

ANNUAL REPORT 2025

e 1 1 i s

UNIT
ALICANTE





Table of contents

Presentation	3
Milestones and awards	5
About ELLIS Europe	12
About ELLIS Alicante	13
Governance	20
Team	22
Programs and research projects	29
Scientific publications 2025	54
Communication Activities and Scientific and Technical Outreach Initiatives	58
PhD thesis defenses	59
Research events	68
Serie: "Distinguished Speakers"	77
Serie: "Student Presentations"	86
Other science outreach activities	89
Media coverage in press, television and radio	101
External relations, collaborations and partnerships	105
Collaborating organizations	113
Funding and Sponsorships	119

Letter by the President of the Board



**CHRISTOF
BEAUPOIL**

President of the Board of
ELLIS Alicante

Dear friends,

We are pleased to present the 2025 Activity Report of the Fundación de la Comunitat Valenciana Unidad ELLIS Alicante.

Throughout 2025, ELLIS Alicante continued to strengthen its position as an international reference in responsible, human-centric Artificial Intelligence research further advancing our mission to develop AI methods that contribute to the social good and to

scientific progress from a humanistic perspective. Our scientific, training and knowledge transfer activities continued to grow thanks to the talent and commitment of our research team, the support of our partner institutions, and our joint work with the ELLIS network.

Among the most significant milestones of 2025, one in particular stands out: the successful defense of the first four doctoral theses by students in the ELLIS doctoral program affiliated with ELLIS Alicante. This achievement marks a historic moment for our institution and provides tangible evidence of the maturity and excellence attained by our international doctoral program. These first theses symbolize not only the academic success of our PhD students, but also the consolidation of a research training model based on scientific excellence, European collaboration and the social impact of Artificial Intelligence. In addition, we strengthened our presence in the international scientific community through high-impact publications and participation in leading conferences.

We also continued our work as active participants in the Horizon Europe ELIAS project, dedicated to research into sustainable Artificial Intelligence; in the RESUMAS Centre of Excellence in Research, focused on Responsible Artificial Intelligence; and in the PERTE for the Language Economy, in collaboration with the University of Alicante. We also welcomed the launch of the Horizon Europe ELLIOT project, dedicated to the development of European multimodal frontier models, as well as our collaboration with the University of Alicante in the framework of the ENIA Chair.

I invite you to read this report as a reflection of the collective effort made throughout 2025 and as an expression of the commitment with which we continue to move forward towards an Artificial Intelligence capable of making a real and positive contribution to social progress.



**NURIA
OLIVER**

ELLIS Alicante Director

It is my pleasure to present the 2025 Annual Report of ELLIS Alicante, a year in which we strengthened our team, expanded our research lines, and consolidated both our scientific presence and our impact on society.

We made significant progress in areas such as privacy and security in AI systems, the detection and mitigation of algorithmic bias, the relationship between human bias and AI, AI-based beauty filters, human-AI interaction, and the opportunities offered by chatbots to foster critical thinking in education. Our scientific output grew both in volume and impact, and we strengthened ties with leading institutions in Europe and at national and local level. In addition, this year we celebrated a particularly important milestone with the successful graduation of the first four students from the ELLIS doctoral program affiliated with ELLIS Alicante, two of whom have already joined us as postdoctoral researchers, reflecting the strength and potential of the talent developed within our environment. Congratulations to them all.

In 2025, our location at the privileged Muelle 5 within the Distrito Digital 5 was a key factor in attracting visiting researchers and in fostering new national and international collaborations. This setting supported a year of open dialogue with society, during which we promoted initiatives to bring Artificial Intelligence closer to the public, encourage scientific vocations from an early age, and contribute to public debate on the ethical development of technology.

I would like to express my deepest gratitude to the ELLIS Alicante team, whose talent, commitment and dedication have been essential in making possible the achievements reflected in this report. I would also like to extend this recognition to the members of our Board of Trustees, whose strategic vision and continued support are fundamental in guiding and strengthening our work. We are deeply grateful to our sponsors and collaborators for their continued trust, and especially to the Government of the Generalitat Valenciana, whose financial and strategic support makes it possible to advance research of excellence with a real impact on society. The combination of all these contributions forms the essential foundation of our mission and of everything we have built.

I invite you to read this report, which sets out in detail the work carried out throughout the year: the progress of our research projects, our scientific outreach initiatives, the strengthening of collaborations and strategic partnerships, as well as the awards and recognition we have received.

In 2025, in a context marked by the use of AI in conflicts and the spread of disinformation, we continued to work with the conviction that a more fairer, human-centric and socially oriented Artificial Intelligence is possible.

Milestones 2025

29

scientific publications

11

awards and recognitions

4

PhD theses

1

dataset

2

European Projects
H2020 (ELIAS & ELLIOT)

1

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

2

Nacional project PERTE de la Economía de
la Lengua: Cátedra ENIA



Milestones 2025

43

new collaboration agreements

6

new institutional/corporate donors



Milestones 2025

45

guest presentations

4

HCML reading group sessions

5

conferences and workshops
organized

Milestones 2025

11

ELLIS PhD students

4

ELLIS postdoctoral
researchers

7

visiting researchers

21

associated researchers



Number ranking on Google for keywords including:

Beauty Filters Bias

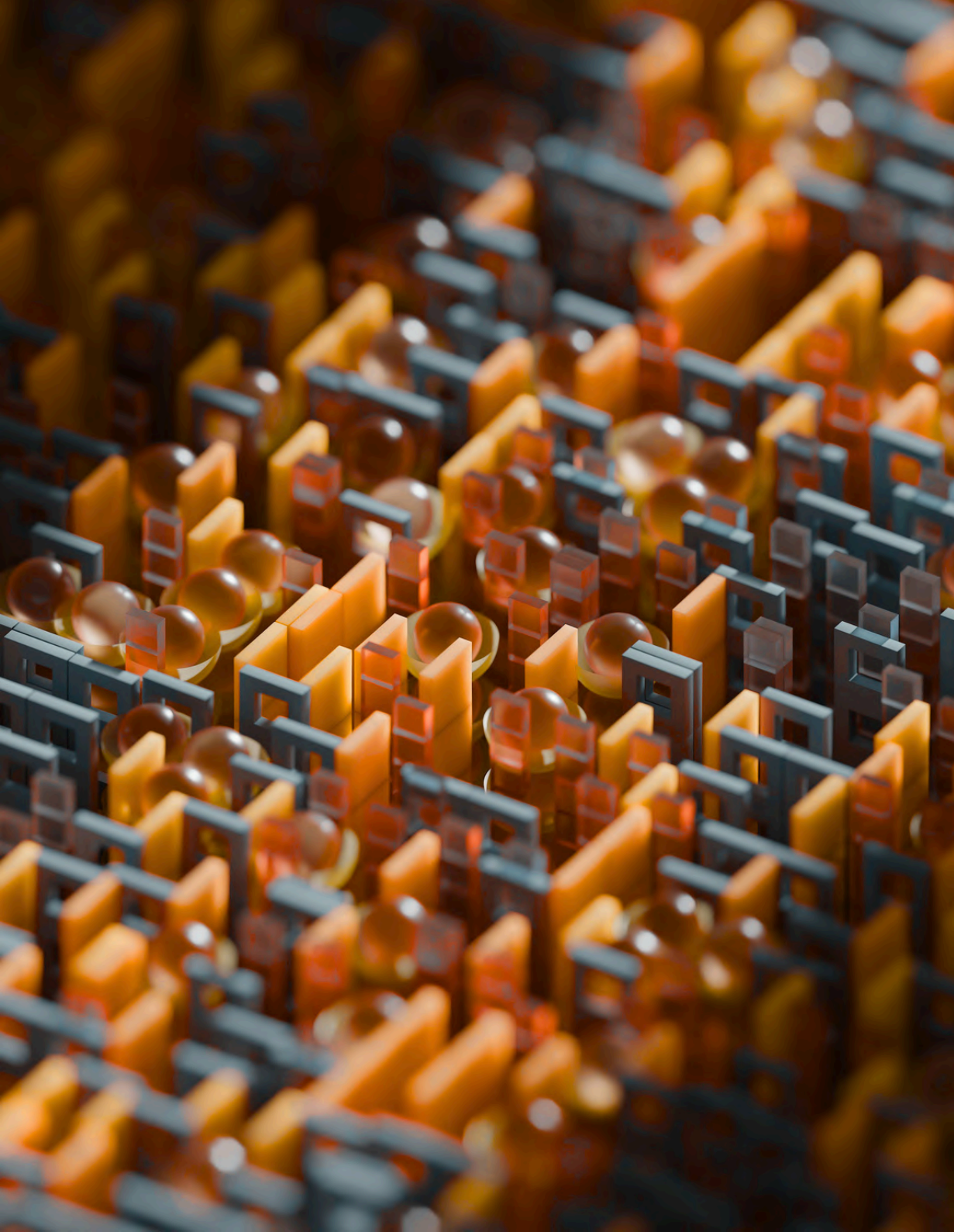


IA para el Bien Social



Art Censorship Social Media





Milestones 2025

2

articles written for INFORMACIÓN

+ 130

media appearances

Awards and recognitions

ELLIS ALICANTE RECEIVES THE 2025 SAN ALBERTO MAGNO AWARD

The Foundation receives the award presented by the Faculty of Sciences at the University of Alicante. This award is granted annually to companies, individuals or institutions that have made an outstanding contribution through their collaboration with the Faculty of Sciences' academic programs and activities.

DR OLIVER RECEIVES THE 2025 TALGO AWARD FOR WOMEN'S EXCELLENCE IN ENGINEERING

This is Talgo's highest accolade for women engineers with outstanding academic and professional careers.

DR OLIVER RECEIVES THE 2025 NATIONAL MATHEMATICS AWARD

Nuria Oliver received the 'Julio Rey Pastor' National Research Award in Mathematics and ICT in recognition of her outstanding career and innovative contributions to Artificial Intelligence, acknowledged for their scientific and social impact, as well as her commitment to ethical, human-centred AI.

ADRIÁN ARNAIZ RODRÍGUEZ WINS THE BEST PAPER AND BEST POSTER AWARDS AT THE MLG WORKSHOP AT ECML-PKDD 2025

Adrián Arnaiz Rodríguez receives the Best Paper and Best Poster awards at the Workshop on Mining and Learning with Graphs (MLG), held as part of the international ECML-PKDD 2025 conference, in recognition of the quality and impact of his research.

ELLIS ALICANTE WINS THE CRITICAL QUESTIONS GENERATION SHARED TASK AT ACL 2025

The ELLIS Alicante team, led by Lucile Favero, wins first prize in an international competition held as part of ACL 2025 for developing AI methods capable of generating critical questions about argumentative texts, thereby promoting critical thinking, rigorous analysis and the responsible use of language models.

DR OLIVER IS NAMED SOCIAL FIGURE OF THE YEAR 2025

Haz magazine named Nuria Oliver its 2025 Social Figure of the Year in recognition of her scientific career and her leadership in advancing ethical, fair and human-centred Artificial Intelligence with social impact and international reach.



ELLIS Europe

ELLIS is a pan-European AI network of excellence. It builds upon machine learning as the driver for modern AI and aims to secure lasting international leadership of AI made in Europe by connecting top researchers in this field and by creating a multi-centric AI research laboratory.

Founded in 2018, ELLIS has quickly grown into a network composed of 2 Institutes, 41 Units, 1 Associate Unit at

world-class institutions in 17 countries, 16 Research Programs and a pan-European PhD & Postdoc program.

The members of ELLIS are committed to shaping the future of AI in Europe by pushing the scientific and technological boundaries for human-centered, beneficial and safe AI.

2000

Members

44

ELLIS Units, ELLIS
Institutes y Associate
ELLIS Unit

560

Fellows & Scholars

16

Cross-Border Research
Programs

100

Associated Institutions

500

PhD Students & Postdocs
ELLIS



e l l i s

17

Countries

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.

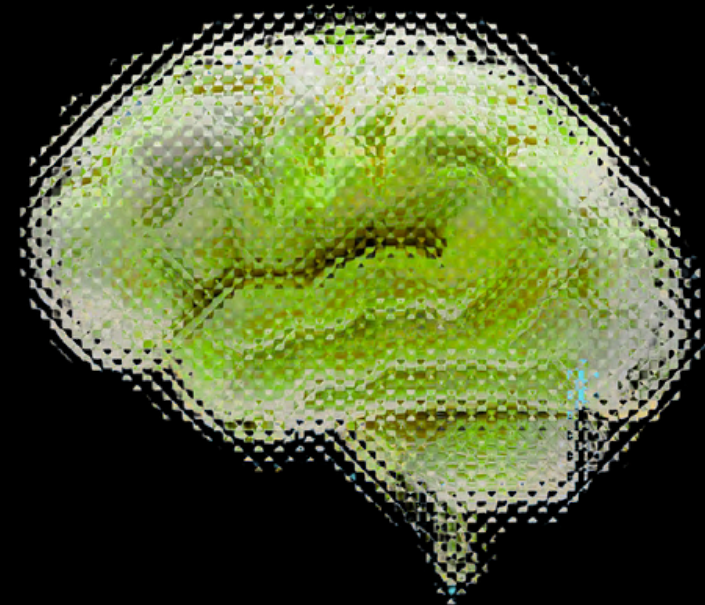


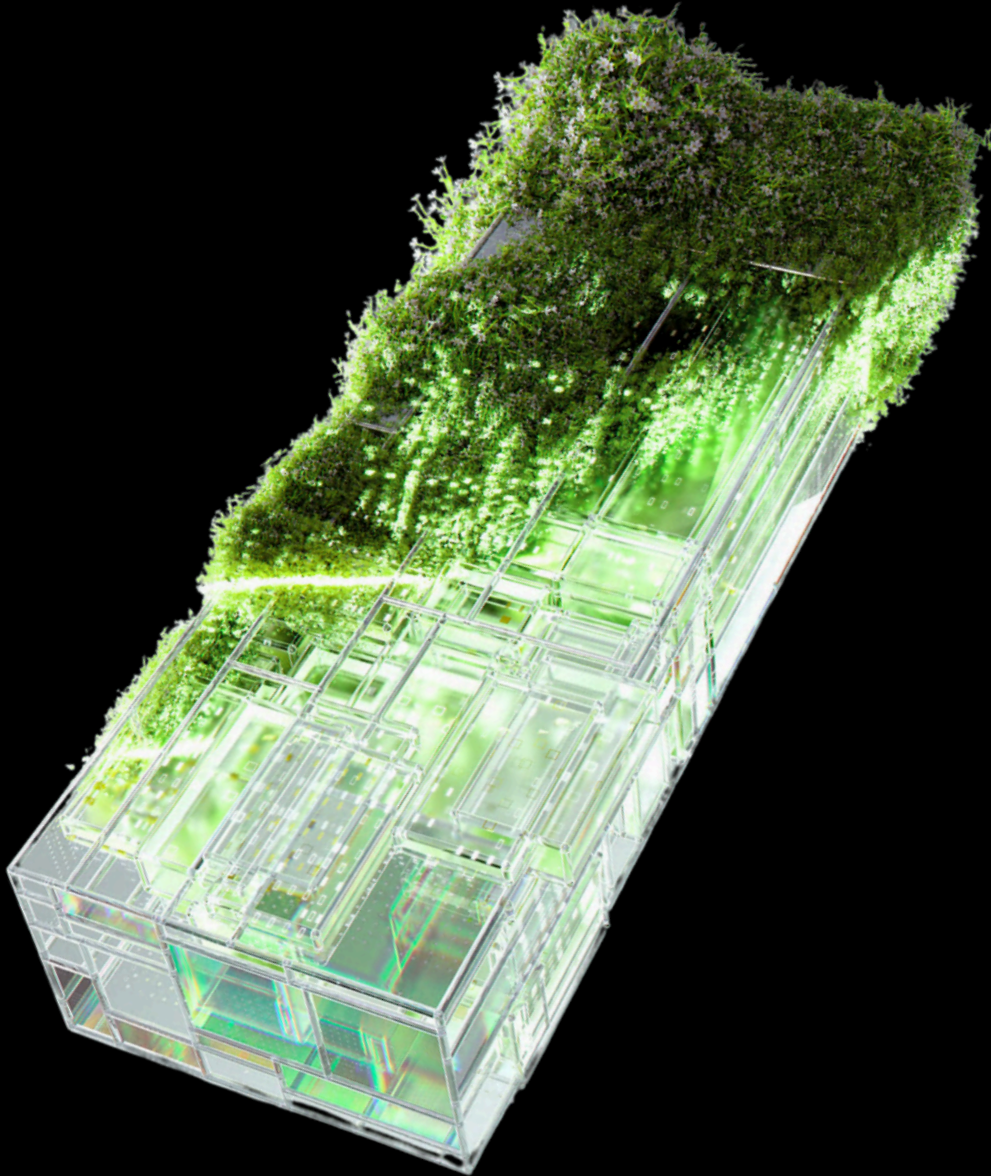
An AI revolution by and for the people

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 program.

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 element to foster creativity and motivation.

Goals

1°

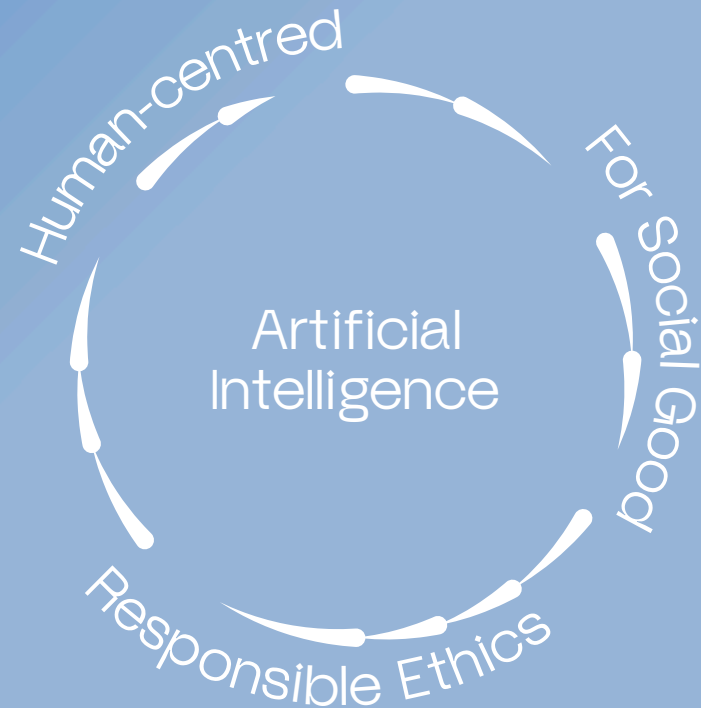
Conduct excellent scientific research in human-centric, responsible, and socially beneficial Artificial Intelligence.

2°

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.

3°

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.



Beneficiaries of our activity

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

The foundation's activities must benefit society as a whole.



