesos laborales y económicos
▸ En 2021, el 80% de los nuevos doctores en IA fueron hombres
▸ 'Inteligencia Artificial' en *Objetivo Igualdad*, el domingo a las 14:25h en Canal 24 horas

14/01/2024 | 09:19 horas   Por   LUCÍA BLÁZQUEZ

laSexta
DÍA DE LA ELIMINACIÓN DE LA DISCRIMINACIÓN RACIAL

Los algoritmos racistas: cuando se equipara "feo" a persona negra y los filtros blanquean caras racializadas

Diferentes expertas hablan sobre cómo la sociedad alimenta buscadores como Google con etiquetas que vinculan ser blanco con éxito social, y ser negro o de otro color de piel con conceptos negativos como pobreza. También critican los filtros de belleza que tienden a blanquear las caras.

informativos Telecinco   Ciencia y Tecnología

Nuria Oliver, sobre la primera Ley de Inteligencia Artificial en Europa: "El punto de partida es que los derechos fundamentales sean respetados"

Radio Coverage

Nuria Oliver liderará uno de los grupos que creará un código de buenas prácticas de la IA en Europa

El borrador final podría estar finalizado para el próximo mes de abril

Nuria Oliver habla sobre la normativa de la IA.

Denis Rodríguez
Alicante – Publicado el 06 oct 2024, 10:00
2 min lectura

Ondacero > Programas > Más de uno > Secciones > La Ciencia con Aparici y Cremades

CON ALBERTO APARICI Y SANTI GARCÍA CREMADES

La relación de Skynet de 'Terminator' con nuestras Inteligencias Artificiales

Con **Alberto Aparici y Santi García Cremades** repasamos algunos de los **futuros que imaginó el cine** para nosotros. ¿Cómo se veía el año 2024 en las películas? ¿Han acertado en algo o no? Hablamos de **Terminator** y su relación con la Inteligencia Artificial que tenemos hoy en día. También, repasamos **'Blade Runner'**, la saga de **'Star Trek'**, **'Fahrenheit 451'** y **'Regreso al Futuro'**. Hablamos de todo esto con **Nuria Oliver, experta en inteligencia artificial**, doctora por el Instituto Tecnológico de Massachusetts (MIT) y cofundadora y directora de ELLIS Alicante, una fundación dedicada a la investigación en IA responsable y para el Bien Social.

ondacero.es
Madrid | 12.09.2024 12:34

LAS MAÑANAS DE RNE - JOSEP CUNÍ

Ley de IA: la importancia de categorizar sus riesgos

13/03/2024 **05:47**

Nuria Oliver, experta en inteligencia artificial: "La regulación es un instrumento para asegurarnos que el desarrollo tecnológico sí va a conllevar un progreso"

Hoy se vota en la Eurocámara el Reglamento Europeo de la Inteligencia Artificial (IA). **Nuria Oliver, co-fundadora y directora de la Fundación ELLIS Alicante y experta en IA**, ha estado en Las Mañanas de RNE para analizar esta norma. Esta norma, según Oliver, es pionera a nivel mundial, y muchos países miran a esta regulación mientras trabajan en sus normas. **"Es el primer esfuerzo regulatorio para regular la inteligencia artificial de manera transversal y en la categorización de los sistemas de inteligencia artificial en base al riesgo que puedan presentar para la sociedad"**, explica esta experta. Oliver considera que no todos los sistemas debería estar sometidos al mismo escrutinio: "La IA es una disciplina transversal, por eso es necesaria esta estimación del nivel de riesgo en función de donde se aplique".

rne

Collaborations and Partnerships

ELLIS Europe

As a member of the European ELLIS network of scientific excellence, ELLIS Alicante has participated in and contributed to the network through various activities and organisational and management committees.

Membership on the ELLIS Europe Board

Firstly, through Dr Oliver's participation in the governing body of ELLIS Europe as Vice-President. This position involves taking part in weekly working meetings to plan and organise the activities and initiatives launched by ELLIS Europe across its various pillars: ELLIS research programmes, the ELLIS doctoral programme, ELLIS units, and industrial alliances.

Human-centric Machine Learning ELLIS Research Program

Secondly, ELLIS Alicante actively contributes to the ELLIS research programme called "Human-Centric Machine Learning" (HCML), since Dr Oliver is co-director of this programme alongside Professor Plamen Angelov from Lancaster University (UK) and Professor Adrian Weller from the University of Cambridge and the Alan Turing Institute (UK).

ELLIS Alicante organised a workshop in June 2024 in Helsinki (Finland) as part of the ELLIS HCML programme, described in the section "Participation in Conferences or Similar Events" of this report.
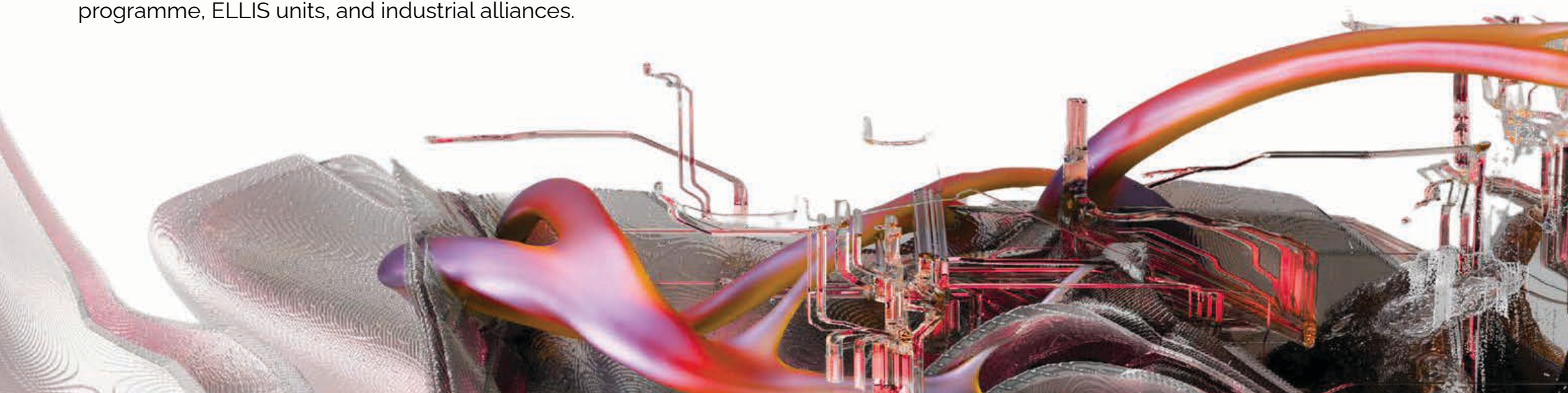
ELLIS PhD / Postdoc Program

Thirdly, ELLIS Alicante is an active participant in the ELLIS doctoral and postdoctoral programme, serving as the primary institution for six of the students enrolled in the programme, one postdoctoral researcher, and as the secondary host institution for three additional students of this prestigious doctoral programme. All students in the ELLIS doctoral programme are co-supervised and affiliated with both a primary and a secondary institution, where they undertake a stay of at least six months.

Hosting these doctoral students and the postdoctoral researcher has involved arranging and financing all logistical procedures for their relocation to Alicante (travel) and handling the necessary administrative processes for their admission to the University of Alicante's doctoral academic programme, as well as their affiliation with ELLIS Alicante: certified copies and sworn translations of academic transcripts, academic registration, etc.