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

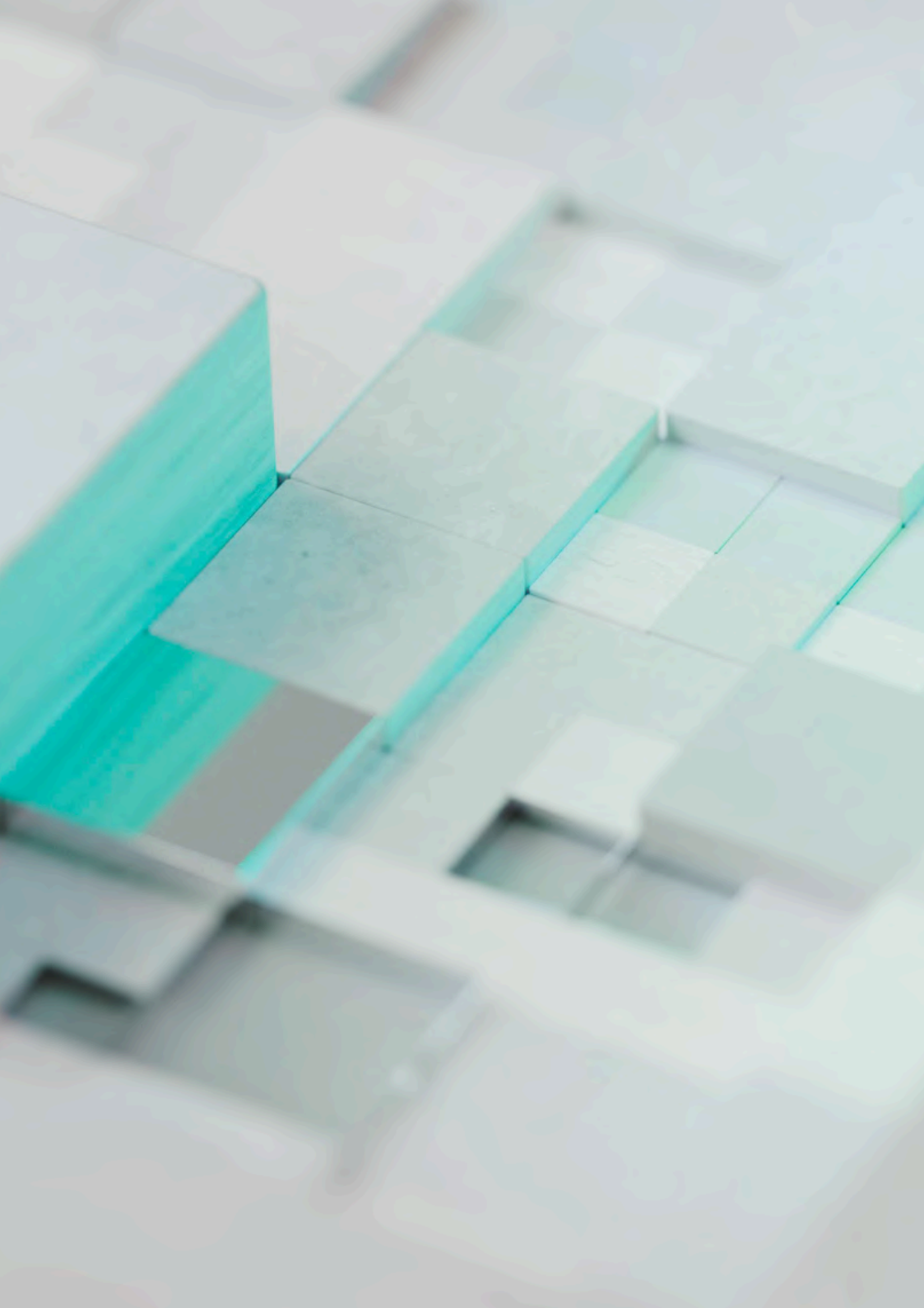
Sara Gómez Martín

Trustee

Elías Fereres Castiel

Trustee

María Victoria Carrascosa
Chagues



Scientific advisory board

Composed of scientists of recognised international 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 López de Mántaras

Massachusetts Institute of Technology (MIT)

Sandy Pentland

Team

Research Team

Internationally renowned senior, junior and pre-doctoral researchers. The Director is an ELLIS Fellow, as well as a Fellow of the ACM, IEEE, ELLIS, AAIA and EurAI. Our pre-doctoral and postdoctoral researchers are part of the highly competitive ELLIS PhD and Postdoctoral Programs. At ELLIS Alicante, we conduct open research in collaboration with researchers from universities and other research institutions.



**NURIA
OLIVER**

ELLIS Alicante Director

Scientific Director and
Co-Founder of ELLIS Alicante

Postdoctoral Researchers



ERIK DERNER

ELLIS Member

January - July 2025



ADITYA GULATI

ELLIS Member

December 2025 -



ADRIÁN ARNAIZ

ELLIS Member

September 2025 -

Associated Researchers



**ANNA
KORHONEN**

University of
Cambridge, United
Kingdom

ELLIS Fellow



**BRUNO
LEPRI**

FBK, Italy

Member of the
Scientific Advisory
Committee

ELLIS Fellow



**CRISTINA
CACHERO**

*Universidad de
Alicante, Spain*



**DALIBOR
KUČERA**

*University of South
Bohemia, Czech
Republic*



**EMMANUEL
LETOUZE**

*Data-Pop Alliance
& UPF, Spain*



**ERIK
DERNER**

*CTU Prague, Czech
Republic*

ELLIS Member



**FLORA
SALIM**

*UNSW Sydney,
Australia*



**GEORGINA
CURTO**

*Notre Dame
University, United
States of America*



**JUAN ANTONIO
PEREZ**

*Universidad de
Alicante, Spain*



**KAISA
VAANÄNEN**

*Tampere University,
Finland*

ELLIS Member



**KRISTINA
DERNER
BATISTIČ**

*Independent
Researcher,
Slovenia*



**MANUEL
GOMEZ
RODRIGUEZ**

MPI-SWS, Germany

ELLIS Fellow



**MARINA
MARTINEZ**

*Universitat Jaume I,
Spain*



**MIGUEL ANGEL
LOZANO**

*Universidad de
Alicante, Spain*



**NOA
GARCIA**

*University of Osaka,
Japan*



**NOVI
QUADRIANTO**

*UoSussex, United
Kingdom*

ELLIS Scholar



**SHIRLEY
OGOLLA**

HIIG, Germany



**TANJA
KASER**

EPFL, Switzerland

ELLIS Member



**THILO
HAGENDORFF**

*University of
Stuttgart, Germany*



**THOMAS
HOFFMANN**

*ETH Zurich,
Switzerland*

ELLIS Fellow



**THOMAS
SERRE**

*Brown University,
United States of
America*

ELLIS Fellow

PhD Students



**ADITYA
GULATI**

ELLIS Alicante

January-
December 2025



**JULIEN
COLIN**

ELLIS Alicante



**ADRIÁN ARNAIZ-
RODRÍGUEZ**

ELLIS Alicante

January-
September 2025



**KIM
ZIERAHN**

ELLIS Alicante

From November
2025



**DIEGO MIGUEL
LOZANO**

ELLIS Alicante

From November
2025



**LUCILE
FAVERO**

ELLIS Alicante



**GERGELY D.
NEMETH**

ELLIS Alicante

January-
September 2025



**PIERA
RICCIO**

ELLIS Alicante

January-
September 2025

PhD Students



MENGTING WEI

University of Oulu, Finland

Secondary ELLIS institution:
ELLIS Alicante



BENEDIKT HÖLTGEN

University of Tübingen, Germany

Secondary ELLIS institution:
ELLIS Alicante

Visiting Scientists



AXEL ABELS

Université Libre de Bruxelles, Belgium



AYOUB BAGHERI

Utrecht University, The Netherlands



MIRIAM DOH

Université de Mons (UMONS), Belgium



NOA GARCIA

University of Osaka, Japan



THILO HAGENDORFF

University of Stuttgart, Germany



JOSÉ LUIS HERNÁNDEZ-RAMOS

Universidad de Murcia, Spain



TOM LENAERTS

Université Libre de Bruxelles, Belgium



RAMON RUIZ-DOLZ

Notre Dame University, Estados Unidos



APURVA SHAH

Université Libre de Bruxelles, Belgium

Administration and Support Team

Composed of professionals with recognized experience, expertise, and a track record in the administrative support of research.



Rebeca
de Miguel
Head of Operations

Cristina
González

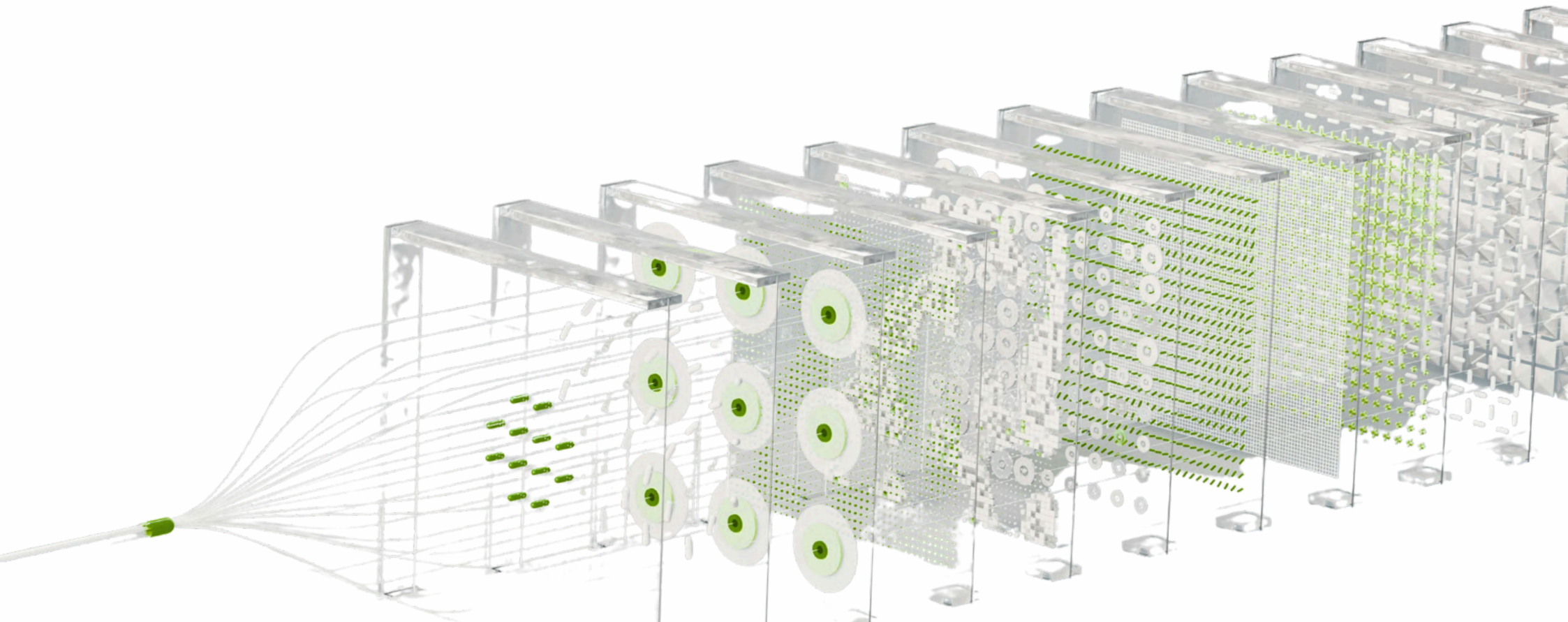
January - September 2025

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

COMPUTATIONAL MODELING OF HUMAN BEHAVIOR

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.



Project - "BIASED"

ARTIFICIAL INTELLIGENCE AND COGNITIVE BIASES

This research project forms part of the doctoral thesis of ELLIS PhD student **Aditya Gulati**, who holds a Master's degree in Computer Science from IIIT Bangalore, India, and has a strong interest in the computational modeling of human behavior. In addition to his work in India, he has collaborated with Professor Neumann in Ulm, Germany, and Professor González at CMU.

Aditya is a member of the prestigious ELLIS PhD program. His thesis is co-supervised by Dr Nuria Oliver (lead supervisor); Dr Bruno Lepri (co-supervisor), a researcher at FBK and an ELLIS Fellow in the ELLIS Human-Centric Machine Learning research program; and Dr Miguel Ángel Lozano, professor and academic tutor at the University of Alicante.

The overarching aim of this line of research is to study human cognitive biases and their impact on Artificial Intelligence systems, in order to understand how these biases are reflected, amplified or transformed by such systems. Building on this knowledge, the project seeks to harness the potential of Artificial Intelligence to identify, analyze and ultimately mitigate these



biases, as well as to design more reliable and transparent intelligent systems and interfaces. intelligence to identify, analyze and ultimately mitigate these biases, as well as to design more reliable and transparent intelligent systems and interfaces.

More specifically, the doctoral thesis examines the existence and manifestation of the so-called "attractiveness halo effect" both in humans and in generative image and multimodal AI systems.

During 2025, the work focused primarily on two studies. The first examined the impact of attractiveness on the behavior of multimodal large language models (MLLMs), which are capable of analyzing both images and text, when making decisions or forming judgements about people.

Scientific Publications:

Doh, M., **Gulati, A.**, & **Oliver, N.** (2025). Attractive by Design: How The Attractiveness Halo Effect Shapes AI Perception. *Collaborative AI and modeling of Humans (CAIHu) - Bridge program at AAAI 2025. Philadelphia, US, February 25th 2025*

Doh, M., **Gulati, A.**, Mancas, M., & **Oliver, N.** (2025). When Algorithms Play Favorites: Lookism in the Generation and Perception of Faces. *European Workshop on Algorithmic Fairness. Eindhoven, The Netherlands. June 30th 2025.*

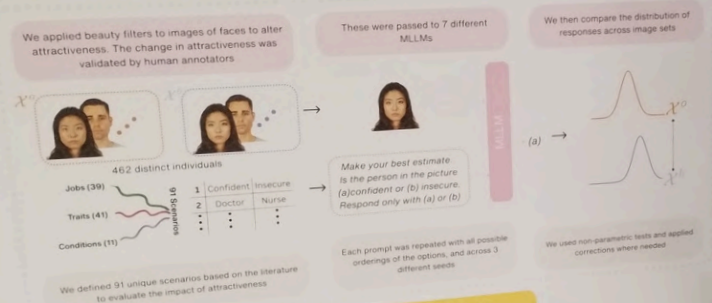
Gulati, A., D'Incà, M., Sebe, N., Lepri, B., & **Oliver, N.** (2025). Beauty and the Bias: Exploring the Impact of Attractiveness on Multimodal Large Language Models. *Eighth AAAI/ACM Conference on AI, Ethics and Society (AIES). Madrid, Spain, October 20th 2025.*

Secondly, Aditya investigated how synthetically generated faces and machine-learning-based gender classification algorithms are affected by attractiveness bias. In experiments involving 12,000 synthetically generated faces, he found that: (1) text-to-image (T2I) systems tend to associate facial attractiveness with unrelated positive traits, such as intelligence and trustworthiness; and (2) gender classification models exhibit higher error rates for "less attractive" faces, particularly among non-white women. These findings raise concerns about fairness in digital identity systems.

Aditya successfully defended his PhD thesis at the University of Alicante in December 2025.

Beauty and the Bias Exploring the Impact of Attractiveness on Multimodal Large Language Models

The **attractiveness halo effect** is the tendency for humans to associate positive traits -- such as intelligence or trustworthiness -- with people they consider attractive, despite the lack of an objective correlation. How are **MLLMs** impacted by **facial attractiveness** when making decisions?



Attractiveness Biases

		Jobs (39)		Traits (41)		Conditions (81)	
		Total (91)	Gender (12)	Other (31)	Other (31)	Geography (4)	Other (2)
			Attractiveness (8)	Attractiveness (8)	Attractiveness (8)	Attractiveness (8)	Attractiveness (8)
Gemini	89.0%	78.9%	91.7%	100.0%	92.9%	75.0%	100.0%
Phi-3.5	78.1%	84.2%	83.3%	75.0%	100.0%	60.0%	50.0%
DeepSeek	90.1%	84.2%	91.7%	100.0%	93.9%	50.0%	60.0%
Mistral	80.2%	68.4%	66.7%	100.0%	90.9%	62.5%	100.0%
Qwen2	81.2%	68.4%	66.7%	100.0%	87.9%	62.5%	100.0%
Pixtral	86.8%	62.2%	75.0%	75.0%	97.0%	75.0%	50.0%
LLaVA 1.5	80.2%	75.7%	79.8%	91.1%	92.6%	71.4%	85.7%
Average	83.8%	75.7%	79.8%	91.1%	92.6%	71.4%	85.7%

Demographic Biases

		Bias	Race
		Gender	Age
Model			
Gemini	228.0***	53.8%	1.0%
Phi-3.5	125.0***	53.8%	1.0%
DeepSeek	183.0***	53.8%	1.0%
Mistral	124.0***	53.8%	1.0%
Qwen2	124.0***	53.8%	1.0%
Pixtral	175.0***	53.8%	1.0%
LLaVA 1.5	124.0***	53.8%	1.0%

Interaction Between Biases

		Model	Halo Effect	Halo Effect	Halo Effect
			Attractiveness	Geography	Other
Gemini	228.0***	1875.0***	1875.0***	1875.0***	1875.0***
Phi-3.5	125.0***	1875.0***	1875.0***	1875.0***	1875.0***
DeepSeek	183.0***	1875.0***	1875.0***	1875.0***	1875.0***
Mistral	124.0***	1875.0***	1875.0***	1875.0***	1875.0***
Qwen2	124.0***	1875.0***	1875.0***	1875.0***	1875.0***
Pixtral	175.0***	1875.0***	1875.0***	1875.0***	1875.0***
LLaVA 1.5	124.0***	1875.0***	1875.0***	1875.0***	1875.0***

Project - "PSYCHAI"

THE INTERSECTION OF PSYCHOLOGY AND AI

This research project is part of the doctoral thesis of **Kim Zierhan**, who joined the ELLIS Alicante team of predoctoral researchers within the ELLIS PhD program in November 2025

Kim holds a Master's degree in Quantitative Methods in Data Science from the University of Tübingen. Her Master's thesis explored the prediction of the incidence of psychological disorders using large language models, comparing their performance with that of machine-learning models. She also holds a Bachelor's degree in Psychology from the University of Münster. Until March 2024, she worked as a research assistant at the Methods Center of the University of Tübingen, at the intersection of data science, machine learning and psychology.

Her doctoral thesis is jointly supervised by Dr Nuria Oliver, as principal supervisor; Dr Anna Korhonen, Professor at the University of Cambridge and an ELLIS Fellow in the ELLIS Natural Language Processing program, as secondary supervisor; and Dr Cristina Cachero, Professor at the University of Alicante, who also serves as her academic tutor.

The project focuses on the intersection of psychology and AI. One area of interest is the modeling of psychological characteristics, such as personality, using AI models, including large language models, as well as the personalisation of human-AI interaction. Another area of interest is the intersection of mental health and AI, together with the psychological implications of dependence on, or the constant use of, chatbots.

As the project was launched only recently, its findings had not yet resulted in any scientific publications by the end of 2025. During the year, the necessary preliminary and foundational research was carried out to support the achievement of these results.



Program 2

ARTIFICIAL INTELLIGENCE THAT WE TRUST

INTERACTIVE INTELLIGENT
SYSTEMS AND THE STUDY OF
HUMAN-AI INTERACTION

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-AI interaction.



Project - "AIDEAS"

CHATBOTS AND HUMAN LEARNING

This research project is being conducted as part of the doctoral thesis of **Lucile Favero**, who joined the ELLIS Alicante team of predoctoral researchers as an ELLIS PhD student 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. Her Master's thesis, supervised by Dr Sylvain Sardy, involved the development of a machine-learning model to optimise heat-pump systems. During her neuroscience studies, she also completed a three-year research placement at the EL-Boustani laboratory under the supervision of Dr Giulio

Matteucci. There, she applied a decision-making model with considerable potential to provide researchers with valuable insights for the design, analysis and improvement of behavioral tasks.

Her doctoral thesis is jointly supervised by Dr Nuria Oliver, as principal supervisor; Dr Tanja Käser, Professor at EPFL as secondary supervisor; and Dr Juan Antonio Pérez-Ortiz, Professor at the University of Alicante, who also serves as her academic tutor.

The project focuses on assessing the effectiveness of using a Socratic approach in the development of educational chatbots, with particular emphasis on fostering active learning, critical thinking and knowledge retention. Drawing on the existing literature on the Socratic method in education and the potential of chatbots as learning tools, this ambitious project aims to develop Maike, an innovative Socratic chatbot for education.

The project will also compare traditional teaching methods, standard chatbots and Maike. The outcome metrics will include depth of reasoning, knowledge retention, engagement with learning and the development of critical-thinking skills. The ultimate goal is to create a learning environment that is both more intellectually stimulating and more emotionally supportive. The findings are expected to provide valuable insights into the design of AI-based conversational agents for learning.



During 2025, the research explored the ability of small, open-weight large language models (LLMs) to perform argument-mining tasks, particularly the automatic segmentation and classification of argument components and the assessment of their quality.

Work was also carried out on the automatic generation of critical questions as a means of improving the quality of arguments. Lastly, the project examined the challenges posed by the indiscriminate use of LLMs and chatbots in education.