The ELLIS doctoral programme aims to generate synergies among all its students and between them and the researchers who are members of the ELLIS network, as an essential part of its objective to create a highly collaborative and interconnected research community across Europe. Consequently, the students from ELLIS Alicante have participated in the programme's most important annual event, the ELLIS Doctoral Symposium (EDS), which the Foundation organised in 2022 and which was held in Paris in 2024
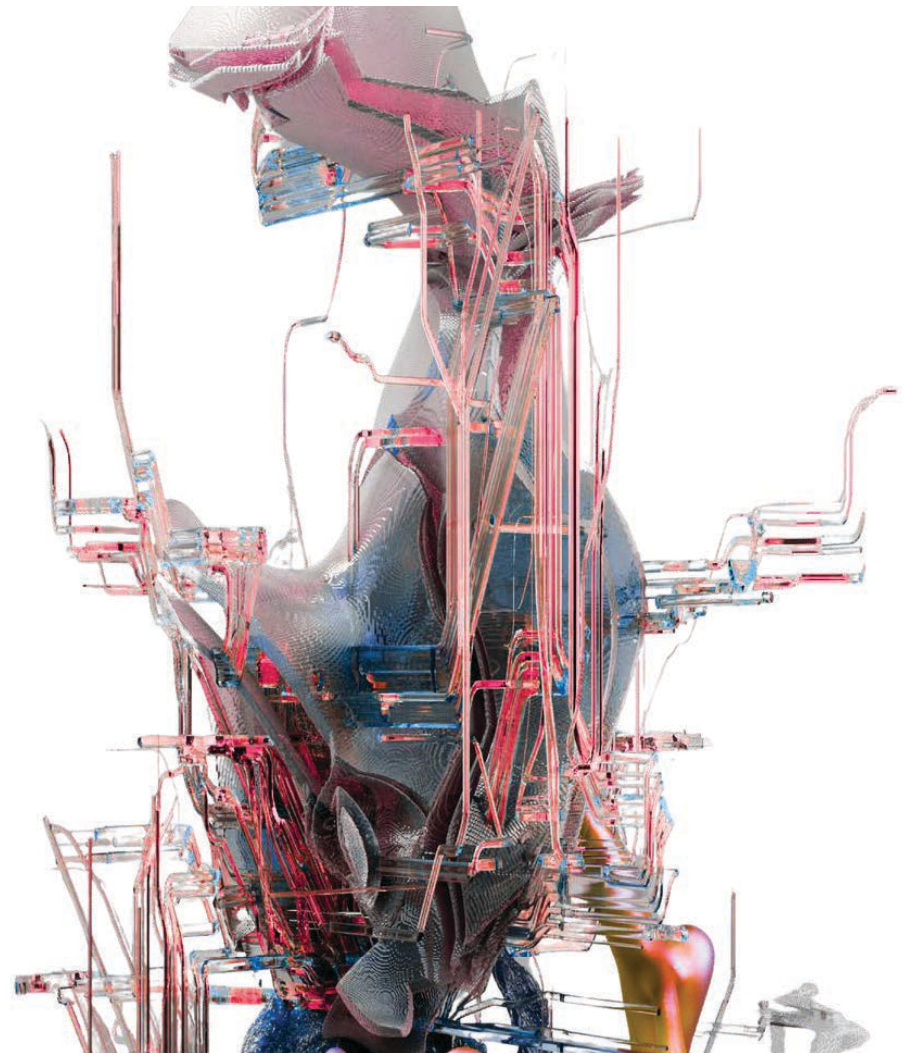
Red NAIXUS

In May 2022, ELLIS Alicante, together with three other globally renowned institutions (IRCAI, CETIC, and Data-Pop Alliance) and under the auspices of UNESCO, launched NAIXUS, an international network of scientific excellence in Artificial Intelligence for Sustainable Development.

By 2030, AI will measurably influence and impact over 8.5 billion people, across all sectors and various human and terrestrial ecosystems on an unprecedented scale.

According to a study published in Nature, AI could help achieve 79% of the Sustainable Development Goals (SDGs); more specifically, it can enable the attainment of 134 targets across all goals, but it may also hinder 59 targets.

Since sustainable development challenges affect all countries differently at both macro and micro levels, this means that each country is developing its own scientific expertise in AI on a specific set of development issues.