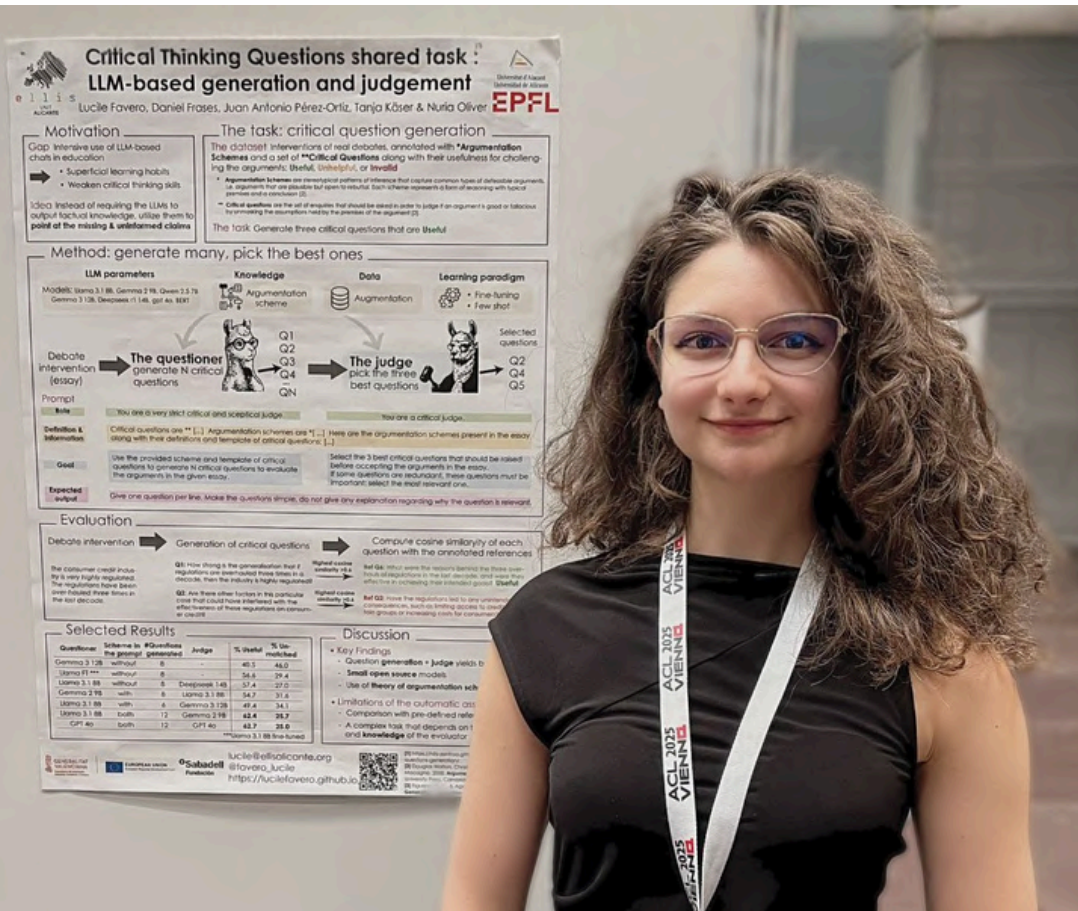
Scientific Publications:

Favero, L. A., Pérez-Ortiz, J. A., Käser, T., & **Oliver, N.** (2025). Leveraging Small LLMs for Argument Mining in Education: Argument Component Identification, Classification, and Assessment. February 20th 2025

Favero, L. A., Pérez-Ortiz, J. A., Käser, T., & **Oliver, N.** (2025). Argument Mining in Education: Exploring the Potential of Open-source Small LLMs for Argument Classification and Assessment. In *AAAI2025 AI for Education - Tools, Opportunities, and Risks in the Generative AI Era*. Philadelphia, USA, March 3rd de 2025

Favero, L. A., Pérez-Ortiz, J. A., Käser, T., & **Oliver, N.** (2025). Do AI tutors empower or enslave learners? Toward a critical use of AI in education. In *Applications of generative AI to support teaching and learning in higher education, co-located with AIED 2025*. Palermo, Italy July 22nd 2025

Favero, L. A., Frases, D., Pérez-Ortiz, J. A., Käser, T., & **Oliver, N.** (2025). ELLIS Alicante at CQs-Gen 2025: Winning the critical thinking questions shared task: LLM-based question generation and selection. In *12th Workshop on Argument Mining, co-located with ACL 2025*. Vienna, Austria. July 31st 2025



Program 3

ARTIFICIAL INTELLIGENCE THAT WE TRUST

CHALLENGES, LIMITATIONS AND ETHICAL AND SOCIAL IMPLICATIONS OF AI

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, PRIVACY AND BIAS

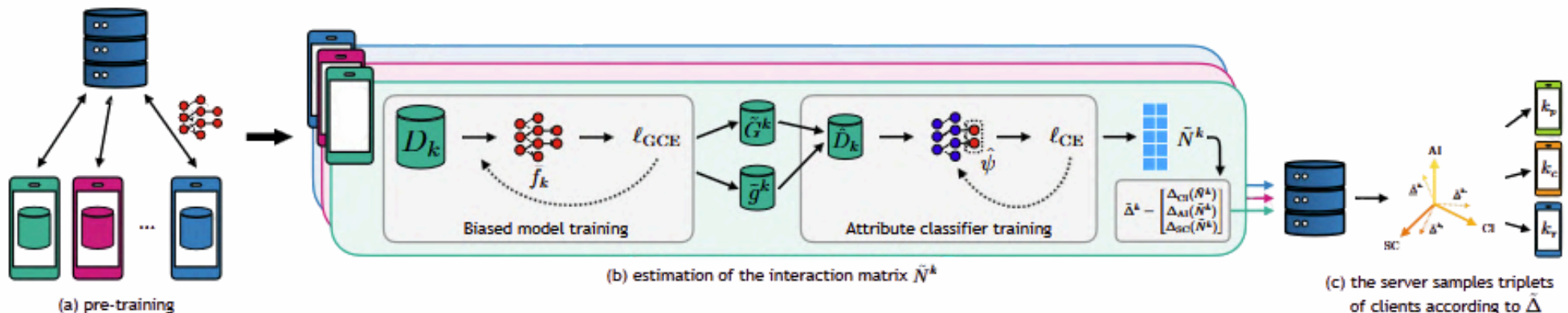
This research project formed part of the doctoral thesis of **Gergely Németh**, a doctoral student in the ELLIS PhD Program.

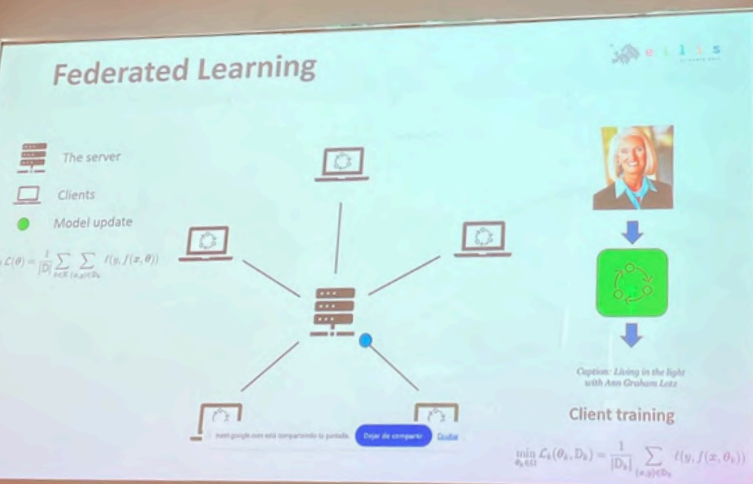
Gergely studied Computer Science at the Budapest University of Technology and Economics and the University of Manchester, where he worked on natural language processing (NLP). He subsequently contributed to applied machine-learning (ML) projects focusing on computer vision (CV) at Asura Technologies, a leading Hungarian start-up specialising in smart traffic applications. He holds a Master's degree in Computer Science and has professional experience in the application of AI in start-up environments. He also has extensive teaching experience.



His thesis was jointly supervised by Dr Nuria Oliver, as principal supervisor, Dr Novi Quadrianto, Professor at the University of Sussex and an ELLIS Scholar in the ELLIS Human-Centric Machine Learning program as secondary supervisor, and Dr Miguel Ángel Lozano, Professor at the University of Alicante, who also served as his academic tutor.

The project focused on federated learning (FL), a machine-learning approach that has emerged as a means of addressing privacy concerns associated with machine-learning applications.





In 2025, the research on privacy in heterogeneous federated learning was completed, and a new line of research was launched to study the impact of data heterogeneity on federated learning. The aim was to propose a new method capable of achieving strong model performance despite differences in the data held by individual clients.

As a preliminary step, a taxonomy was proposed to characterise data heterogeneity across clients. Lastly, the potential of the proposed method to improve algorithmic fairness was explored in scenarios involving protected attributes and biased data distributions across clients.

The project concluded in September 2025, following the successful defense of Gergely's doctoral thesis at the University of Alicante.

Scientific Publications:

Németh, G. D., Lozano, M. A., Quadrianto, N., & **Oliver, N.** (2025). Privacy and Accuracy Implications of Model Complexity and Integration in Heterogeneous Federated Learning. *IEEE Access*, 13, 40258-40274. February 27th 2025

Németh, G. D., Fani, E., Ng, Y. J., Caputo, B., Lozano, M. A., **Oliver, N.,** & Quadrianto, N. (2025). FedDiverse: Tackling Data Heterogeneity in Federated Learning with Diversity-Driven Client Selection. In *3rd IEEE International Conference on Federated Learning Technologies and Applications (FLTA25)*. October 14th 2025

Project - "SOCIALAI"

SOCIAL IMPLICATIONS OF ARTIFICIAL INTELLIGENCE ON SOCIAL MEDIA PLATFORMS AND BIAS IN GENERATIVE AI MODELS



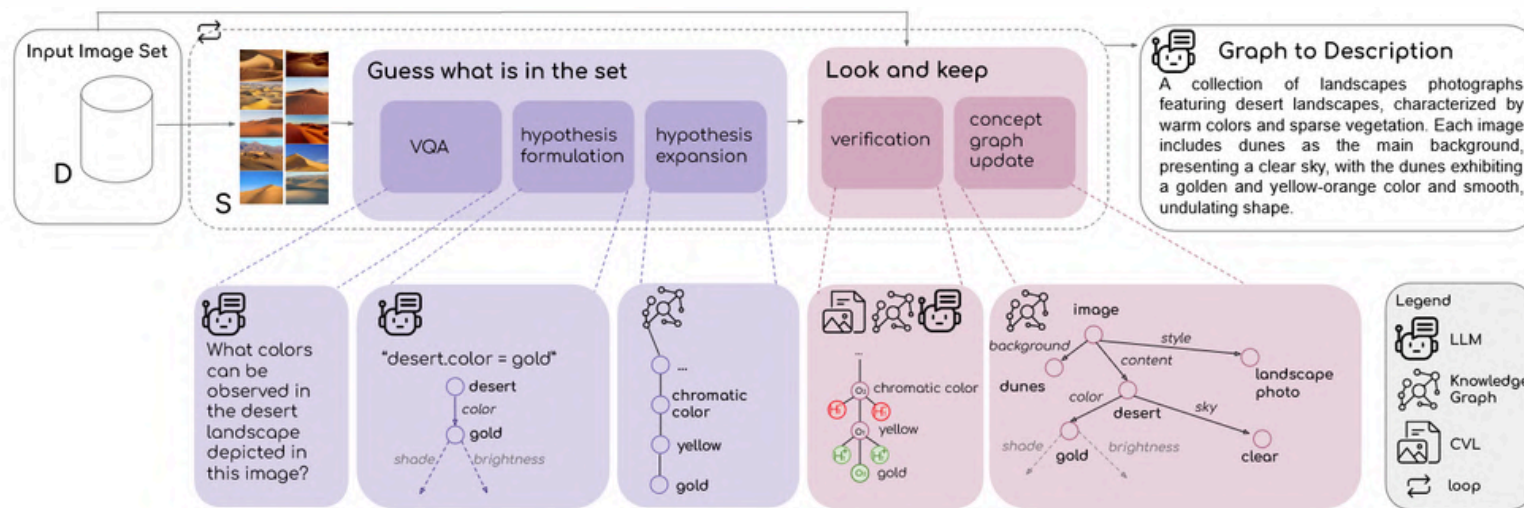
This research project formed part of the doctoral thesis of **Piera Riccio**, a doctoral student in the ELLIS PhD Program.

Piera holds a Bachelor's degree in Cinema and Media Engineering from the Politecnico di Torino (2018), a Master's degree in ICT for Smart Societies from the Politecnico di Torino (2021), and a Master's degree in Data Science and Engineering from Télécom Paris – EURECOM (2021). In 2020, she was affiliated with metaLAB (at) Harvard. In 2021, she worked as a research assistant at Oslo Metropolitan University.

Her thesis was jointly supervised by Dr Nuria Oliver, as principal supervisor; Dr Thomas Hofmann, Professor at ETH Zurich and an ELLIS Fellow in the ELLIS Human-Centric Machine Learning program as secondary supervisor; and Dr Miguel Ángel Lozano, Professor at the University of Alicante, who also served as her academic tutor.

The project focused on studying the societal implications of the AI algorithms embedded in social media platforms used by billions of people, as well as on developing new methods for describing collections of images, including both human-created artworks and images generated by AI algorithms.

In 2025, the project addressed the previously unresolved challenge of automatically generating textual descriptions of image collections, with the aim of characterising sets of images on the basis of their shared features.

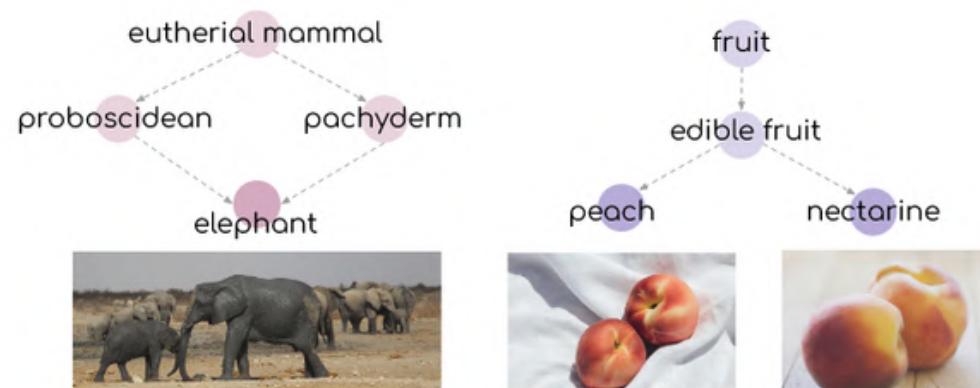


Piera also collaborated with visiting ELLIS doctoral researcher Ben Höltgen on a project examining the implications of using racial labels in relation to algorithmic discrimination, potential biases in genera-

tive AI models that generate human faces, and the effects of beauty filters.

The project concluded in September 2025, following the successful defense of Piera's doctoral thesis at the University of Alicante.

To this end, a novel method was proposed that combined the capabilities of large language models (LLMs) and multimodal large language models (MLLMs) with a knowledge representation. The method was validated on image collections of different types. This included a study of the differences identified by the method between human-created artworks and images generated by generative AI systems, which could help uncover biases or patterns in AI-generated images. This research was carried out in collaboration with Professor Noa García of Osaka University, where Piera completed a four-month research stay.





Scientific Publications:

Riccio, P., Galati, F., Schweighofer, K., Garcia, N., & **Oliver, N.** (2025). ImageSet2Text: Describing Sets of Images through Text. *arXiv preprint arXiv:2503.19361*. March 25th 2025.

Doh, M., Höltgen, B., **Riccio, P.**, & **Oliver, N.** (2025) Position: The Categorization of Race in ML is a Flawed Premise. In *ICML'25: International Conference on Machine Learning*. **Poster Spotlight** Vancouver, Canada, July 17th 2025.

Riccio, P., Domene Garcia, L., & **Oliver, N.** (2025) Towards Accessible Image Set Descriptions with ImageSet2Text. *Workshop on Vision Foundation Models and Generative AI for Accessibility: Challenges and Opportunities*. Honolulu, USA, August 22nd 2025.

Riccio, P (2025). Human aesthetics under the representational power of Artificial Intelligence. In PhD Thesis, University of Alicante. September 22nd 2025.

Project - "SOCIALAI"

BIAS IN GENERATIVE AI MODELS FOR HUMAN FACES

This research project is part of the doctoral thesis of **Mengting Wei**, a doctoral student in the ELLIS PhD Program. ELLIS Alicante served as her secondary institution for part of her international research stay, from 15 September to 15 December 2025.

Mengting holds a Bachelor's degree in Biomedical Engineering from Hebei University of Technology and a Master's degree from Southeast University, completed under the supervision of Professor Wenming Zheng. Her current research interests focus on the automatic generation and editing of human faces using Artificial Intelligence techniques.

Mengting is a doctoral student in the ELLIS PhD Program at the University of Oulu, where she works at the Center for Machine Vision and Signal Analysis (CMVS) under the supervision of Professor Guoying Zhao, an ELLIS Fellow in the ELLIS Human-Centric Machine Learning program. Dr Nuria Oliver is her secondary supervisor within the ELLIS PhD Program.



During her three-month stay at ELLIS Alicante, Mengting investigated biases relating to gender, age, race and perceived attractiveness in the faces generated by eight generative AI models, four developed in the United States and four in China. The study examined the generation of faces displaying a range of emotions: happiness, sadness, anger, surprise, fear and a neutral expression.

As her stay at ELLIS Alicante lasted only three months, the research had not yet resulted in any scientific publications by the end of 2025.

Project - "FAIRNESS"

FAIR AND TRANSPARENT MACHINE LEARNING MODELS

This research project formed part of the doctoral thesis of **Adrián Arnaiz Rodríguez**, a doctoral student in the ELLIS PhD Program.

Adrián holds a Bachelor's degree in Computer Engineering from the University of Burgos and a Master's degree in Artificial Intelligence from the Universitat Oberta de Catalunya. His doctoral thesis was jointly supervised by Dr Nuria Oliver, as principal supervisor; Dr Manuel Gómez, a researcher at a Max Planck Institute and an ELLIS Fellow in the ELLIS Human-Centric Machine Learning program as secondary supervisor; and Dr Miguel Ángel Lozano, Professor at the University of Alicante.



The project focuses on algorithmic fairness and non-discrimination from both technical and regulatory perspectives. Machine learning models are becoming key tools for addressing complex societal challenges. They are also increasingly being deployed to make or support decisions that affect people in many important areas of their lives, ranging from the justice system to healthcare. It is therefore essential to ensure that these algorithms do not discriminate, as well as to study effective forms of human-AI collaboration.

During 2025, Adrián completed his research into the potential of spectral graph theory to measure social capital within social networks and the isolation of different social groups, as well as to mitigate such isolation. He also completed his research into the implications of deep ensembles for algorithmic fairness.

In addition, he carried out research on human-AI collaboration, with the aim of ensuring that AI algorithms complement rather than replace humans. The goal was for human-AI systems to perform better than either humans or algorithms working independently. A scientific paper was also written on the use of discrete categories in machine-learning systems.

The project concluded in September 2025, following the successful defense of Adrián's doctoral thesis at the University of Alicante.

Scientific Publications:

Arnaiz-Rodríguez, A., Curto, G., & **Oliver, N.** (2025). Structural Group Unfairness: Measurement and Mitigation by means of the Effective Resistance. In *International AAAI Conference on Web and Social Media (ICWSM)*. Copenhagen, Denmark, July 7th 2025.

Schweighofer, K., **Arnaiz-Rodríguez, A.**, Hochreiter, S., & **Oliver, N.** (2025). The Disparate Benefits of Deep Ensembles. In *ICML'25: International Conference on Machine Learning*. Vancouver, Canada, July 18th 2025.

Arnaiz-Rodríguez, A., & Errica, F. (2025). Oversmoothing, Oversquashing, Heterophily, Long-Range, and more: Demystifying Common Beliefs in Graph Machine Learning. In *22nd International Workshop on Mining and Learning with Graphs at ECML-PKDD 2025*. **Premio al Mejor Artículo en MLG 2025 en ECML-PKDD 2025**. Porto, Portugal, September 15th 2025.

Arnaiz-Rodríguez, A., Corvelo Benz, N., Thejaswi, S., **Oliver, N.**, & Gomez Rodriguez, M. (2025). Towards Human-AI Complementarity in Matching Tasks. In *The Third Workshop on Hybrid Human-Machine Learning and Decision Making (HLDM'25) at ECML-PKDD 2025*. Porto, Portugal, September 15th 2025.

Arnaiz-Rodríguez, A. (2025). A Sociotechnical Approach to Trustworthy AI: from Algorithms to Regulation. In PhD Thesis, Uni. September 25th 2025.

Arnaiz-Rodríguez, A., Baidal, M., Derner, E., Layton Annable, J., Ball, M., Ince, M., Perez Vallejos, E., & Oliver, N. (2025). Between Help and Harm: An Evaluation of Mental Health Crisis Handling by LLMs. In Under review. September 29th 2025.

Spatial Rewiring - Dynamic
Drew [Gutteridge et al 2023]

- Closer nodes should interact earlier in the architecture
- Rewire to modulate **not only if** nodes interact, **but also when**.
- Multi-hop rewiring that evolves during the layers.
 - Nodes interact from a certain depth (hop $l+1$ is only aggregated in layer l)
 - Nodes interact with Delay (nodes interact with previous states)

Delay for each hop to interact with the previous states of the nodes

Separate aggregation for each k hop neighborhood

$$a_{i,k}^{(l)} = \text{AGG}_k^{(l)} \left(\{h_j^{(l-r_k(k))} : j \in N_k^{(l)}(i)\} \right), 1 \leq k \leq l+1$$
$$h_i^{(l+1)} = \text{UP}_k^{(l)} \left(h_i^{(l)}, a_{i,1}^{(l)}, \dots, a_{i,l+1}^{(l)} \right)$$

Hop $l+1$ is only aggregated in layer l

(b) Diffusion

(c) Diffusion

of dynamic rewiring but rather than having one rewarded graph on its layer on its layer of the graph.



Project - "FAIRNESS"

FAIR MACHINE LEARNING PREDICTIONS IN SOCIETY

This research project formed part of the doctoral thesis of **Benedikt Höltgen**, a doctoral student in the ELLIS PhD Program for whom ELLIS Alicante served as a secondary institution, with Dr Nuria Oliver as secondary supervisor.



His research visit to ELLIS Alicante took place as part of the mandatory research stays required by the ELLIS PhD Program and ran from October 2024 to March 2025.

Ben holds a Bachelor's degree in Mathematics and Philosophy and a Master's degree in Logic and Philosophy of Science from LMU Munich. He also completed a Master's degree in Computer Science at the University of Oxford, where he subsequently continued to collaborate externally with the OATML research group.

In 2022, he began a PhD in Machine Learning at the University of Tübingen and joined both the IMPRS-IS and ELLIS PhD programs.

His research focuses on machine-learning-based algorithmic systems used to support high-stakes decisions affecting individuals. These technologies are increasingly deployed in areas such as predicting the risk of reoffending or assessing credit default risk, thereby influencing decisions relating to parole or the granting of loans.

In recent years, numerous cases have emerged in which such systems have perpetuated or even amplified existing inequalities and forms of structural discrimination.

As part of his doctoral research, Ben studies how these issues can be better understood and how methods can be developed to mitigate them. His thesis is supervised by Professor Bob Williamson of the University of Tübingen.

During his visit to ELLIS Alicante, Ben worked on two research projects. First, he collaborated with Piera Riccio and Miriam Doh on a paper examining the implications of using racial categories to develop AI models.

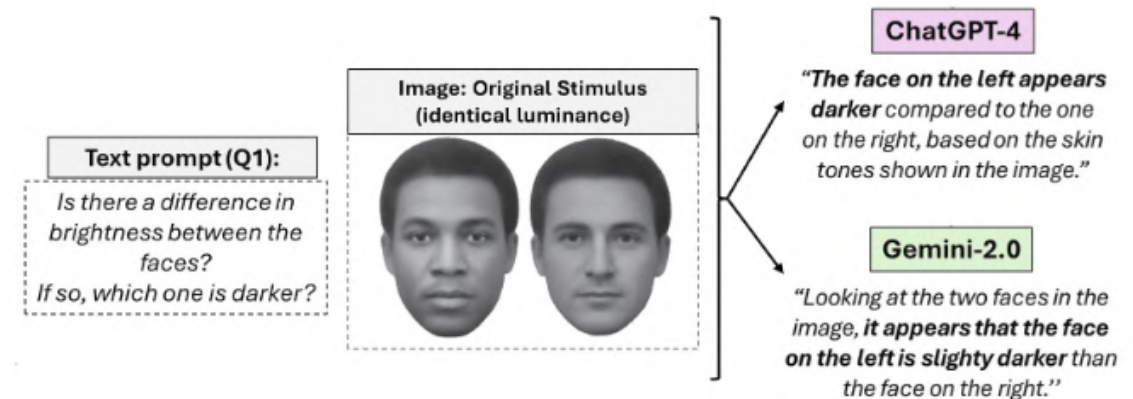


Scientific Publications:

Doh, M., **Höltgen, B.**, **Riccio, P.**, & **Oliver, N.** (2025). Position: The Categorization of Race in ML is a Flawed Premise. In *ICML'25: International Conference on Machine Learning*. Vancouver, Canada July 7th 2025.

Höltgen, B., & **Oliver, N.** (2025). Reconsidering Fairness Through Unawareness from the Perspective of Model Multiplicity. In *EAAMO 2025*. May 11th 2025

Second, he carried out novel research into the implications for algorithmic fairness of the widely discussed concept of fairness through unawareness (FTU). This approach is based on the idea that discrimination against demographic groups can be avoided by excluding group membership from decisions or predictions. The study demonstrated, both theoretically and empirically, that FTU can reduce algorithmic discrimination without necessarily compromising accuracy. The findings suggest that FTU should be considered in practical applications, particularly in high-stakes settings, and that the use of protected attributes such as gender in predictive models should be supported by a clear and well-founded justification.



Project - "XAI"

EXPLAINABLE ARTIFICIAL INTELLIGENCE

This research project is part of the doctoral thesis of **Julien Colin**, a doctoral student in the ELLIS PhD Program.

Julien holds a Bachelor's degree in Physics and Chemistry from the University of Lorraine, France (2019), and a Master's degree in Cognitive Science: Natural and Artificial Cognition from Grenoble INP (2021). Before beginning his doctoral studies, he worked as a research assistant, first at ANITI in Toulouse for six months in 2021 and subsequently at Brown University in Providence for five months in 2022.

His doctoral thesis is jointly supervised by Dr Nuria Oliver, as principal supervisor; Dr Thomas Serre, Professor at Brown University and an ELLIS Fellow in the ELLIS Natural Intelligence program as secondary supervisor; and Dr Miguel Ángel Lozano, Professor at the University of Alicante, who also serves as his academic tutor.

The project focuses on algorithmic transparency and explainability. Explainable Artificial Intelligence (XAI) seeks to develop methods that help people understand how AI models behave. Among the most widely used approaches are attribution methods.



Recent studies evaluating attribution methods in computer vision with human participants have shown that these methods are sufficiently effective at revealing where a model is looking. However, the field must now develop complementary explainability methods that can reveal what the model is seeing. Better benchmarks are also needed to evaluate explainability methods, as the benchmark that currently leads the field is insufficiently aligned with human understanding. The project draws inspiration from neuroscience to develop benchmarks that provide a clearer understanding of the practical value of explainability methods.

During 2025, Julien investigated two aspects of XAI. First, previous research had sought to understand the visual features learned by computer-vision models by analyzing individual neurons within their latent representation spaces. This had led to the view that advances in vision models had reduced their interpretability. Julien investigated a more recent hypothesis: that models should instead be analyzed after their representations have been projected onto a sparse distributed basis, rather than through the analysis of individual neurons.

Through user studies comparing the two approaches, Julien contributed the first empirical evidence that features within sparse distributed representations are significantly more interpretable than neuron-based features. The research also showed that the choice of basis has a substantial impact when comparing the interpretability of different models.

Second, Julien investigated how interpretability can be quantified through user studies, an area that remains an open challenge. In a comprehensive analysis of existing experimental protocols, he identified a key confounding factor in the standard protocol commonly used in the field. Once controlled for, this factor can reverse previous findings suggesting that newer architectures are less interpretable than older ones. To address these limitations, a new experimental design was proposed in

which the categorization task was replaced with an identification task, enabling fairer comparisons between models. The proposed paradigm was validated through additional control experiments.

In 2025, Julien also completed a six-month international research stay as part of the ELLIS PhD Program. The stay took place at Brown University in the United States under the supervision of Professor Thomas Serre, an ELLIS Fellow and co-supervisor of Julien's doctoral thesis.



Project - "VLMBE"

BIAS IN LARGE LANGUAGE MODELS

This research project was carried out by ELLIS postdoctoral researcher **Dr Erik Derner**.

Dr Derner obtained his PhD in Robotics and Artificial Intelligence from the Czech Technical University (CTU) in the Czech Republic in 2022. His doctoral thesis received the 2023 Werner von Siemens Award in the Industry 4.0 category, as well as the Dean's Award from the Faculty of Electrical Engineering at CTU.



He specialises in the societal and ethical implications of large language models (LLMs). Dr Derner was a researcher in the ELLIS Postdoctoral Program, with Professor Robert Babuška of CTU in Prague serving as his secondary postdoctoral supervisor.

The project focused on evaluating biases in the corpora used to train large language models, proposing mitigation methods, and studying human-LLM interaction from ethical and privacy perspectives. The aim of the research was to contribute to the development of safe, robust and user-friendly language models, with particular emphasis on low-resource languages such as Valencian, which is especially relevant to the Valencian Community.

During 2025, Dr Derner developed a novel method for measuring gender representation bias in the corpora used to train large language models. The method was applied to both high-resource languages, such as English, and low-resource languages, such as Valencian and Czech. He also studied the impact of gender representation bias on the text generated by large language models and explored potential mitigation strategies.

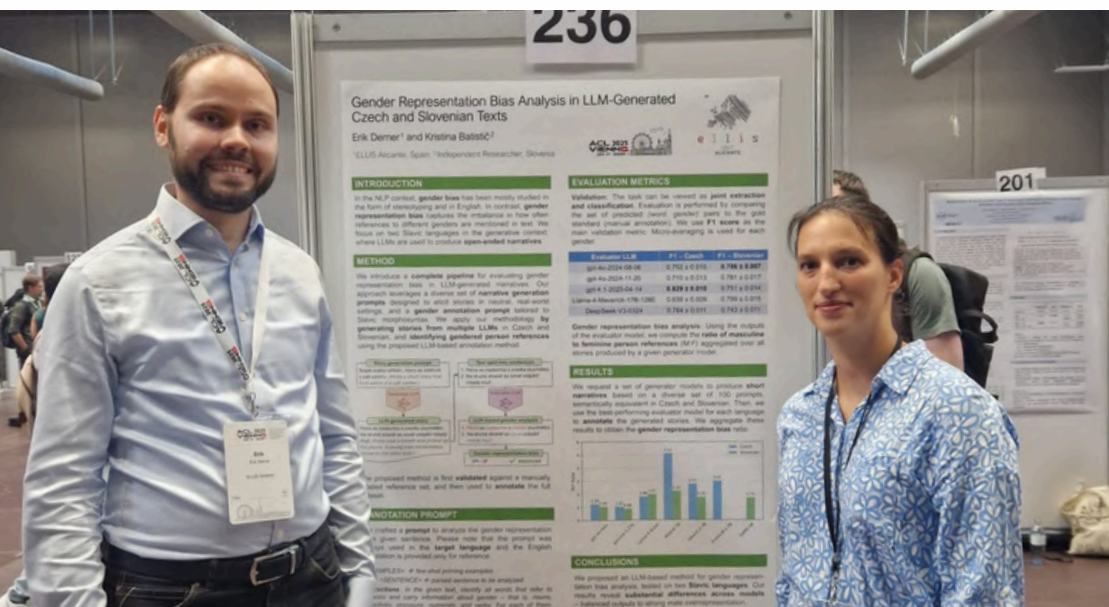
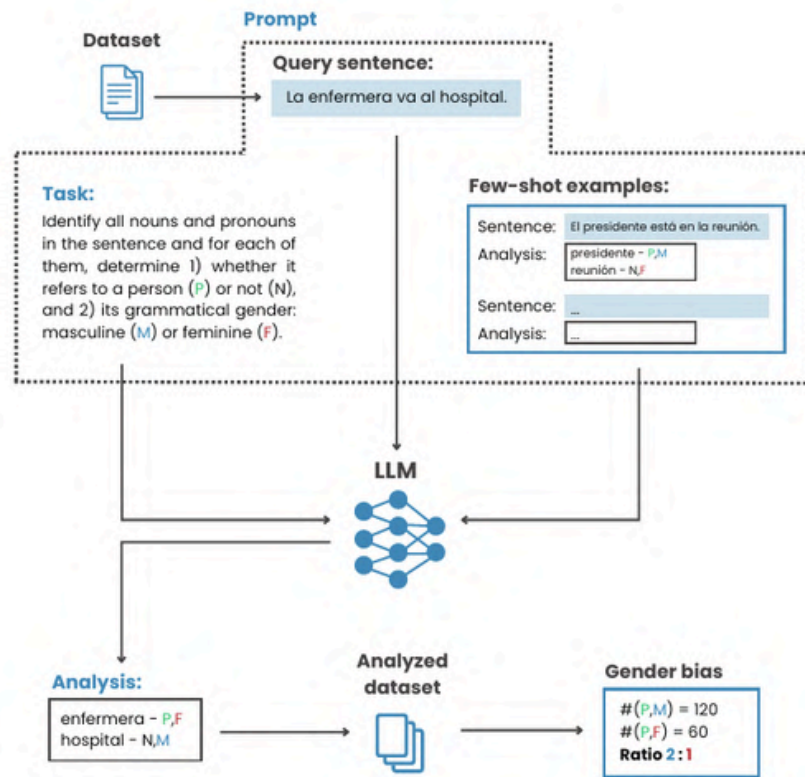
Dr Derner successfully completed his postdoctoral research period at ELLIS Alicante on 31 August 2025.

Scientific Publications:

Derner, E., Sansalvador de la Fuente, S., Gutiérrez, Y., Moreda, P., & **Oliver, N.** (2025). Leveraging Large Language Models to Measure Gender Representation Bias in Gendered Language Corpora. Collaborative AI and modeling of Humans (CAIHu) - Bridge program at AAAI 2025 Philadelphia, USA, February 25th 2025.

Derner, E., & Batistič, K. (2025). Gender Representation Bias Analysis in LLM-Generated Czech and Slovenian Texts. *10th Workshop on Slavic Natural Language Processing (Slavic NLP 2025)*, ACL 2025. Vienna, Austria. July 30th 2025.

Derner, E., Sansalvador de la Fuente, S., Gutiérrez, Y., Moreda, P., & **Oliver, N.** (2025). Leveraging Large Language Models to Measure Gender Representation Bias in Gendered Language Corpora. 6th Workshop on Gender Bias in Natural Language Processing (GeBNLP), ACL 2025. Vienna, Austria. August 1st 2025.



Project - "HCLM"

HUMAN-CENTRIC LANGUAGE MODELS AND LARGE LANGUAGE MODEL SAFETY

Dr Erik Derner's second research project focused on the human and ethical implications arising from the widespread use of large language models, such as ChatGPT, as well as the security vulnerabilities of these systems.

During 2025, Dr Derner investigated security risks associated with large language models, including their potential to be manipulated into revealing dangerous or privileged information. As part of this research, he carried out a pioneering project on vulnerabilities in multimodal large language models when used with low-resource languages, in collaboration with Kristina Batistič, an associate researcher at ELLIS Alicante.

He also conducted a user study examining the impact of communication style on human-chatbot interaction and launched a new line of research into the challenges and opportunities associated with the use of chatbots in mental-health contexts. This latter project was carried out in collaboration with Miguel Baidal, a university student of a MSc

program in Health Data Science, who completed his curricular internship at ELLIS Alicante. The project illustrates ELLIS Alicante's ability to attract outstanding young talent to research and deliver impactful results.

Scientific Publications:

Baidal, M., Derner, E., & Oliver, N. (2025). Guardians of Trust: Risks and Opportunities for LLMs in Mental Health. Fourth Workshop on NLP for Positive Impact (NLP4PI), ACL 2025. Vienna, Austria. July 31st 2025.

Derner, E., & Batistič, K. (2025). Beyond Words: Multilingual and Multimodal Red Teaming of MLLMs. First Workshop on LLM Security (LLMSEC), ACL 2025. Vienna, Austria. August 1st 2025.

Project - "SAFEAI"

AI SAFETY

This research project is being conducted as part of the doctoral thesis of **Diego Miguel**, a doctoral student in the ELLIS PhD Program.

Diego holds a Bachelor's degree in Computer Engineering from the University of Burgos and a Master's degree in Computer Science from the Technical University of Munich (TUM). He specialises in natural language processing (NLP). After several years working in industry, he began his doctoral studies at ELLIS Alicante in November 2025.

Throughout his academic career, Diego has received several awards, including the University of Burgos Award for Outstanding Academic Achievement. He also won first place at hackaTUM 2025, one of Europe's largest hackathons.

His doctoral thesis is jointly supervised by Dr Nuria Oliver, as principal supervisor; Dr Erik Derner, ELLIS Member, currently affiliated with the Czech Technical University (CTU), as secondary supervisor; and Dr Juan Antonio Pérez, Professor at the University of Alicante, who also serves as his academic tutor.

The project focuses on the safety, reliability and limitations of large language models (LLMs) and multimodal large language models (MLLMs).

As the project was launched in November 2025, the research had not yet resulted in any scientific publications by the end of the year. During this initial period, Diego carried out a review of the existing literature and the preliminary and foundational research required to support the next stages of the project.



Scientific Publications 2025

Favero, L. A., Pérez-Ortiz, J. A., Käser, T., & **Oliver, N.** (2025). Leveraging Small LLMs for Argument Mining in Education: Argument Component Identification, Classification, and Assessment. *In arXiv:2502.14389.*

Derner, E., Sansalvador de la Fuente, S., Gutiérrez, Y., Moreda, P., & **Oliver, N.** (2025). Leveraging Large Language Models to Measure Gender Representation Bias in Gendered Language Corpora. *In Collaborative AI and modeling of Humans (CAIHu) - Bridge program at AAAI 2025.*

Doh, M., **Gulati, A.,** & **Oliver, N.** (2025). Attractive by Design: How The Attractiveness Halo Effect Shapes AI Perception. *In Collaborative AI and modeling of Humans (CAIHu) - Bridge program at AAAI 2025.*

Németh, G. D., Lozano, M. A., Quadrianto, N., & **Oliver, N.** (2025). Privacy and Accuracy Implications of Model Complexity and Integration in Heterogeneous Federated Learning. *IEEE Access, 13, 40258-40274.*

Favero, L. A., Pérez-Ortiz, J. A., Käser, T., & **Oliver, N.** (2025). Argument Mining in Education: Exploring the Potential of Open-source Small LLMs for Argument Classification and Assessment. *In AAAI2025 AI for Education - Tools, Opportunities, and Risks in the Generative AI Era.*

Riccio, P., Galati, F., Schweighofer, K., Garcia, N., & **Oliver, N.** (2025). ImageSet2Text: Describing Sets of Images through Text. *arXiv preprint arXiv:2503.19361.*

Oliver, N. (2025, Marzo). Big Mobile Data for Social Good: Opportunities and Challenges. *In Human Development and the Data Revolution (Vol. 1, pp. 51-71).*

Lepri, B., **Oliver, N.,** & Pentland, S. (2025). Reinventing the Social Contract in an Age of Big Data. *Human Development and the Data Revolution, 1, 136-145.*

Doh, M., Canali, C., & **Oliver, N.** (2025). Filters of Identity: AR Beauty and the Algorithmic Politics of the Digital Body. In *CHI'25: International Conference on Human-computer Interaction. Workshop on Body Politics: Unpacking Tensions and Future Perspectives for Body-Centric Design Research in HCI.*

Arnaiz-Rodríguez, A., Curto, G., & **Oliver, N.** (2025). Structural Group Unfairness: Measurement and Mitigation by means of the Effective Resistance. In *International AAAI Conference on Web and Social Media (ICWSM).*

Doh, M., Canali, C., & **Oliver, N.** (2025). What TikTok Claims, what Bold Glamour Does: a Filter's Paradox. In *ACM Conference on Fairness, Accountability and Transparency, FAccT 2025.*

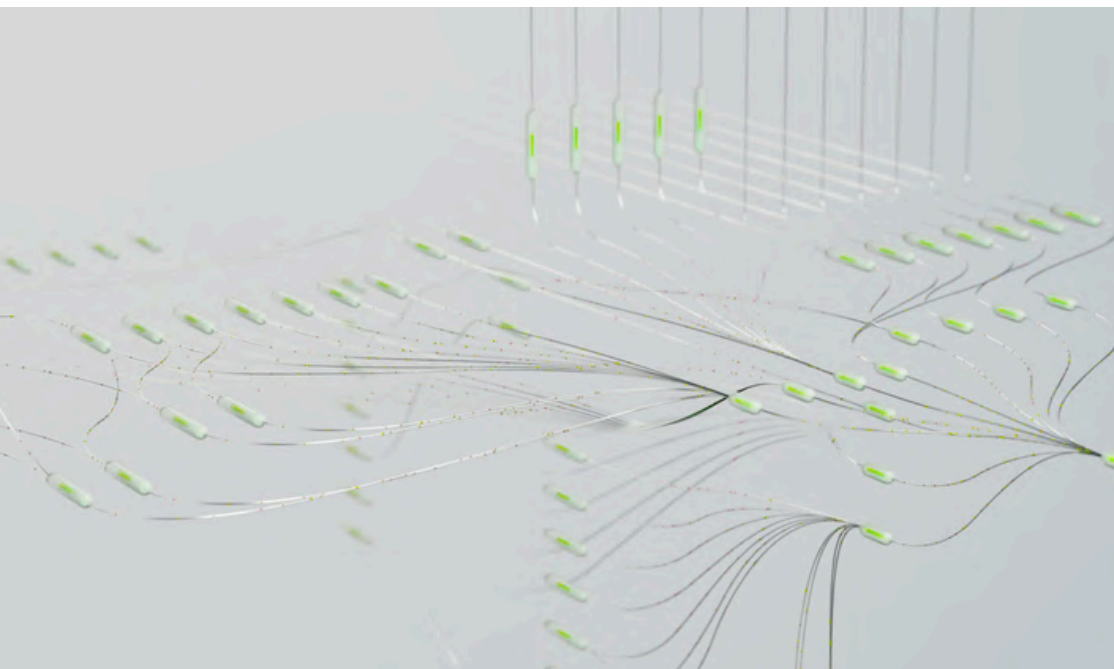
Doh, M., Canali, C., & **Oliver, N.** (2025). Filters of Identity: AR Beauty and the Algorithmic Politics of the Digital Body. In *CHI 2025, Workshop on Body Politics: Unpacking Tensions and Future Perspectives For Body-Centric Design Research in HCI.*

Doh, M., **Gulati, A.**, Mancas, M., & **Oliver, N.** (2025). When Algorithms Play Favorites: Lookism in the Generation and Perception of Faces. In *European Workshop on Algorithmic Fairness.*

Doh, M., **Höltgen, B.**, **Riccio, P.**, & **Oliver, N.** (2025). Position: The Categorization of Race in ML is a Flawed Premise. In *ICML'25: International Conference on Machine Learning.*

Schweighofer, K., **Arnaiz-Rodríguez, A.**, Hochreiter, S., & **Oliver, N.** (2025). The Disparate Benefits of Deep Ensembles. In *ICML'25: International Conference on Machine Learning.*

Favero, L. A., Pérez-Ortiz, J. A., Käser, T., & **Oliver, N.** (2025). Do AI tutors empower or enslave learners? Toward a critical use of AI in education. In *Applications of generative AI to support teaching and learning in higher education, co-located with AIED 2025.*



Baidal, M., **Derner, E.**, & **Oliver, N.** (2025). Guardians of Trust: Risks and Opportunities for LLMs in Mental Health. In Fourth Workshop on NLP for Positive Impact (NLP4PI), ACL 2025.