Some of these overlap across different continents and regions, while others are location-specific. Therefore, there is an opportunity to connect research groups working on these topics and leverage diverse solutions, case studies, skills, and competencies in the field of development and AI. That is the aim of NAIXUS — to promote the use of AI for Social Good. Throughout 2024, ELLIS Alicante has continued its membership in the NAIXUS Network.
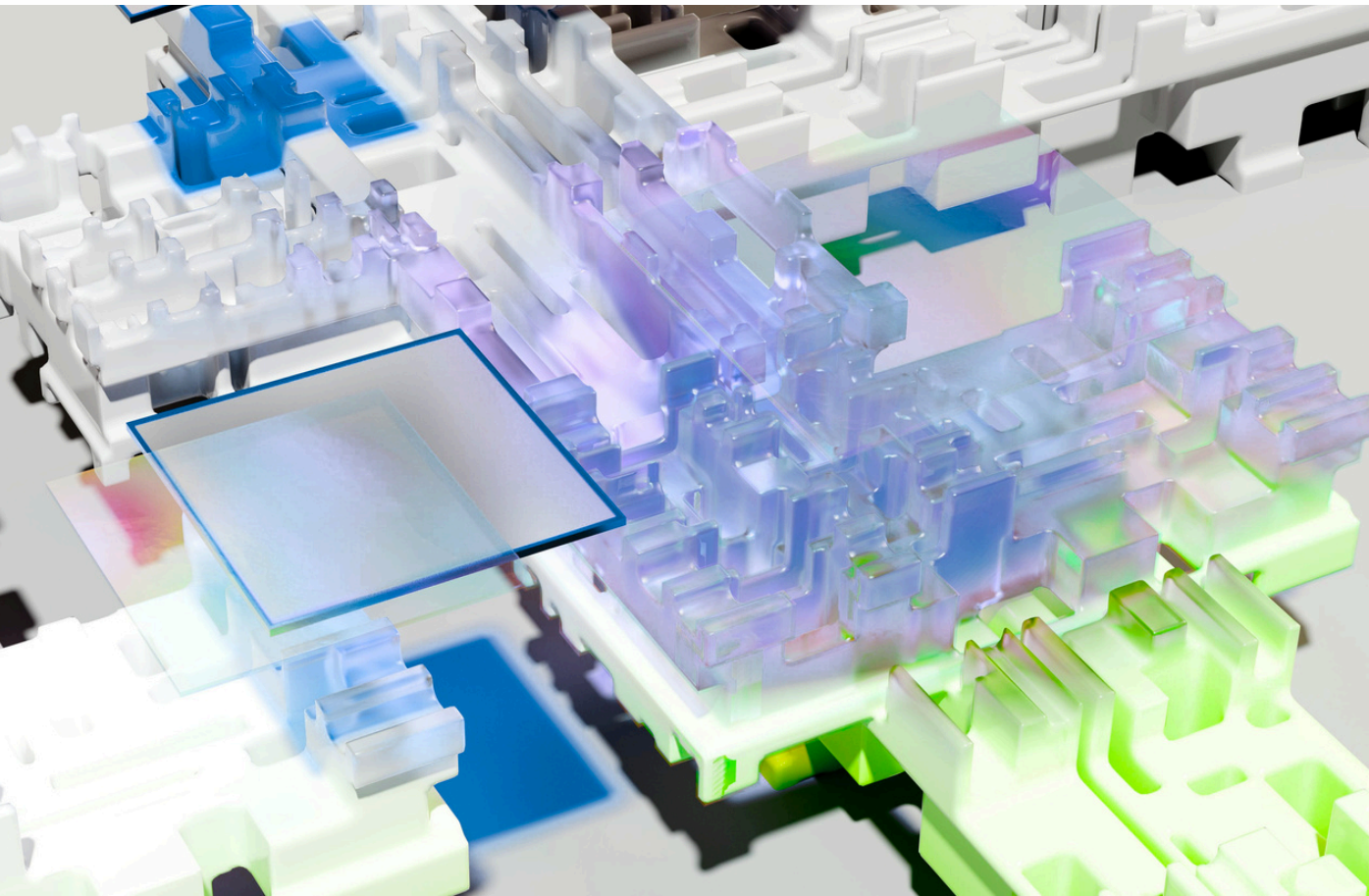
Collaborations

During 2024, ELLIS Alicante has maintained existing collaborations and established new partnerships and alliances with researchers affiliated with the following academic, scientific, and corporate entities.

Scientific Collaborations

Scientific Advisory Board

Alan Turing Institute and University of Cambridge

Spanish National Research Council (CSIC)

Cornell University

European Commission

Bruno Kessler Foundation

Max Planck Institute

MIT Media Lab

Pompeu Fabra University

ELLIS secondary institutions

Czech Technical University in Prague

Swiss Federal Institute of Technology Zurich (ETH Zurich)

Bruno Kessler Foundation (FBK)

Johannes Kepler University Linz

Max Planck Institute for Software Systems

University of Tübingen

University of Sussex

École Polytechnique Fédérale de Lausanne (EPFL)

Visiting research institutions

University of Turin

Max Planck Institute for Software Systems

Bruno Kessler Foundation

University of Notre Dame

Utrecht University

University of South Bohemia in České Budějovice

Polytechnic University of Turin

Tampere University

University of Mons

Czech Institute of Informatics, Robotics and Cybernetics

D3 Center, Osaka University

Institutional Relations

Collaboration agreements

Cardenal Herrera University

CaixaBank

Official Medical Association of Alicante

Banco Sabadell

Sabadell Bank Foundation

Max Planck Institute for Software Systems

Alfonso X El Sabio University

Esperanza Pertusa Foundation

IESE Business School

Intel

ONCE Foundation

Ministry of Innovation, Industry, Trade and Tourism of the Generalitat Valenciana

Nippon Gases

University of Alicante

Miguel Hernández University

University of Alicante Language Centre

University of Trento

Memberships

Climate Foundations

Registry of Innovative Entities on the APTE Private Blockchain Network

Spanish Technology Platform for Disruptive Technologies

Digital District of the Valencian Community

NAiXUS

Research

ELIAS – European Lighthouse of AI for Sustainability (EU H2020)

Aalto University

Bitdefender SRL

Robert Bosch GmbH

Robert Bosch Kft

Hellas Research and Technology Centre

Czech Technical University in Prague (CTU)

ETH Zurich

Engineering Group

Bruno Kessler Foundation

Hasso Plattner Institute

IBM Ireland

Italian Institute of Technology

Jozef Stefan Institute

IDIAP Research Institute

National Institute for Research in Digital Science and Technology

Paris Polytechnic Institute

Polytechnic University of Bucharest

Max Planck Institute

University of Manchester

NCBR Ideas

Politecnico di Milano

University of Copenhagen

University of Modena and Reggio Emilia (Unimore)

University of Milan

University of Tübingen

Umeå University

University of Trento

University of Amsterdam

University of Toulouse

Scientific Publications

University of South Bohemia, Czech Republic

Czech Technical University in Prague (CTU), Czech Republic

University of Alicante, Spain

École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

University of Notre Dame, USA

King County Public Health Department, Seattle, USA

Alexander von Humboldt Institute for Internet and Society, Germany

Swiss Federal Institute of Technology Zurich (ETH Zurich), Switzerland

Jaume I University, Spain

Polytechnic University of Catalonia, Spain

State Labour and Social Security Inspectorate, Spain

Brown University, USA

Harvard University, USA

University of Toulouse, France

National Centre for Scientific Research (CNRS), France

Interuniversity Microelectronics Centre (IMEC), Belgium
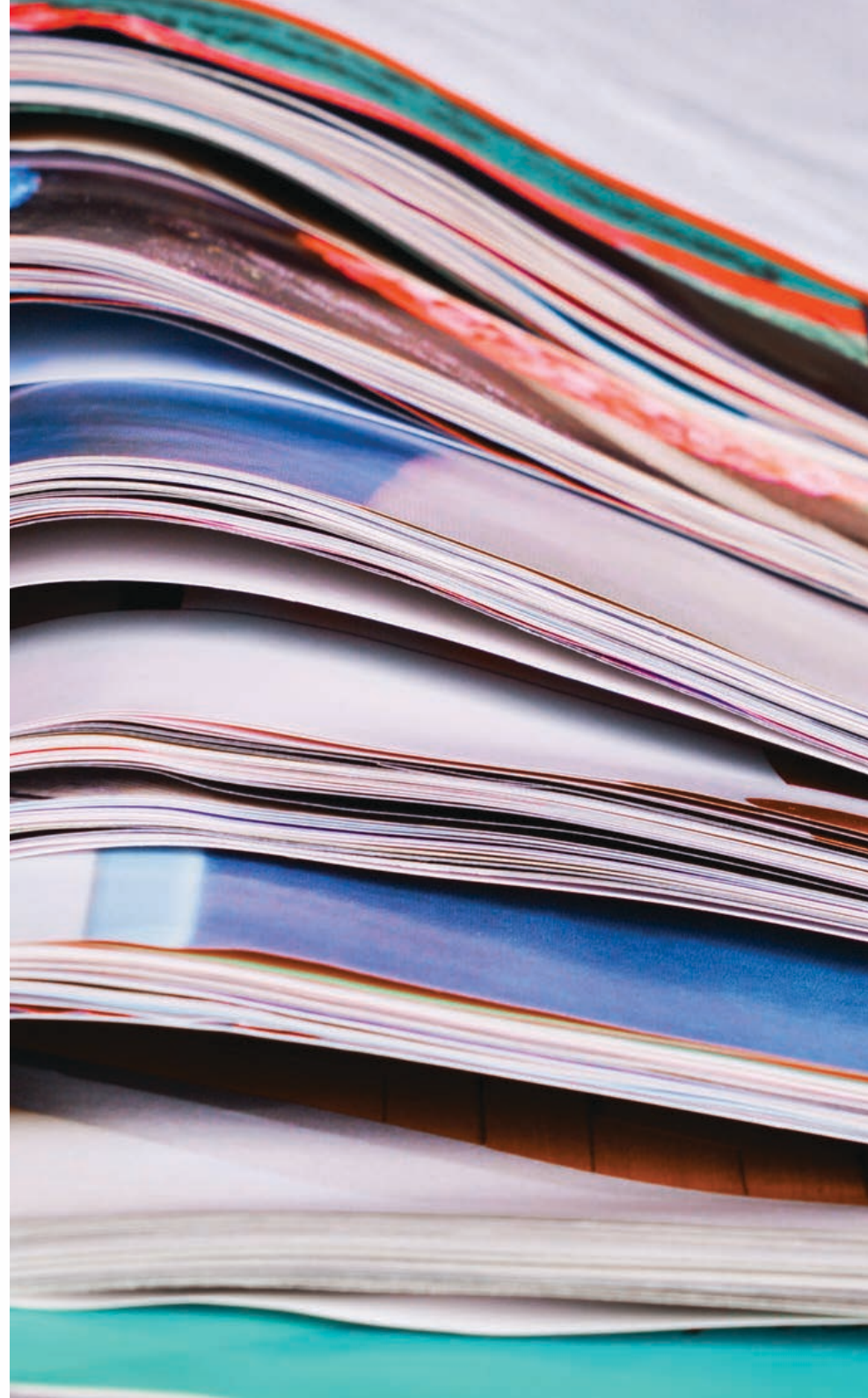
Funding and sponsorships