Derner, E., & Batistič, K. (2025). Gender Representation Bias Analysis in LLM-Generated Czech and Slovenian Texts. In *10th Workshop on Slavic Natural Language Processing (Slavic NLP 2025)*, ACL 2025.

Favero, L. A., Frases, D., Pérez-Ortiz, J. A., Käser, T., & **Oliver, N.** (2025). ELLIS Alicante at CQs-Gen 2025: Winning the critical thinking questions shared task: LLM-based question generation and selection. In 12th Workshop on Argument Mining, co-located with ACL 2025.

Derner, E., & Batistič, K. (2025). Beyond Words: Multilingual and Multimodal Red Teaming of MLLMs. In First Workshop on LLM Security (LLMSEC), ACL 2025.

Derner, E., Sansalvador de la Fuente, S., Gutiérrez, Y., Moreda, P., & **Oliver, N.** (2025). Leveraging Large Language Models to Measure Gender Representation Bias in Gendered Language Corpora. In 6th Workshop on Gender Bias in Natural Language Processing (GeBNLP), ACL 2025.

Hagendorff, T., **Derner, E.**, & **Oliver, N.** (2025). Large Reasoning Models Are Autonomous Jailbreak Agents. In arXiv:2508.04039.

Riccio, P., Domene Garcia, L., & **Oliver, N.** (2025). Towards Accessible Image Set Descriptions with ImageSet2Text. Workshop on Vision Foundation Models and Generative AI for Accessibility: Challenges and Opportunities.

Németh, G. D. (2025). Relaxing Core Assumptions: the Impact of Data, Model and Participation Heterogeneity on Performance, Privacy and Fairness in Federated Learning. In PhD Thesis, University of Alicante.

Arnaiz-Rodríguez, A., & Errica, F. (2025). Oversmoothing, Oversquashing, Heterophily, Long-Range, and more: Demystifying Common Beliefs in Graph Machine Learning. In 22nd International Workshop on Mining and Learning with Graphs at ECML-PKDD 2025.

Arnaiz-Rodríguez, A., Corvelo Benz, N., Thejaswi, S., **Oliver, N.**, & Gomez Rodriguez, M. (2025). Towards Human-AI Complementarity in Matching Tasks. In The Third Workshop on Hybrid Human-Machine Learning and Decision Making (HLDM'25) at ECML-PKDD 2025.

Riccio, P. (2025). Human aesthetics under the representational power of Artificial Intelligence. In PhD Thesis, University of Alicante.

Arnaiz-Rodríguez, A. (2025). A Sociotechnical Approach to Trustworthy AI: from Algorithms to Regulation. In PhD Thesis, University of Alicante.

Arnaiz-Rodríguez, A., Baidal, M., **Derner, E.**, Layton Annable, J., Ball, M., Ince, M., Perez Vallejos, E., & **Oliver, N.** (2025). Between Help and Harm: An Evaluation of Mental Health Crisis Handling by LLMs. In Under review.

Németh, G. D., Fani, E., Ng, Y. J., Caputo, B., Lozano, M. A., **Oliver, N.**, & Quadrianto, N. (2025). FedDiverse: Tackling Data Heterogeneity in Federated Learning with Diversity-Driven Client Selection. In 3rd IEEE International Conference on Federated Learning Technologies and Applications (FLTA25).

Gulati, A., D'Inca, M., Sebe, N., Lepri, B., & **Oliver, N.** (2025). Beauty and the Bias: Exploring the Impact of Attractiveness on Multimodal Large Language Models. In Eighth AAAI/ACM Confernece on AI, Ethics and Society (AIES).

Höltgen, B., & **Oliver, N.** (2025). Reconsidering Fairness Through Unawareness from the Perspective of Model Multiplicity. In EAAMO 2025.

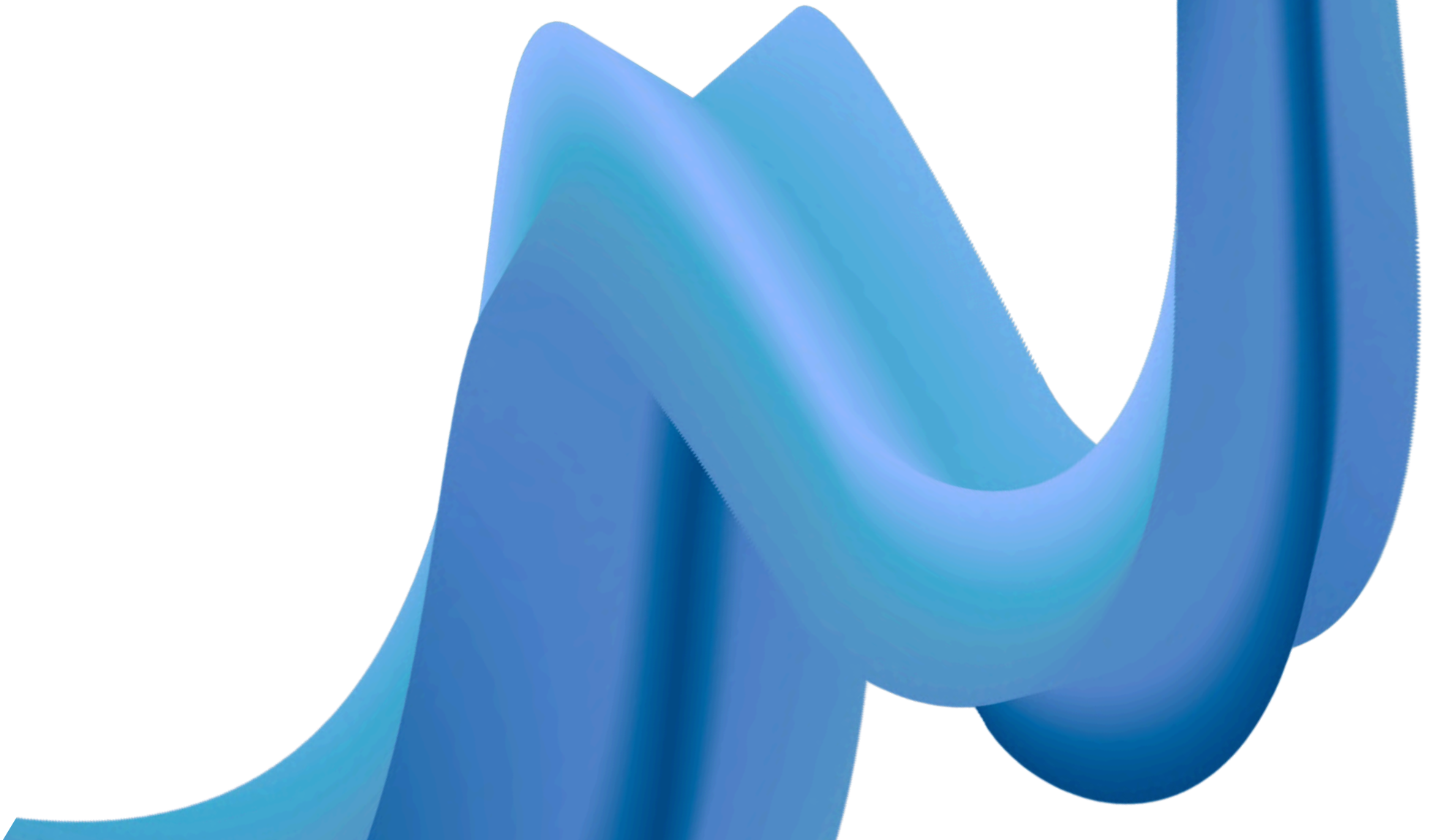
Favero, L. A., Pérez-Ortiz, J. A., Käser, T., & **Oliver, N.** (2025). Leveraging Small LLMs for Argument Mining in Education: Argument Component Identification, Classification, and Assessment. In The 25th International Workshop on Computational Models of Natural Argument (CMNA'25).

Gulati, A. (2025). Judging Books by Their Cover: The Impact of Facial Attractiveness on Humans and AI. In PhD Thesis, University of Alicante.

Communication Activities and Scientific and Technical Outreach Initiatives



Doctoral Theses



PhD Thesis Defense Gergely Németh

September 12th 2025

“Relaxing Core Assumptions: the Impact of Data, Model and Participation Heterogeneity on Performance, Privacy and Fairness in Federated Learning”

Our PhD student Gergely Németh successfully defended his thesis on 12 September 2025 and was awarded the International Doctorate distinction.

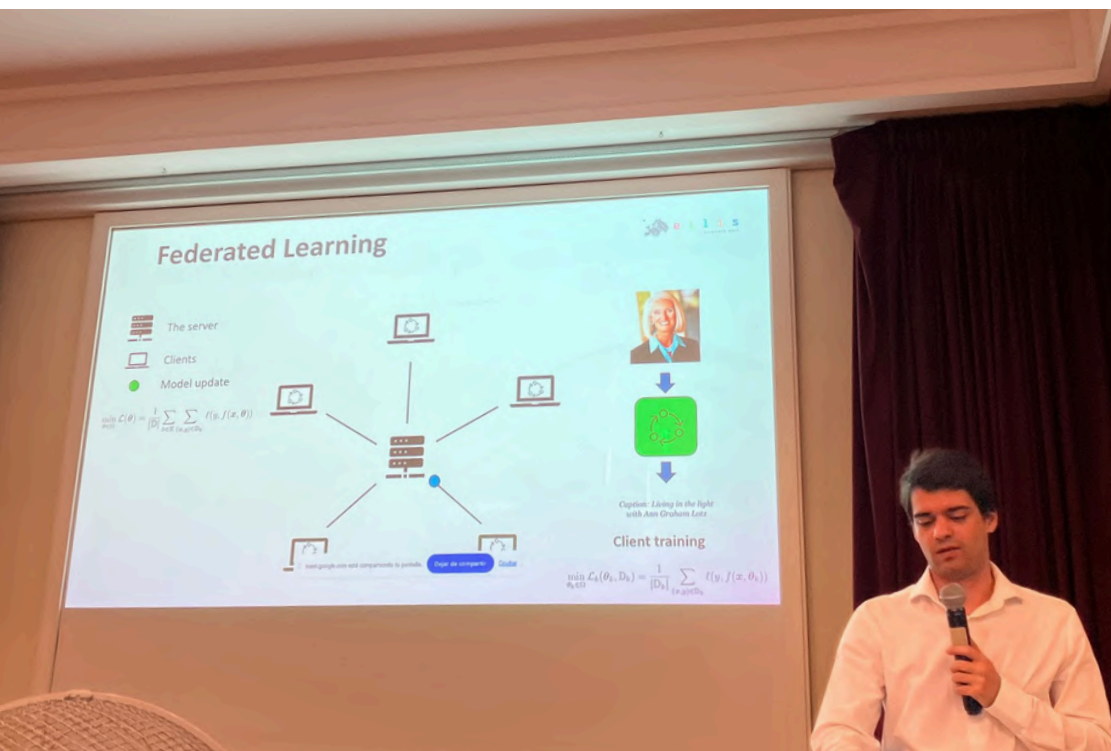
Gergely's PhD thesis explores federated learning (FL), a paradigm that enables machine-learning models to be trained on decentralised, distributed data while preserving privacy by design. In an FL system, individual clients train models using private data, while a central server aggregates their contributions to produce a global model based on consensus. Although the ideal scenario assumes that data and computational resources are distributed independently and identically across clients, substantial differences arise in real-world settings, both in the available data and in the computational capacity of each participant. The thesis focuses specifically on these heterogeneous environments, paying particular attention to how the server can adapt its aggregation methods to address such differences across clients.

The first line of research analyzes existing client-selection methods and proposes a new taxonomy of FL approaches in which the server actively manages client participation to achieve a global objective while accounting for client heterogeneity.



The thesis then examines model heterogeneity as a strategy for including clients with limited resources. More specifically, it investigates the privacy implications of computational constraints when clients use less complex models.

Lastly, the thesis proposes a solution to the problem of data heterogeneity through distribution-aware client selection. This approach helps to mitigate spurious correlations and improve algorithmic fairness in federated-learning systems.



Overall, the thesis contributes new strategies for making federated-learning systems more inclusive, robust and fair in real-world scenarios characterised by substantial differences across clients, both in terms of computational resources and data distribution.



PhD Thesis Defense Piera Riccio

September 22nd 2025

“Human aesthetics under the representational power of Artificial Intelligence”

Our PhD student Piera Riccio successfully defended her thesis on 22 September 2025. She was awarded the International Doctorate distinction and a cum laude distinction.

Piera's PhD thesis investigates how Artificial Intelligence (AI) technologies mediate human representation in contemporary visual culture. The work draws on Don Ihde's philosophical framework, which examines different forms of human-technology relations: embodiment, hermeneutic and alterity relations. By combining technical contributions with critical reflections on the socioethical and artistic dimensions of these systems, the thesis offers an interdisciplinary exploration of the role of AI in visual culture.

First, the thesis analyzes augmented-reality (AR) beauty filters as a form of embodiment relation, in which technology becomes transparent and changes the way people perceive and present their own faces. Through the introduction of new datasets, FairBeauty and B-LFW, and the OpenFilter tool, the research demonstrates how these filters reproduce Eurocentric beauty standards, subtly transforming identity in ways that reinforce historically rooted and racialized aesthetics.





Lastly, the thesis explores text-to-image (T2I) generative systems through the lens of the alterity relation, highlighting how users interact with technologies that produce outputs perceived as novel and autonomous. By auditing leading T2I platforms and introducing ImageSet2Text, a new method for summarizing image collections using vision-language models, the research identifies stylistic patterns and cultural biases embedded in AI-generated representations of people.

The thesis then examines the algorithmic censorship of artistic nudity as a hermeneutic relation, focusing on how moderation systems interpret and assess the obscenity of the human body. Using a mixed-methods approach that combines qualitative and quantitative contributions, this section reveals the limitations of current moderation technologies and calls for greater transparency, cultural sensitivity and accountability in content-moderation governance.



PhD Thesis Defense

Adrián Arnaiz

September 26th 2025

"A Sociotechnical Approach to Trustworthy AI: from Algorithms to Regulation"



Our PhD student Adrián Arnaiz successfully defended his thesis on 26 September 2025. He was awarded the International Doctorate distinction and a cum laude distinction.

Adrián Arnaiz's PhD thesis proposes a socio-technical framework for the effective implementation of trustworthy Artificial Intelligence (TAI), addressing the technical, human and regulatory dimensions of the harms arising from the use of AI. Rather than treating TAI as an exclusively technical objective, the research emphasises its interdisciplinary nature, aligning algorithmic development with societal needs and legal requirements throughout the life cycle of AI systems. The thesis focuses particularly on the potential harms associated with these systems and on how they can be mitigated through algorithmic design, regulatory development and effective implementation.

From a technical perspective, the thesis analyzes the harms arising from discrimination in algorithmic decision-making. First, it introduces FairShap, a new data-valuation method that quantifies the contribution of individual training examples to fairness metrics in decision-making systems. This approach enables biases and discrimination in high-risk systems to be diagnosed and mitigated more precisely, in line with the auditing obligations established by the EU AI Act.



Moving beyond traditional definitions of fairness in decision-making, the thesis proposes ERG, a graph-based approach for measuring and reducing structural inequalities relating to social capital within social networks. This proposal responds to emerging regulatory requirements, such as those set out in the EU Digital Services Act, which require online platforms to assess and mitigate systemic risks.

In the context of the practical use of algorithms, the thesis designs and evaluates a human-AI complementarity framework for collaborative decision-making in critical resource-allocation tasks. By combining human and algorithmic decisions through bandit-based strategies, the research explores how to design semi-automated systems capable of outperforming either humans or algorithms working independently. This approach follows key TAI principles, including technical robustness, human oversight and harm minimisation, as reflected in regulations such as the EU General Data Protection Regulation (GDPR) and in TAI guidelines.

Lastly, in the area of governance, the thesis examines the use of AI in workforce management in the context of Spanish labour law. The study identifies the legal frameworks applicable throughout the life cycle of AI systems, analyzes the relationship between TAI principles, the EU AI Act and employment-related obligations, and highlights existing tensions, such as the gap between correlation-based models and the legal requirement to provide causal justification for certain decisions.

Taken together, these contributions show that ensuring trust in AI systems requires far more than isolated algorithmic improvements. The thesis presents trustworthiness as the outcome of a complex socio-technical system shaped by the interaction between data, algorithms, institutions and regulatory constraints. It also offers practical insights for stakeholders seeking to develop AI systems that are technically robust, socially aligned and legally compliant.



PhD Thesis Defense

Aditya Gulati

December 12th 2025

"Judging Books by Their Cover: The Impact of Facial Attractiveness on Humans and AI"



Our PhD student Aditya Gulati successfully defended his thesis on 12 December 2025. Aditya was awarded the International Doctorate distinction.

Aditya's PhD thesis examines how human perception, memory and decision-making are shaped by numerous cognitive biases and heuristics that influence our actions and judgements. Despite their widespread presence, these biases are rarely taken into account in the design of human-AI systems. The thesis argues that the future of effective collaboration between humans and Artificial Intelligence will require the computational modeling and systematic understanding of these cognitive biases and, in certain cases, their deliberate replication. To this end, the thesis first introduces a conceptual framework that organises known cognitive biases into five categories, illustrated through representative examples and accompanied by open questions about their role in human-AI interaction.

The second part of the thesis focuses on a specific bias: the attractiveness halo effect, namely the tendency to attribute positive traits, such as intelligence or trustworthiness, to physically attractive individuals, even when attractiveness is irrelevant to the assessment. The research examines how this bias manifests itself in the digital age by analyzing the influence



of beauty filters on people's assessments of different attributes. This large-scale study, the most extensive conducted on this subject to date, resulted in a dataset of 924 images with high-quality benchmark ratings for attractiveness, intelligence, trustworthiness, sociability, happiness and other variables.