Public Entities

Generalitat Valenciana

To comply with the Valencian Strategy for Artificial Intelligence, published in 2019 and approved by the Consell on 13 May 2022, the Generalitat Valenciana committed, among its actions, to financially support with €1.5 million annually the creation of an ELLIS unit (European Laboratory for Learning and Intelligent Systems). This unit aims to enable collaboration with leading researchers across Europe and foster synergy between AI research, the entrepreneurial ecosystem, and technology transfer.

Thus, the Generalitat Valenciana has become a key ally, alongside other European public institutions, in achieving the ambitious international objectives of the ELLIS network—particularly in establishing a European regional counterbalance that retains, attracts, and trains outstanding research talent in Artificial Intelligence, thereby contributing to the improved economic development of our society

This commitment was formalised in the Resolution dated 22 June 2020 by the Minister of Innovation, Universities, Science and Digital Society, which approved the Strategic Grants Plan for the period 2020–2023. Currently, it is reaffirmed by the Resolution dated 8 May 2023 from the same Ministry, approving the Strategic Grants Plan for the period 2023–2025.

In line with this commitment and the aforementioned Resolution, Law 8/2023, dated 27 December, on the Generalitat's Budget for the 2024 financial year established under Subprogram 121H00 – Advancement of the Digital Society, specifically in the Details of Transfers and Subsidies, two earmarked funding lines: S1600, "Support for the operation and current expenses of the ELLIS Alicante Node", amounting to €400,000; and S1233, "Support for excellent artificial intelligence research at the ELLIS Node", amounting to €100,000. In both cases, the beneficiary entity is the Fundación de la Comunitat Valenciana Unidad ELLIS Alicante.

The funding obligations under these lines are governed by an agreement signed on 4 December 2024 between the Generalitat Valenciana, through the Ministry of Innovation, Industry, Commerce and Tourism, and the Fundación de la Comunitat Valenciana Unidad ELLIS Alicante (ELLIS Alicante), entitled "Agreement between the Generalitat, via the Ministry of Innovation, Industry, Commerce and Tourism, and the Fundación de la Comunitat Valenciana Unidad ELLIS Alicante for the support of its activities and investments during the 2024 financial year."

European Commission

ELIAS (The European Lighthouse of AI for Sustainability) is a Horizon 2020 project funded by the European Commission, in which ELLIS Alicante participates as one of 34 institutions from 17 countries.

The scientific excellence of the consortium is supported by the prominent standing of its members in their respective fields, as measured by their international visibility and impact.

ELIAS is firmly committed to advancing fundamental research in Artificial Intelligence, aiming to address key questions such as how AI can reduce computational costs, model the social impacts of policy decisions, and affect individuals on a personal level. This approach involves deep integration between academic research and industry-driven applications, creating a synergy that drives progress in the field.

The project defines three core research areas for sustainable AI: sustainable AI for the planet, sustainable AI for society, and sustainable AI for people. ELLIS Alicante contributes to the last of these—sustainable AI for people—with projects focused on algorithmic discrimination and bias, human cognitive biases, and the relationship between AI and privacy, particularly in the realm of federated learning.

The project spans 48 months, beginning in September 2023, with a budget exceeding €11 million. Its ultimate goal is to position Europe as a global leader in AI research, ensuring that environmental impact, social relevance, and individual well-being are central considerations throughout the entire development process.

Private Institutions

In order to help the ELLIS Alicante Unit Foundation carry out its mission, there are different possible forms of contributions and donations from companies and other collaborating institutions, aligned with the ELLIS Europe sponsorship program.

We are tremendously grateful for the generous contributions of our donors and patrons. Your support of ELLIS Alicante's mission, vision and activity is instrumental to the success of our organization.

GOLD

SILVER

BRONZE

AMIGOS DE ELLIS ALICANTE

MICRO-FINANCIACIÓN

Sponsorship levels

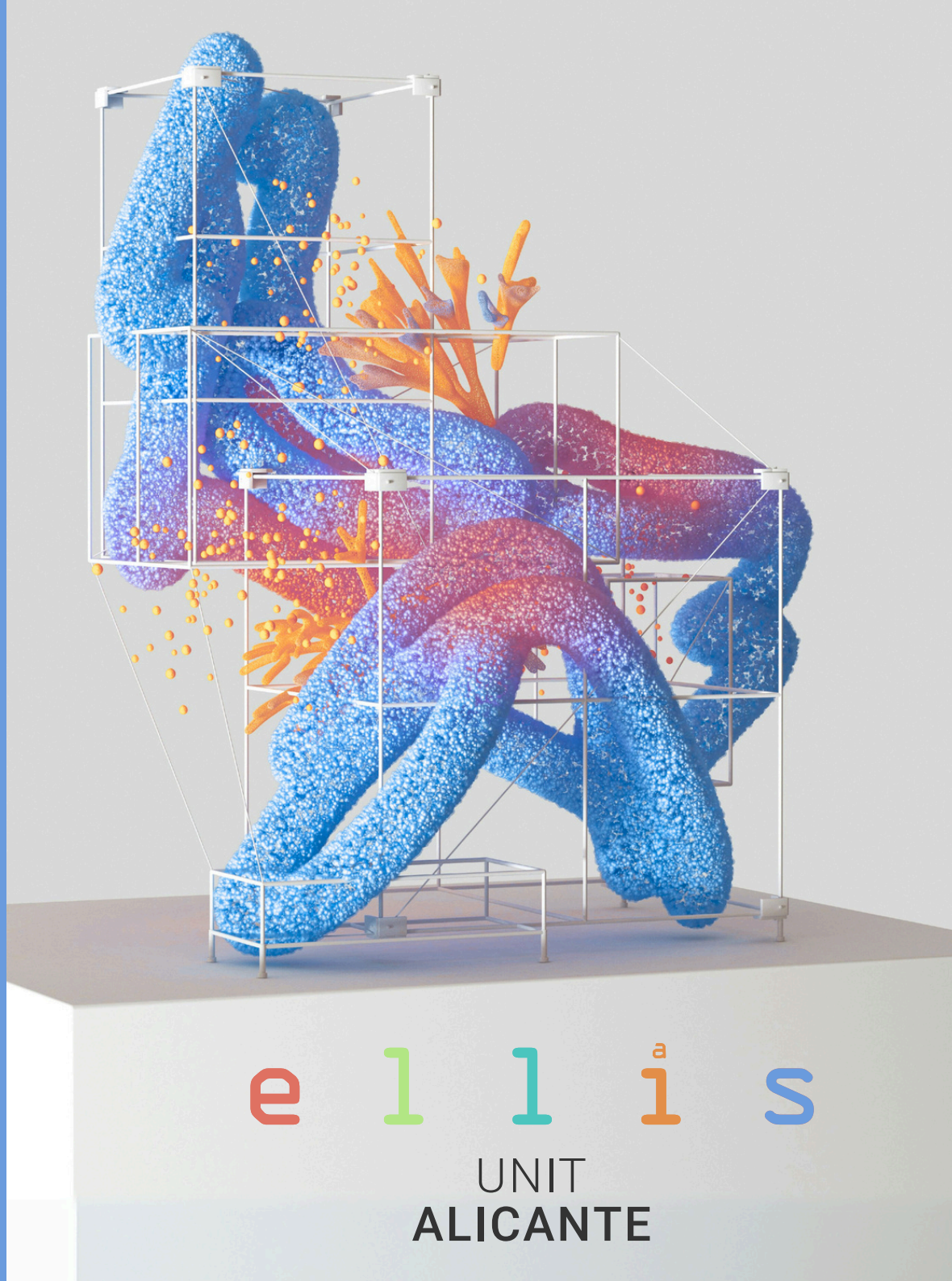| Level	Annual contribution (+VAT)	Benefits
PLATINUM	500,000 € first year 200,000 € next years (min 2 years) 50,000 € ELLIS Europe	• Option to have a seat on the board of trustees of ELLIS Alicante • Logo in all communication material • Annual gala dinner with distinguished guests • Reputational impact and CSR/ESG • Speaking slot at ELLIS Doctoral Symposium • Job postings in the ELLIS Europe newsletter • Subscription to the ELLIS Europe newsletter
GOLD	100,000 € per year 20,000 € ELLIS Europe	• Logo in all communication material • Visit to ELLIS Alicante • Reputational impact and CSR/ESG • Speaking slot at ELLIS Doctoral Symposium • Job postings in the ELLIS Europe newsletter • Subscription to the ELLIS Europe newsletter
SILVER	50,000 € per year 10,000 € ELLIS Europe	• Logo in all communication material • Reputational impact and CSR/ESG • Up to 5 job postings in the ELLIS Europe newsletter • Subscription to the ELLIS Europe newsletter
BRONZE	25,000 € per year 2,000 € ELLIS Europe	• Logo in all communication material • Reputational impact and CSR/ESG • Subscription to the ELLIS Europ newsletter
FRIENDS	5,000 - 30,000 € per year	• Recognition on the website and annual report
MICRO-FINANCING	Up to 5,000 € per year	• Recognition on the website and annual report