Building on these findings, the thesis introduces the concept of algorithmic lookism, understood as the tendency of algorithms, particularly AI systems, to discriminate on the basis of physical

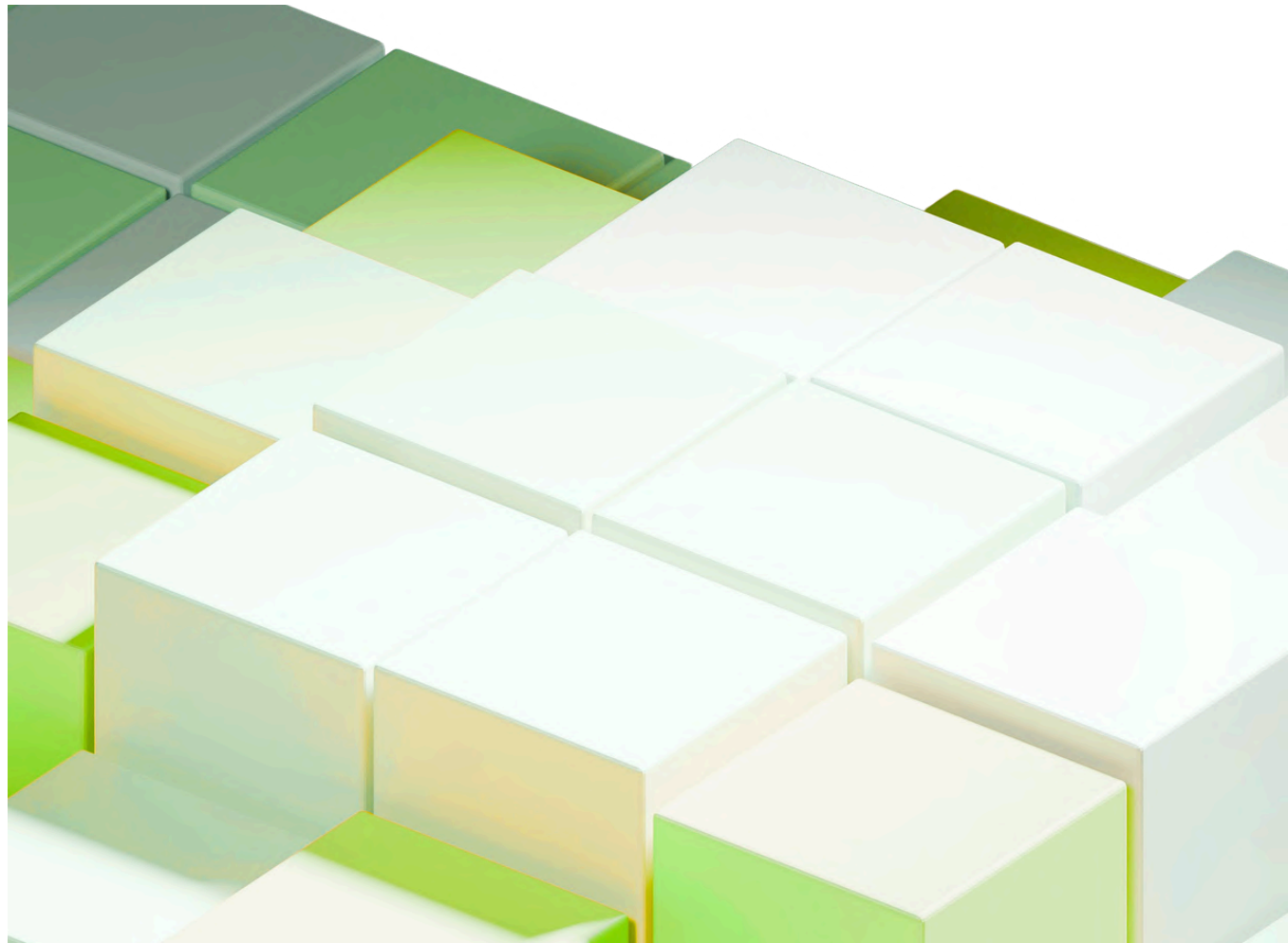


attractiveness. This phenomenon remains relatively underexplored. The research investigates this effect empirically in seven open multimodal large language models (MLLMs), demonstrating that these systems use attractiveness as an implicit criterion in their decision-making processes.

The thesis also analyzes synthetically generated faces and shows that images created from descriptions of positive traits are significantly more likely to depict people perceived as attractive. In addition, it assesses the subsequent impact of these AI-generated images on related tasks. The findings reveal critical mechanisms through which cognitive biases can propagate through AI systems, highlighting the need to develop ethical and equitable design principles for their deployment in real-world settings.



Research Events



9th Project Meeting ELIAS - European Lighthouse of AI for Sustainability

ELLIS Alicante
September 2nd-3rd 2025



The 9th ELIAS Project Meeting was organized by ELLIS Alicante and took place at our headquarters on September 2nd-3rd, 2025.

The meeting brought together around thirty participants and was divided into eleven sessions. These included updates on the status of each of the project's Work Packages, presentations on the use cases and discussions of the next steps.

ELLIS / ULB-MLG Workshop 2025

ELLIS Alicante
February 28th 2025

ELLIS Alicante hosted a workshop with visiting researchers from the Machine Learning Group (MLG) at the Université libre de Bruxelles (ULB), Belgium. The talks delivered as part of the workshop formed part of the Distinguished Talks series.

The workshop took place at the ELLIS Alicante premises on February 28th, and featured three MLG researchers: Professor Tom Lenaerts, Dr Axel Abels and Apurva Shah.

The Machine Learning Group (MLG)

Founded in 2004 by G. Bontempi, the Machine Learning Group (MLG) is a research unit within the Department of Computer Science at the Université libre de Bruxelles (ULB) in Brussels, Belgium. It forms part of the Faculty of Sciences and is currently co-directed by Professor Gianluca Bontempi and Professor Tom Lenaerts.

The MLG focuses its research on machine learning and behavioral intelligence, with particular emphasis on time-series analysis, large-scale data mining, causal inference, network inference, decision-making models and behavioral

analysis. Its work has applications in areas including data science, medicine, molecular biology, cybersecurity and social dynamics relating to cooperation and emotions.



ON DEFAULTS AND DELEGATION IN DECISION-MAKING – INTRODUCING THE LABS AND SOME OF THE ONGOING HUMAN-AI INTERACTION PROJECTS

Prof. Tom Lenaerts
Université libre de Bruxelles
Brussels, Belgium

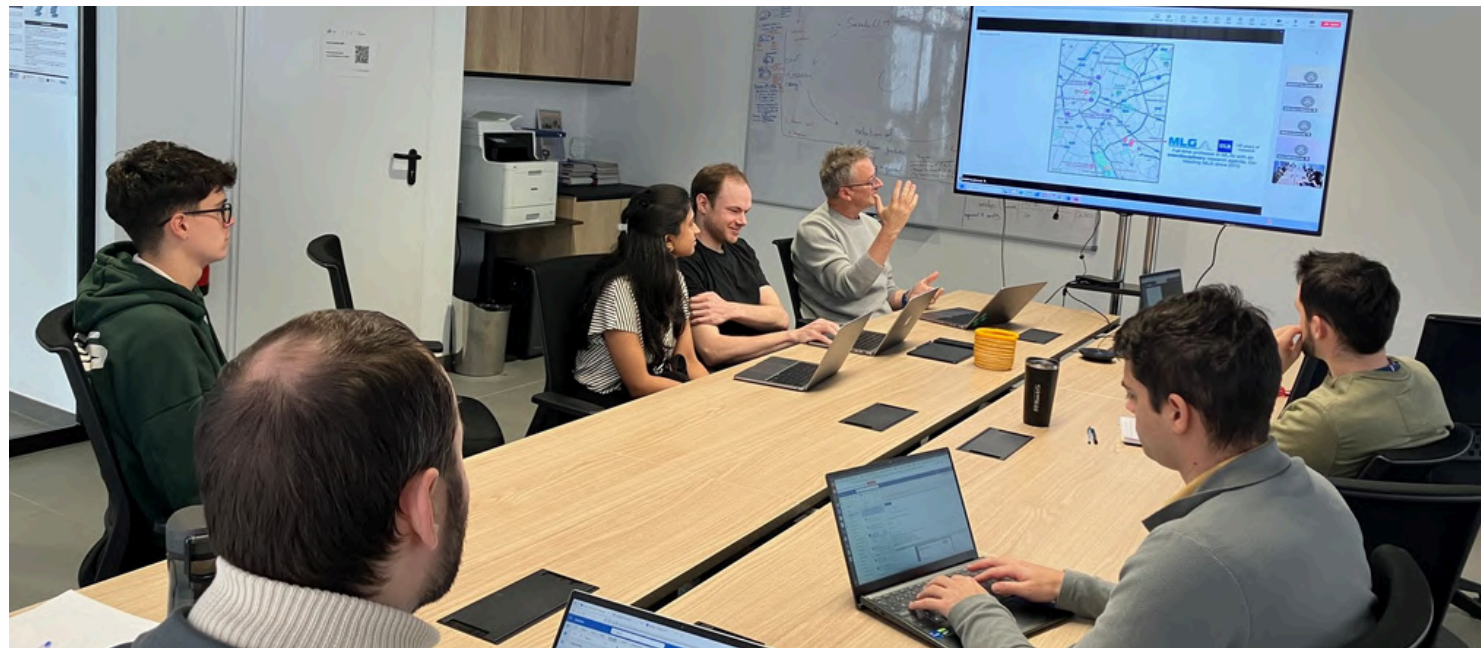
The first talk of the “ELLIS / ULB-MLG Workshop 2025” was delivered by Tom Lenaerts, Professor at the Université libre de Bruxelles (ULB) and co-director of the Machine Learning Group (MLG).

His research focuses on machine learning, multi-agent systems and collective decision-making, with particular emphasis on artificial social intelligence.

During the talk, Lenaerts examined the role of default options in decision-making, analyzing how they can implicitly influence agents' behavior. Although

defaults often go unnoticed, they can have a significant impact on outcomes, particularly depending on whether they are designed to promote individual interests or prosocial goals. Their implications were considered in collective-action contexts, such as the management of shared resources.

The talk also addressed the delegation of decisions to Artificial Intelligence systems, exploring how interactions between human and artificial agents can affect collective outcomes. Particular attention was paid to situations involving potential conflicts of interest, as well as to the impact that the partial replacement of human agents with AI systems may have in such scenarios. Taken together, these issues provide a basis for examining the role of decision-making mechanisms and their influence on behavior and outcomes in complex social environments.



ENHANCING COLLECTIVE INTELLIGENCE THROUGH LEARNED AGGREGATION

Axel Abels
Université libre de Bruxelles
Brussels, Belgium

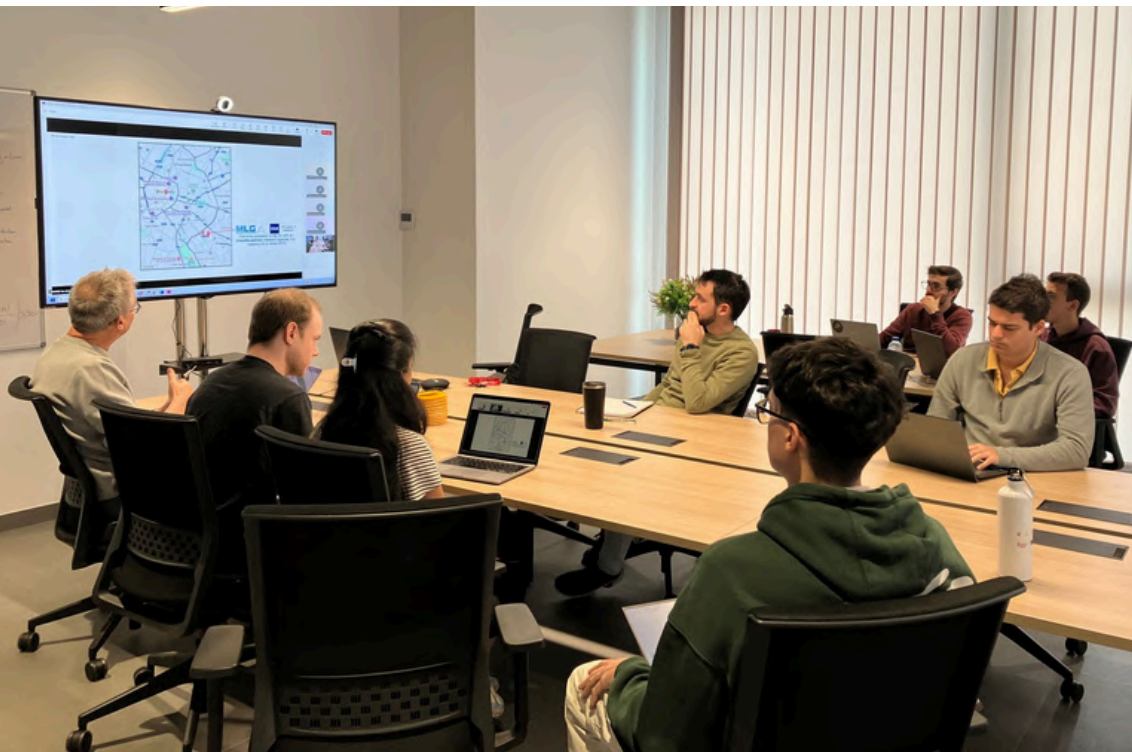
The second talk of the “ELLIS / ULB-MLG Workshop 2025” was delivered by Axel Abels, a postdoctoral researcher affiliated with the Machine Learning Group (MLG) at the Université libre de Bruxelles (ULB) and the AI Lab at Vrije Universiteit Brussel.

His doctoral research focused on collective decision-making (CDM), and more specifically on how machine learning and reinforcement learning can be used to harness the “wisdom of crowds”. His current research builds on this work, exploring fairness in collective decision-making and the benefits of hybrid human-AI systems in greater depth.

During his talk, Dr Abels discussed the inherent limitations of human decision-making, particularly those arising from individual and social biases that can lead to suboptimal outcomes in collective settings.

In this context, he highlighted the need to explore alternatives to traditional consensus-based mechanisms in order to improve the quality of group decisions.

He presented an approach based on the aggregation of independent judgements as a strategy for mitigating effects such as groupthink and herding. From this perspective, the talk examined how optimising the processes used to aggregate collective knowledge can make it more difficult for individual biases to shape the final decision, thereby improving collective-intelligence outcomes.



FAIRNESS AND ALIGNMENT IN COLLECTIVE DECISION- MAKING UNDER UNCERTAINTY

Apurva Shah
Université libre de Bruxelles
Brussels, Belgium

Apurva Shah delivered the final talk of the “ELLIS / ULB-MLG Workshop 2025”.

Apurva Shah is a joint PhD student at the Université libre de Bruxelles and Vrije Universiteit Brussel in Belgium. His research focuses on fairness and alignment in collective decision-making under uncertainty.

The talk addressed collective decision-making as a process involving divergent objectives and preferences, which raises alignment challenges in contexts where explicit feedback is not available. Against this background, Shah explored aggregation and preference-elicitation methods aimed at ensuring fairness and managing uncertainty as a basis for designing more robust decision-making systems.



Miniworkshop: Mobile and Social Computing Lab @ FBK

ELLIS Alicante
April 9th 2025

ELLIS Alicante hosted a workshop with two visiting researchers from Fondazione Bruno Kessler (FBK) in Trento, Italy: Dr Massimiliano Luca and PhD student Nicolò Penzo. The talks delivered as part of the workshop formed part of the ELLIS Alicante Distinguished Talks series.

The mini-workshop took place at the ELLIS Alicante premises on April 9th, 2025, and featured two researchers from the Mobile and Social Computing Lab.

Mobile and Social Computing Lab

The Mobile and Social Computing Lab at Fondazione Bruno Kessler develops computational methods for analyzing individual and collective human behavior using digital data from sources such as mobile devices, social media and banking transactions.

Its research covers human mobility, urban and social phenomena, graph neural networks, and cooperative AI systems based on agents and large language models (LLMs).



URBAN SAFETY PERCEPTION THROUGH THE LENS OF LARGE MULTIMODAL MODELS

Dr. Massimiliano Luca
Bruno Kessler Foundation
Trento, Italy

On April 9th 2025, Dr Massimiliano Luca was invited by ELLIS Alicante to participate in the Miniworkshop: Mobile and Social Computing Lab, held as part of the Distinguished Speaker Series.

Dr Massimiliano Luca is a senior researcher at the Mobile and Social Computing Lab, specifically at the Center for Augmented Intelligence of the Fondazione Bruno Kessler (FBK). His research focuses on Artificial Intelligence applied to mobility, transport and urban science.

During his talk, Dr Massimiliano Luca presented an approach to analyzing perceptions of safety in urban environments through the use of large-scale multimodal models.

In particular, he explored the use of street view images as a source of information for assessing how urban spaces are perceived, proposing these models as a scalable and efficient alternative to traditional survey-based methods.

He also analyzed how perceptions of safety vary according to sociodemographic factors, simulated through the use of prompts based on different “personas”. The results showed significant differences depending on variables such as age, gender and nationality, as well as the influence of visual features of the environment, such as isolation or the state of maintenance, on perceptions of insecurity. This highlighted the potential of these models for urban analysis and public policy planning.



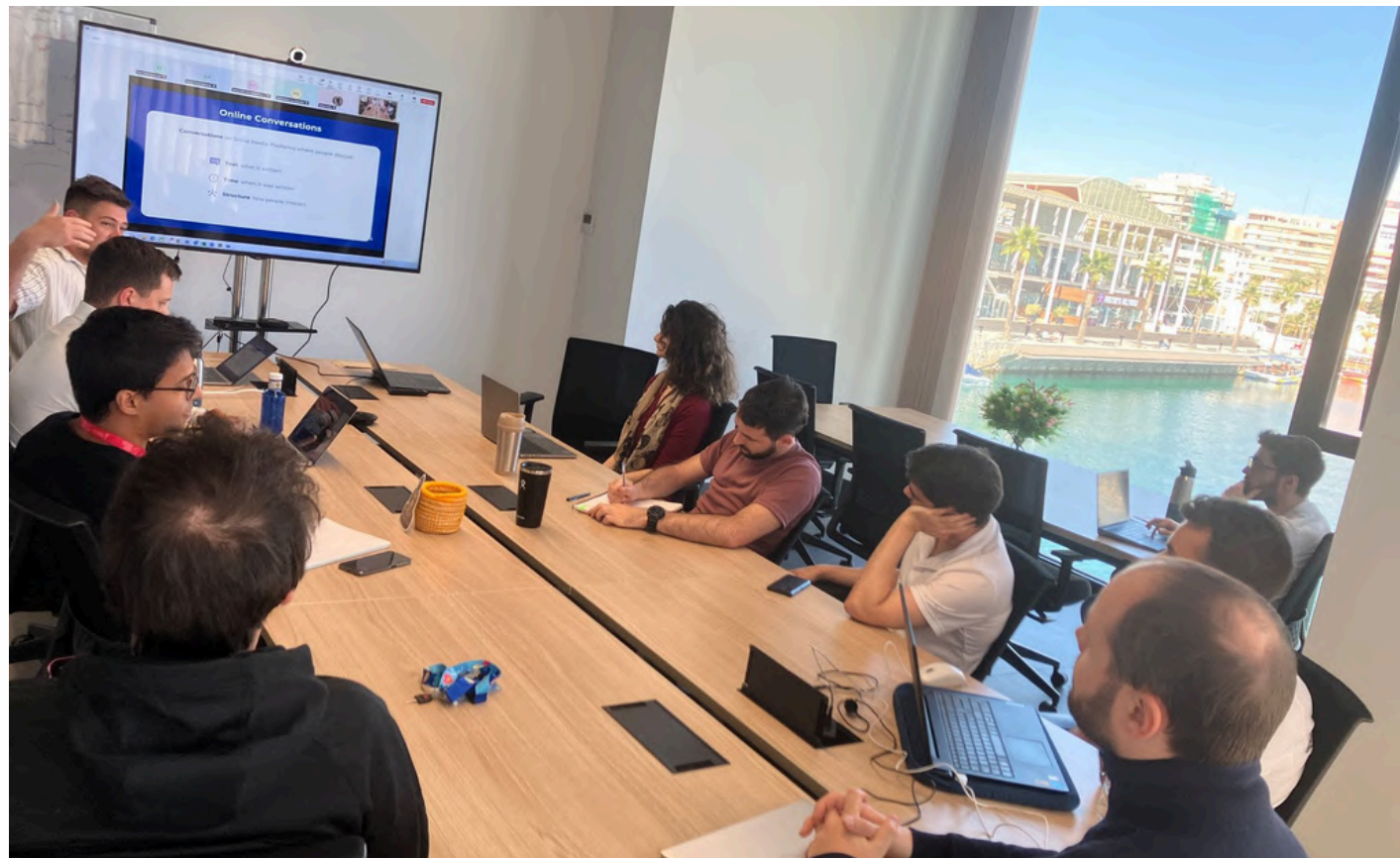
INTERACTING IN NATURAL LANGUAGE IN THE ERA OF LLMs: NLP FOR ONLINE DISCUSSIONS AND MULTI-PARTY CONVERSATIONS

Nicolò Penzo
Predoctoral Researcher
Universidad de Trento
Trento, Italy

On April 9th 2025, Nicolò Penzo was invited by ELLIS Alicante to participate in the Student Talks series, as part of the ELLIS / FBK-MobS Workshop.

Nicolò Penzo is a predoctoral researcher at the Mobile and Social Computing Lab (MobS) of the Fondazione Bruno Kessler (FBK). His research focuses on natural language processing and the analysis of Multi-Party Conversations in digital environments.

Penzo discussed the challenges involved in analyzing written conversations between multiple participants in digital environments, highlighting the influence of interaction dynamics on their meaning. In this context, he presented the challenges faced by models when processing Multi-Party Conversations and proposed an approach based on synthetic data generation, together with an evaluation framework for comparing the effectiveness of different strategies.



Serie: "Distinguished 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.



THE ROLE OF ARTIFICIAL INTELLIGENCE IN ACHIEVING THE SUSTAINABLE DEVELOPMENT GOALS

Dr Ricardo Vinuesa
KTH Royal Institute of Technology
Stockholm, Sweden

On February 20th 2025, ELLIS Alicante held the 2nd "Distinguished Lectures Mini-Workshop", featuring two professors: Professor Ricardo Vinuesa from KTH Royal Institute of Technology and Professor Alberto Conejero from the Universidad Politénica de Valencia.

Dr Vinuesa's talk focused on the intersection between Artificial Intelligence and the 17 Sustainable Development Goals. The emergence of Artificial Intelligence (AI), together with its increasingly widespread impact across numerous sectors, requires an assessment of its effect on the achievement of the Sustainable Development Goals. Through a consensus-based expert consultation process, the study found that AI could support the achievement of 134 targets across all the goals, but could also hinder 59 of them. However, current research approaches overlook important aspects.