Learn more about ELLIS Alicante

Video created by Rudy van Os
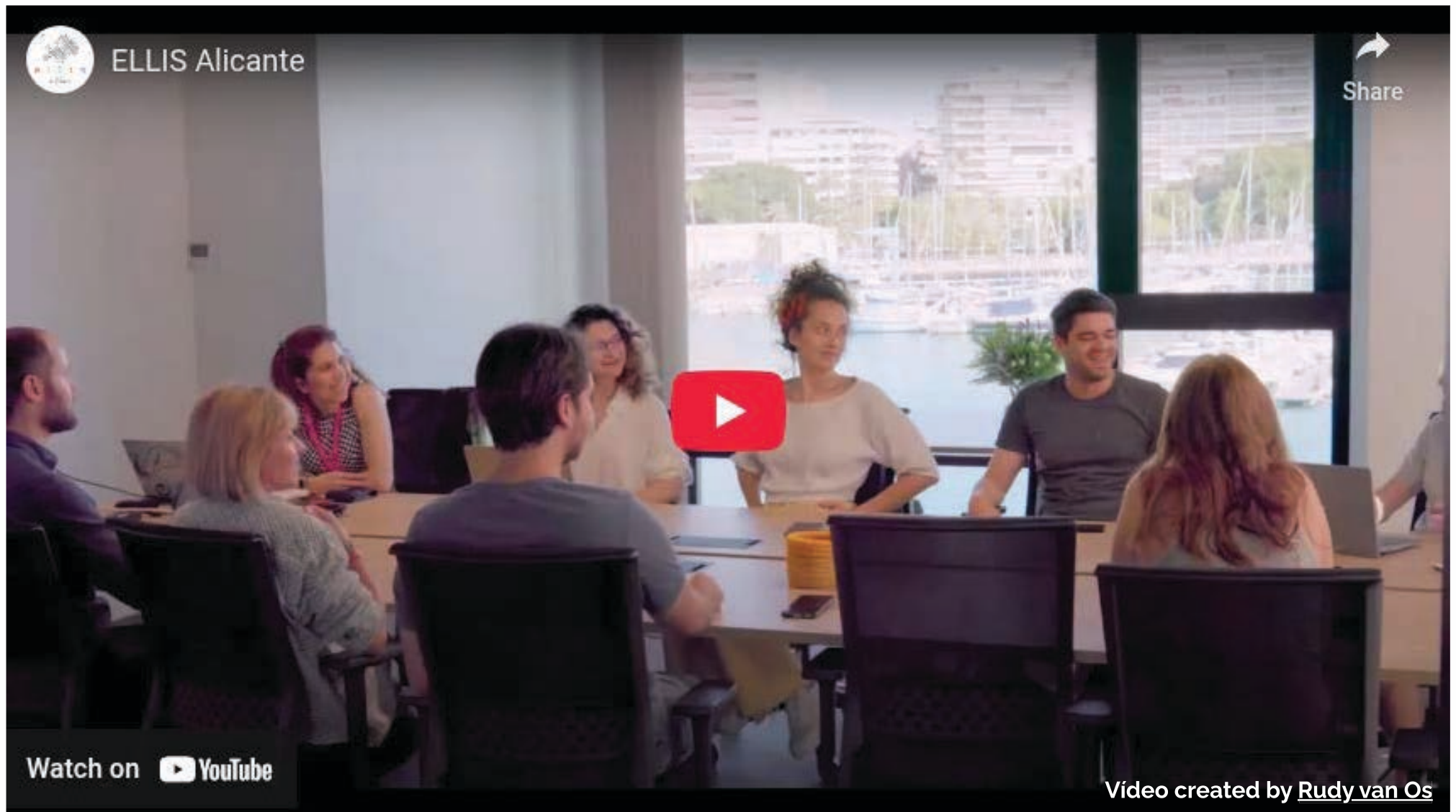
Thank you for making ELLIS Alicante possible

ellis
UNIT
ALICANTE