The rapid development of AI must be accompanied by the necessary regulatory insight and oversight to ensure that AI-based technologies contribute to sustainable development. Otherwise, shortcomings may arise in relation to transparency, safety and ethical standards.



ANOMALOUSNET: A HYBRID APPROACH WITH ATTENTION U-NETS AND CHANGE POINT DETECTION FOR ACCURATE CHARACTERIZATION OF ANOMALOUS DIFFUSION IN VIDEO DATA

Dr J. Alberto Conejero
Universitat Politècnica de València
Valencia, Spain

The second speaker at the 2nd “Distinguished Lectures Mini-Workshop” was Professor Alberto Conejero.

During his talk, Professor Conejero addressed the presence of anomalous diffusion in a range of systems, including the transport of proteins within cells, the movement of animals in complex environments, the dispersion of pollutants in groundwater and the behavior of nanoparticles in synthetic materials. He explained that accurately estimating the anomalous diffusion exponent and the diffusion coefficient from particle trajectories makes it possible to gain a better

understanding of the underlying dynamics of a system, as well as to identify behavioral patterns and detect changes in diffusion states.

The talk also highlighted the challenges that short, noisy video sequences, which often generate heterogeneous and incomplete trajectories, pose for traditional statistical methods.

In this context, Conejero presented a data-driven method that integrates particle tracking, an Attention U-Net architecture and change-point detection. The methodology achieved strong results in the 2º Anomalous Diffusion (AnDi) Challenge, ranking among the top two proposals for single-trajectory tasks in the video-based track and among the top six for the ensemble task in the single-trajectory track.



TRUSTWORTHY AI AND COMPOSITIONALITY IN MULTIMODAL LANGUAGE MODELS

Dr Adam Dahlgren Lindström
Umeå University
Umeå, Sweden

On March 26th 2025, ELLIS Alicante invited Dr Adam Dahlgren Lindström to give a talk as part of the ELLIS Alicante Distinguished Talks series.

Dr Adam Dahlgren Lindström is a postdoctoral researcher in the Department of Computing Science at Umeå University in Sweden, where he is a member of the Responsible AI group and the AI Policy Lab.

His research focuses on multimodal models, reasoning and compositional generalisation, as well as the evaluation and development of trustworthy AI systems and human-AI collaboration.



Dahlgren Lindström presented two complementary perspectives on trustworthy AI: a technical perspective and a sociotechnical one.

From a technical perspective, he discussed visual reasoning experiments with multimodal models, highlighting limitations in their compositional generalisation capabilities and inconsistencies in problem-solving, based on evaluations using synthetic benchmarks from the CLEVR domain.

He also examined a sociotechnical approach centred on the use of Reinforcement Learning from Human Feedback (RLHF) as an alignment mechanism. In this context, he discussed the limitations of RLHF as a method and raised questions concerning the very concept of alignment in AI systems, highlighting the challenges involved in developing trustworthy multimodal models.

MINDS IN MACHINES? ON THE PSYCHOLOGY AND DECEPTION CAPABILITIES OF LARGE LANGUAGE MODELS

Dr Thilo Hagendorff
University of Stuttgart
Stuttgart, Germany

On May 9th 2025, Dr Thilo Hagendorff was invited by ELLIS Alicante to participate in the Distinguished Speaker Series.

Dr Thilo Hagendorff is a researcher at the University of Stuttgart. His research focuses on the safety and ethics of Artificial Intelligence and on the intersection between machine learning and psychology.

During his talk, Dr Hagendorff explored the analysis of large language models (LLMs) from a psychological perspective, proposing that they be treated as participants in experiments originally designed for humans. In this context, he showed how methodologies from the field of psychology can be applied to study the behavior of these systems and evaluate their capabilities in different scenarios.

In particular, he focused on the study of deception in LLMs, highlighting its implications for the safety and alignment of AI systems. This approach underlines the need to develop new evaluation frameworks to gain a better understanding of the behavior of generative models, particularly as their use expands into increasingly complex contexts that more closely resemble human interaction.



THE IMPACT OF RANKING INTERVENTIONS AND TASK-LEVEL FACTORS IN RECRUITMENT INTERFACES FOR SHORTLISTING

Prof. Carlos Castillo
Universitat Pompeu Fabra
Barcelona, Spain

On May 15th 2025, Prof. Carlos Castillo was invited by ELLIS Alicante to participate in the 3rd Distinguished Lectures Mini-Workshop.

Dr Castillo is an ICREA Research Professor at the Universitat Pompeu Fabra, where he leads the Web Science and Social Computing group. Their research focuses on information retrieval, web mining and the analysis of sociotechnical systems, with particular attention to fairness and credibility in digital environments.

During their talk, he analyzed the role of automated ranking systems in recruitment processes, which are increasingly common in recruitment interfaces. In this context, Prof. Castillo presented a series of user studies aimed at understanding how



different factors, such as the order in which candidates are presented, the characteristics of the tasks and the representation of information, influence decision-making when selecting candidates from ranked lists.

In particular, Prof. Castillo showed that equitable visibility in rankings does not necessarily guarantee equitable outcomes in the final selection. These findings highlight the complexity of designing fair ranking systems and have direct implications for the development of algorithms, interfaces and strategies aimed at promoting diversity in recruitment processes.

FROM BIOLOGY TO MACHINE LEARNING AND VICE VERSA

Marina Martínez-García
Universidad Pompeu Fabra
Barcelona, Spain

On May 15th 2025, Professor Marina Martínez-García was invited by ELLIS Alicante to participate in the 3rd Distinguished Lectures Mini-Workshop.



Dr Martínez-García is a lecturer in the Department of Mathematics at the Universitat Jaume I in Castellón and a researcher at IMAC. Her research focuses on machine learning and deep learning, with applications in areas such as neuroscience and complex data analysis.

During her talk, she explored the bidirectional relationship between biology and machine learning, highlighting how concepts and methodologies from both fields can be transferred and used to enrich one another. In this context, she presented examples of how biological principles can inspire the development of new models and approaches in machine learning.

Professor Martínez-García also discussed how machine learning techniques can be applied to the study of complex biological systems, facilitating data analysis and the identification of patterns in contexts where traditional methods are limited. This interdisciplinary approach highlights the potential of integrating both fields to advance our understanding of complex systems.



FROM BIOLOGY TO MACHINE LEARNING AND VICE VERSA

Dr Ramon Ruiz-Dolz
University of Dundee
Dundee, Scotland, United Kingdom

On June 4th 2025, Professor Ramon Ruiz-Dolz was invited by ELLIS Alicante to participate in the Distinguished Speaker Series.

Dr Ramon Ruiz-Dolz is a Lecturer in Computing at the University of Dundee. His research focuses on computational argumentation and natural language processing, with a particular interest in analyzing the reasoning capabilities of large language models (LLMs).

During his talk, he analyzed the role of reasoning capabilities in LLMs, highlighting how these are currently evaluated through benchmarks covering different types of reasoning. In this context, he situated the recent development of advanced models within the race towards a purported Artificial General Intelligence (AGI), focusing on the need to better understand what it actually means for these systems to “reason”.

He also explored the integration of concepts from argumentation theory into natural language processing algorithms as a means of improving the explainability of these models. This approach makes it possible not only to analyze their reasoning capabilities in greater depth, but also to develop tools aimed at fostering critical thinking in the use of LLM-based systems.

NLP APPLICATIONS AT UPV TO FIGHT AGAINST DISINFORMATION AND SEXISM

Prof. Paolo Rosso
Universitat Politècnica de València
Valencia, Spain

On June 12th 2025, Professor Paolo Rosso was invited by ELLIS Alicante to participate in the Distinguished Speaker Series.

Dr Rosso is a Professor of Computer Science at the Universitat Politècnica de València (UPV). His research focuses on the detection of harmful information on social media, including disinformation, conspiracy theories and hate speech, with particular attention to subtle forms such as irony and sarcasm.

Professor Rosso analyzed the role of social media as a channel for the spread of disinformation, examining how malicious actors use fake news and conspiracy narratives to generate negative emotions among the public. In this context, he presented recent cases related to the dissemination of misleading content on platforms such as X and TikTok, highlighting the challenges these phenomena pose for automatic detection.

He also examined the limitations of current AI approaches in distinguishing between critical thinking and conspiratorial thinking, a key distinction in the design of content moderation systems. Finally, he mentioned ongoing initiatives focused on identifying sexist content on social media, including datasets and competitions aimed at improving NLP capabilities in this area.





"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.

EMPOWERING AFRICAN LANGUAGES THROUGH NLP: THE MASAKHANE PROJECT

Ayodele Awokoya
Predoctoral Researcher
University of Ibadan, Nigeria

On January 24th 2025, Ayodele Awokoya was invited by ELLIS Alicante to participate in the Student Presentation series.

Ayodele Awokoya is a lecturer in the Department of Computer Science at McPherson University in Nigeria and a PhD candidate at the University of Ibadan. Her research focuses on NLP and machine translation for low-resource African languages, with the aim of reducing the linguistic divide in technology.

During her talk, she presented the Masakhane project, a collaborative, open-source initiative led by African researchers and aimed at developing natural language processing tools for African languages. In this context, she highlighted the significant underrepresentation of these languages in current technologies, despite accounting for approximately one third of the world's languages. This limits their presence in digital systems and AI models.

FAIRNESS AND ALIGNMENT IN COLLECTIVE DECISION-MAKING UNDER UNCERTAINTY

Mengting Wei
ELLIS Predoctoral Researcher
Universidad de Oulu
Oulu, Finland

On December 10th 2025, Mengting Wei was invited by ELLIS Alicante to participate in the Student Presentation series.

Mengting Wei is a PhD student at the University of Oulu, where she conducts research at the Center for Machine Vision and Signal Analysis (CMVS) under the supervision of Prof. Guoying Zhao. She is a doctoral student at the ELLIS PhD Program and is co-supervised by Dr Nuria Oliver. Her research focuses on face generation and editing.

During her talk, she analyzed how generative AI systems respond to different human emotions when generating images, exploring how the distribution of generated images varies in terms of attributes such as age, gender and ethnicity depending on the emotions represented.

Reading Group: ELLIS / ELLIOT Human-Centric Machine Learning

The **ELLIS-ELLIOT Human-Centric Machine Learning Reading Group** is an international reading group organized since 2022 by researchers from ELLIS Alicante and focused on human-centric machine learning.

Its aim is to bring together researchers and students interested both in gaining a broad overview of the field and in exploring it in greater depth. By reading research papers on different topics within this area and discussing new ideas, approaches and sources of bias, the group seeks to better understand how algorithmic and human decisions influence one another.

In 2025, the Human-Centric Machine Learning Reading Group became part of the ELLIS-ELLIOT initiative, in line with the aim of strengthening collaboration between the two initiatives and promoting a greater diversity of perspectives.

January 2025

On the Privacy Risks of Algorithmic Fairness (Hongyan Chang and Reza Shokri), 2021

Presenter: **Gergely D. Nemeth**

[Paper Link](#)

March 2025

Aligned LLMs Are Not Aligned Browser Agents (Priyanshu Kumar, Elaine Lau, Saranya Vijayakumar, Tu Trinh, Elaine T Chang, Vaughn Robinson, Shuyan Zhou, Matt Fredrikson, Sean M. Hendryx, Summer Yue, Zifan Wang), 2025

Presenter: **Erik Derner & Kristina Batistič**

[Paper Link](#)

March 2025

iLLuMinaTE: An LLM-XAI Framework Leveraging Social Science Explanation Theories Towards Actionable Student Performance Feedback (Vinitra Swamy, Davide Romano, Bhargav Srinivasa Desikan, Oana-Maria Camburu, Tanja Käser), 2025

Presenter: **Vinitra Swamy**

[Paper Link](#)

November 2025

Personality Traits in Large Language Models (Greg Serapio-García, Mustafa Safdari, Clément Crepy, Luning Sun, Stephen Fitz, Peter Romero, Marwa Abdulhai, Aleksandra Faust, Maja Matarić), 2025.

Presenter: **Kim Zierahn**

[Paper Link](#)

Other scientific outreach activities



Participation in conferences and events

Members of the ELLIS Alicante team have taken part as organizers, speakers or invited panellists in numerous national and international conferences and events, including the following:

January 2025

Dr Oliver was an invited speaker at an event organized by Fundación ONCE and ELLIS Alicante, where she gave a talk on Artificial Intelligence and Accessibility. Madrid, Spain.

Piera Riccio was appointed Workshop Chair of the International Conference on Computational Creativity (ICCC) 2025. Campinas, Brazil.

Piera Riccio presented at the Ethics and Aesthetics of Artificial Intelligence 2025 Conference". Venice, Italy.

Dr Oliver was an invited speaker at the high-level event co-organized by EMBL and ELLIS, Shaping Europe AI Leadership: Insights from Science, where she gave a talk on ELLIS. Brussels, Belgium.

February 2025

Adrian Arnaiz gave a talk on "IA para el Bien Social" at the Facultad de Filosofía of the Universidad de Sevilla, as part of the Proyecto de Innovación Docente "Proyectos Fin de Carrera en Filosofía en el Contexto de la IA". Seville, Spain.

Adrian Arnaiz gave a talk on his thesis, "A Holistic View of Algorithmic Fairness and Trustworthy AI: Society and Algorithms", at the Centro Común de Investigación de la Comisión Europea (JRC-EU). Seville, Spain.



Lucile Favero presented Dr Erik Derner's work at the Collaborative AI and Modeling of Humans Bridge at AAAI25, Philadelphia, United States.

Dr Oliver was an invited speaker at HUMANIZAR LA EDUCACIÓN EN LA ERA DIGITAL, where she gave a talk on Artificial Intelligence and Education. Salamanca, Spain (online).

Dr Oliver was an invited panellist at the AI, Science and Society conference to discuss the Informe Internacional sobre la Seguridad de la Inteligencia Artificial. Paris, France.

Piera Riccio gave a talk as part of the MilanNLP Seminar Series. Milan, Italy (online).



Dr Oliver was an invited speaker at the Rotary Lucentum club, where she gave a talk on ELLIS Europe and ELLIS Alicante. Alicante, Spain.

Fundacion Anna Lindh and Casa Mediterraneo visited ELLIS Alicante.

Lucile Favero was the lead organiser of the Collaborative AI and Modeling of Humans Bridge Program at AAAI25, Philadelphia, United States.





March 2025

Dr Oliver, Adrian Arnaiz, Gergely Nemeth, Aditya Gulati and Rebeca de Miguel took part in the plenary meeting of the H2020 ELIAS project. Bucharest, Romania, and online.

Piera Riccio gave a talk as part of the ELLIS Turin Talk at Politecnico di Torino. Turin, Italy, and online.

Dr Oliver was an invited speaker in the CSIC's "Cerebro vs IA" series, where she gave a talk entitled "Hacia una IA socialmente sostenible". Alicante, Spain.

April 2025

Piera Riccio gave a talk at the Universidad de Osaka. Osaka, Japan.

Aditya Gulati and Dr Erik Derner delivered a talk and workshop on university studies, research and Artificial Intelligence for students at Colegio Santo Domingo. Orihuela, Spain.

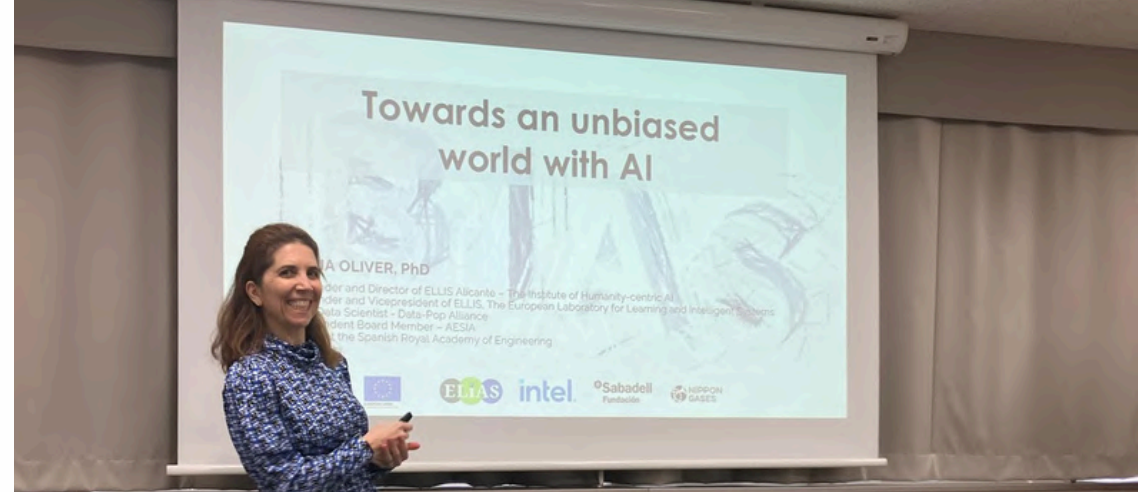
Adrián Arnaiz and Erik Derner delivered a talk and workshop on university studies, research and Artificial Intelligence for students at IES Mutxamel. Mutxamel, Spain.



The ELLIS Alicante team **ranked 1st** in the Critical Questions Generation Shared Task organized by the 12th Workshop on Argument Mining, co-located with ACL 2025.

Julien Colin gave a talk at Brown University on his work on evaluating the interpretability of computer vision systems. Providence, United States.

Dr Oliver gave an invited talk at the University of Osaka entitled "Towards an unbiased world with AI". Osaka, Japan.



May 2025

Dr Oliver gave an invited talk at The Hong Kong Polytechnic University entitled "Hacia un mundo sin sesgos con IA". Hong Kong.

Dr Oliver gave an invited talk at the Club de Roma entitled "ELLIS: Hacia la Soberania Europea en Investigacion en IA". Pamplona, Navarre (online).

Lucile Favero presented her work on the shared task "Preguntas de pensamiento crítico: generación y evaluación con modelos LLM" at the Jornada del Doctorado en Informática, JDI2025, held at the Universidad de Alicante. Alicante, Spain.

ELLIS Alicante co-organized a workshop in collaboration with the UCLM on Artificial Intelligence and labour law in the context of the Carta de Derechos Digitales project. Alicante, Spain.





July 2025

Piera Riccio was an invited speaker at an MPI-SP seminar. Germany (online).

Lucile Favero gave the opening talk at the GenAIHED workshop, co-organized with AIED2025, entitled "¿Los tutores de IA empoderan o esclavizan a los estudiantes? Hacia un uso crítico de la IA en la educación". Palermo, Italy.

Banco Sabadell visited the ELLIS Alicante headquarters.

June 2025

Dr Oliver was an invited speaker at the Curso de Dirección Estratégica de Universidades, where she gave a talk entitled "Limitaciones y Retos Éticos de la IA". Madrid, Spain.

Dr Oliver was an invited panellist at the II Foro Económico y Social del Mediterráneo, where she took part in a panel on Artificial Intelligence. Málaga, Spain.

Dr Oliver was an invited speaker at WomenNOW, where she took part in a conversation on Artificial Intelligence. Madrid, Spain.

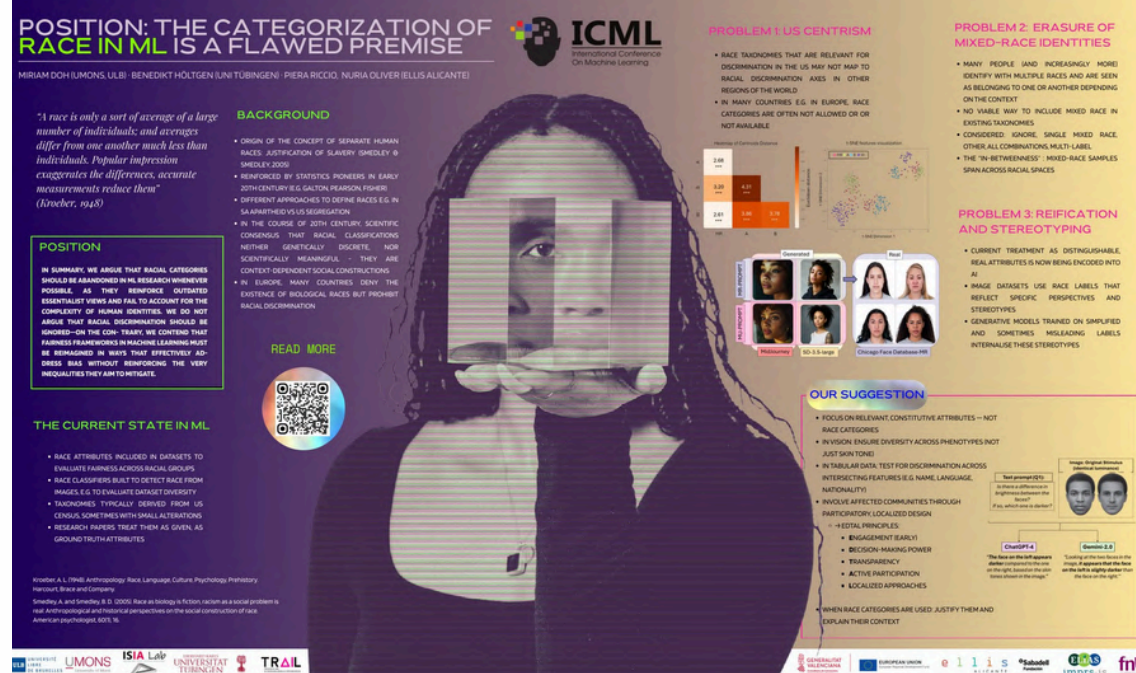
Adrian Arnaiz presented his work "Structural Group Unfairness" at ICWSM 2025. Copenhagen, Denmark.



Lucile Favero presented her work, "ELLIS Alicante at CQs-Gen 2025: Winning the critical thinking questions shared task: LLM-based question generation and selection", at the 12th Workshop on Argument Mining, co-organized with ACL2025, Vienna, Austria.

Dr Erik Derner and Miguel Baidal presented their work, "Guardians of Trust: Risks and Opportunities for LLMs in Mental Health", at the 4th Workshop on NLP for Positive Impact (NLP4PI), held as part of ACL 2025, Vienna, Austria.

Dr Erik Derner and associate researcher Kristina Batistič presented their work, "Gender Representation Bias Analysis in LLM-Generated Czech and Slovenian Texts", at the 10th Workshop on Slavic Natural Language Processing (Slavic NLP 2025), held as part of ACL 2025, Vienna, Austria.

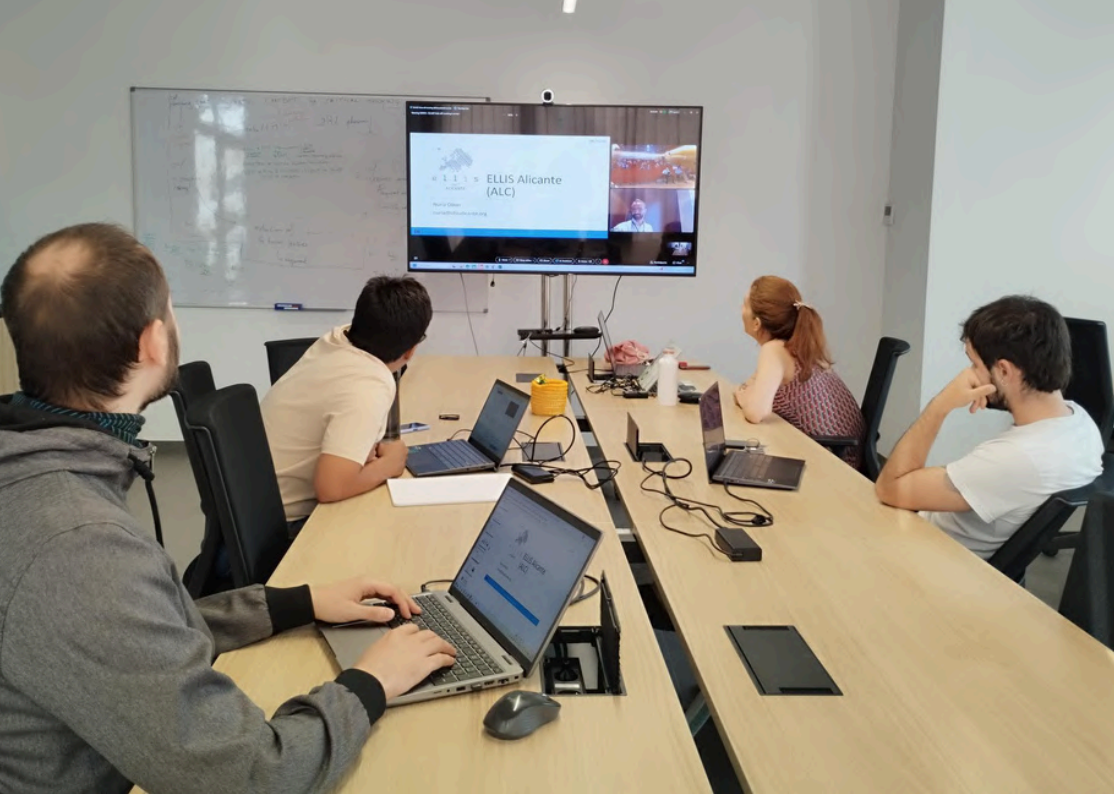


August 2025

Dr Erik Derner and Kristina Batistič, an associate researcher at ELLIS Alicante, presented their work, "Beyond Words: Multilingual and Multimodal Red Teaming of MLLMs", at the 1st Workshop on LLM Security (LLMSEC), held as part of ACL 2025, Vienna, Austria.

Dr Erik Derner presented his work, "Leveraging Large Language Models to Measure Gender Representation Bias in Gendered Language Corpora", at the 6th Workshop on Gender Bias in Natural Language Processing (GeBNLP), held as part of ACL 2025, Vienna, Austria.





September 2025

Adrian Arnaiz presented his work, "Demystifying Common Beliefs in Graph Machine Learning", at the GLOW Reading group (Graph Learning on Weds). Online.

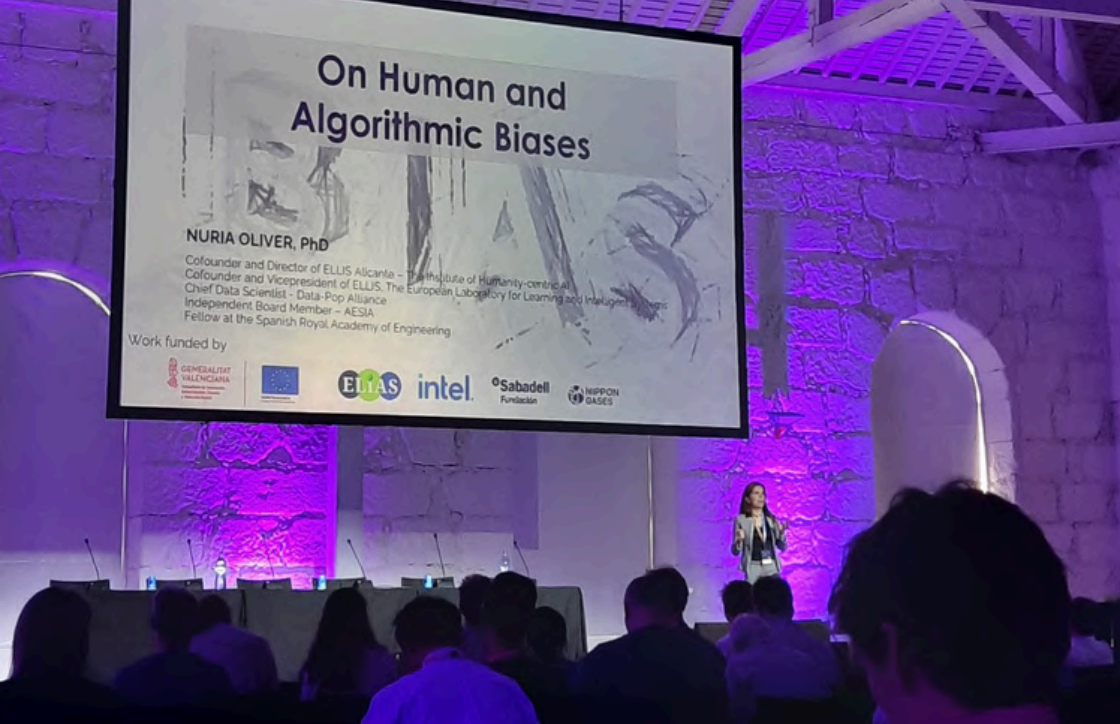
ELLIS Alicante hosted the plenary meeting of the Horizon Europe ELIAS project. Alicante, Spain.

Dr Oliver gave an invited talk at the 33rd edition of the IEEE Requirements Engineering 2025 conference entitled "Towards a fairer world - Unconverging and addressing human and algorithmic biases". Valencia, Spain.

Dr Oliver gave an invited talk at Microsoft Research entitled "Towards a fairer world - Unconverging and addressing human and algorithmic biases". Redmond, WA, United States.

Dr Oliver gave a keynote talk at the ELLIS Doctoral Symposium (EDS 2025) entitled "Towards a fairer world with AI". Warsaw, Poland.





Piera Riccio visited the Data Analytics Lab at ETH. Zurich, Switzerland.

Dr Oliver presented ELLIS Alicante at an OsloMet seminar. Oslo, Norway (online).

Lucile took part in the round table "¿Puede hacernos la IA más humanos?" at the Samsung Innovation Campus Summit: AI for a better world. Madrid, Spain.

Lucile took part in a round table at La cultura del talento, organized by Fundación Banco Sabadell. Barcelona, Spain.

Gergely Németh defended his doctoral thesis entitled "Relaxing Core Assumptions: the Impact of Data, Model and Participation Heterogeneity on Performance, Privacy and Fairness in Federated Learning" at the Universidad de Alicante. Alicante, Spain.

Adrian Arnaiz presented his work, "Demystifying Common Beliefs in Graph Machine Learning", at the MLG workshop at ECML-PKDD 2025. Porto, Portugal.

Adrian Arnaiz presented his work, "Towards Human-AI Complementarity in Matching Tasks", at the HLDM'25 workshop at ECML-PKDD 2025. Porto, Portugal.



Dr Oliver gave a keynote talk at ECML-PKDD 2025 entitled "Towards a fairer world - Unconverging and addressing human and algorithmic biases". Porto, Portugal.

Piera Riccio defended her doctoral thesis entitled "Human aesthetics under the representational power of Artificial Intelligence" at the Universidad de Alicante. Alicante, Spain.

Adrian Arnaiz defended his doctoral thesis entitled "A Sociotechnical Approach to Trustworthy AI: from Algorithms to Regulation" at the Universidad de Alicante. Alicante, Spain.



October 2025

Lucile Favero gave a talk at the Jornada "JÓVENES, PANTALLAS Y VÍNCULOS: CONEXIÓN / DESCONEXIÓN?". Murcia, Spain.

Dr Gergely Németh presented his work, "FedDiverse: Tackling Data Heterogeneity in Federated Learning with Diversity-Driven Client Selection", at FLTA'25. Dubrovnik, Croatia.





November 2025

Dr Oliver moderated a session with Prof. Yoshua Bengio on the risks of AI as part of the IV Foro de Inteligencia Artificial. Alicante, Spain.

ELLIS Alicante was a co-organisier of the IV Foro de Inteligencia Artificial. Alicante, Spain.

Aditya Gulati presented his work at the encuentro de jóvenes investigadores ELLIOT. Online.

Dr Oliver gave an invited talk on Artificial Intelligence at IA-GOBCAN 2025. Las Palmas, Spain (online).

Dr Adrián Arnaiz gave an invited talk at the VII Conferencia "Innovar hoy para liderar mañana", organized by La Gran Fábrica IBI. Ibi (Alicante), Spain.

Aditya Gulati presented his work, "Beauty and the Bias: Exploring the Impact of Attractiveness on Multimodal Large Language Models", at AIES 2025. Madrid, Spain.

Lucile Favero presented the Maike project to the Centro Gallego de Innovación Educativa y Digital. Alicante, Spain.





December 2025

Aditya Gulati defended his doctoral thesis entitled “Judging Books by Their Cover: The Impact of Facial Attractiveness on Humans and AI” at the Universidad de Alicante. Alicante, Spain.

Lucile Favero presented her work, “Leveraging Small LLMs for Argument Mining in Education: Argument Component Identification, Classification, and Assessment”, at the 25th International Workshop on Computational Models of Natural Argument. Online.

Dr Adrian Arnaiz gave a talk on “Sesgos y Justicia Algorítmica en la Toma de Decisiones” at ADITECH and NAIR. Pamplona, Spain.



Top Media Coverage in Press, Television and Radio

The work of ELLIS Alicante has been featured in **more than a hundred** print and online publications, as well as receiving extensive television and radio coverage, both nationally and internationally.

Here is a small selection of its media impact in 2025.



Top press coverage:

20 años de compromiso

revista **haz** INNOVACIÓN SOCIAL

Nuria Oliver, Personaje Social 2025 por poner a las personas en el centro de la IA

Stephen Hawking señaló en 2016 que la inteligencia artificial sería “lo mejor o lo peor que le pase a la humanidad”; la científica Nuria Oliver lleva más de 30 años trabajando para que la tecnología se incline hacia el lado bueno de la balanza. Esta es su motivación, “poder decir que he contribuido a que sea lo mejor que nos ha pasado”, resume en una entrevista concedida a esta revista, en el año en el que la IA generativa se ha instalado para quedarse definitivamente en nuestras vidas.

Esther Barrio @Esther_Bame | 18 diciembre 2025

INFORMACIÓN

Educación

+ Ellis recibe el galardón San Alberto Magno de la Universidad de Alicante

La Facultad de Ciencias reconoce la implicación de la entidad dirigida por Nuria Oliver en los programas académicos y de prácticas curriculares



Nuria Oliver sobre la IA: “Tenemos que asegurarnos que es una oportunidad que contribuye al progreso y no todo desarrollo tecnológico contribuye al progreso”

PI STUDIO | FOTO: ÁLEX ZEA

valenciaplaza

La alicantina Nuria Oliver, Premio Nacional de Matemáticas por sus contribuciones innovadoras a la IA



INFORMACIÓN

+ La Fundación ELLIS de Alicante participa en una investigación europea de IA a gran escala

Nuria Oliver destaca la oportunidad de que Europa pueda competir con las grandes tecnológicas “poniendo en el centro el bienestar de las personas y el planeta”



Real
Academia
de Ingeniería

Nuria Oliver Premio Talgo a la Excelencia Profesional de la Mujer en la Ingeniería y Premio Nacional de Investigación 2025 Viernes, 17 octubre, 2025



EL PAÍS | 50 Opinión

YANN LECUN | NURIA OLIVER | BERNHARD SCHÖLKOPF

10 FEB 2025 - 12:10 CET

TRIBUNA > 1

E La inteligencia artificial y el futuro de Europa: un momento decisivo

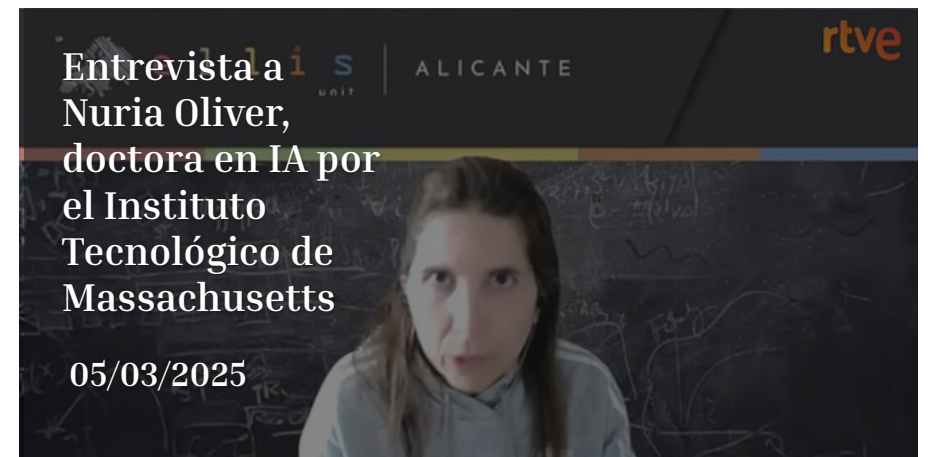
La clave para la innovación está en identificar y atraer a las mentes más brillantes brindándoles autonomía y apoyo financiero a largo plazo

Top television coverage

rtve noticias



¿Cómo afecta la IA a nuestro trabajo? | Entrevista con Nuria Oliver, cofundadora de ELLIS Alicante fundación **Hermes**



rtve play

Top radio coverage:



CON ELLAS. MUJERES QUE CUENTAN

Núria Oliver Ramírez: Experta en Inteligencia Artificial

26 de enero de 2025 00:10:52

Núria Oliver Ramírez es una de las mayores expertas en Inteligencia Artificial de España. Sus artículos han sido citados más de 24.000 veces. Trabaja siempre en proyectos que puedan ayudar a las personas por eso sus conclusiones están presentes en muchas de las tecnologías que utilizamos a diario.

[IR AL PROGRAMA](#)

[● ESCUCHAR RADIO 5](#)

[+ DIRECTOS](#)



radio 5



A GOLPE DE BIT

Nuria Oliver, la investigadora española pionera en IA

10 de septiembre de 2025 00:30:06

Nos acompaña una figura clave en la inteligencia artificial en Europa, la **Dra. Nuria Oliver**, ingeniera de telecomunicaciones y doctora en IA por el MIT, reconocida por su investigación en **comportamiento humano a partir de datos**, también es directora de la Fundación ELLIS Alicante y cofundadora de la red europea ELLIS, que agrupa más de 30 centros de excelencia en IA en 14 países, con un fuerte enfoque en investigación ética y aplicaciones con impacto social positivo. Con ella charlamos sobre el **uso que se le da a la IA en la actualidad**. Además, descubrimos el **desarrollo de un tinte fotocrómico que transforma la luz en electricidad** sin comprometer la transparencia del material.

[IR AL PROGRAMA](#)

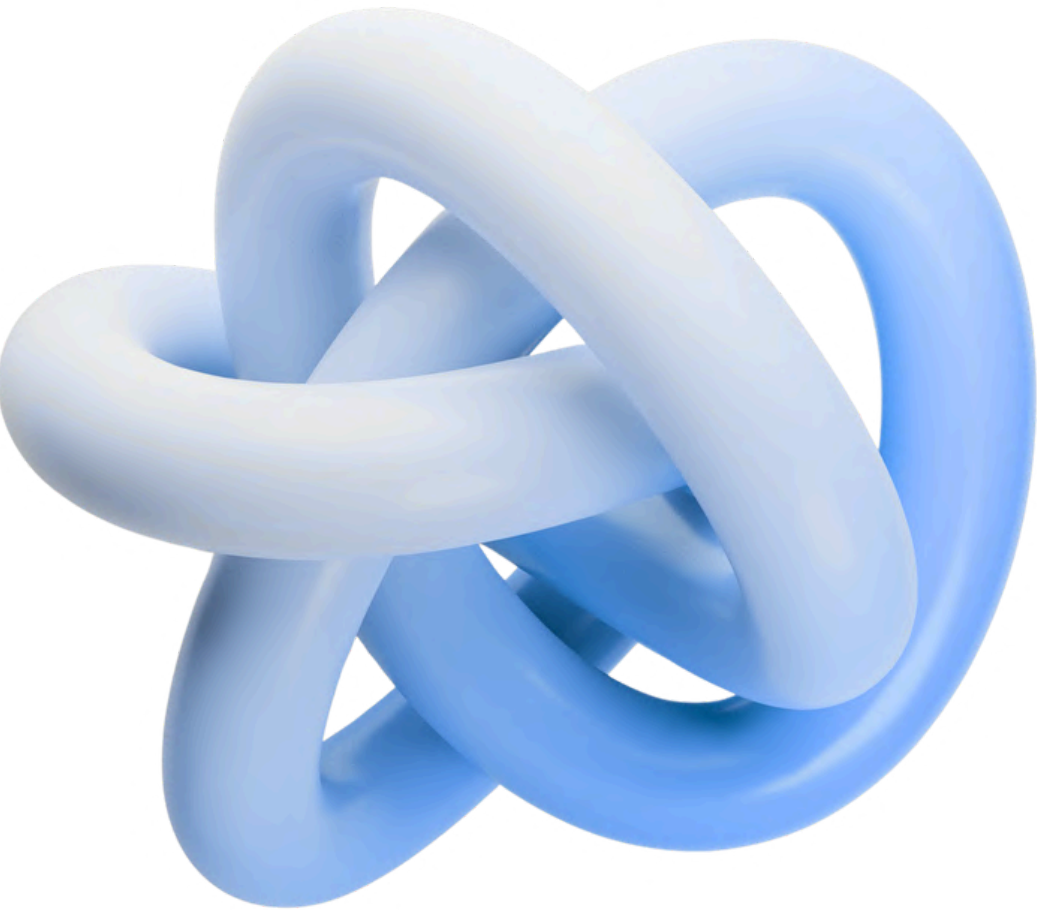
[● ESCUCHAR RADIO EXTERIOR](#)

[+ DIRECTOS](#)



radio e

External relations,
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 organizational and management committees.

Membership of the ELLIS Europe Board

Firstly, through Dr Oliver's participation in the governing body of ELLIS Europe as Vice-President. This role involves taking part in weekly working meetings to plan and organise the activities and initiatives launched by ELLIS Europe across its different pillars: ELLIS research programs, the ELLIS PhD Program, ELLIS units and industrial partnerships.

Human-centric Machine Learning ELLIS Research Program

Secondly, ELLIS Alicante actively contributes to the ELLIS research program "Human-Centric Machine Learning" (HCML), as Dr Oliver co-directs the program alongside Professor Plamen Angelov of Lancaster University in the United Kingdom and Prof. Adrian Weller of the University of Cambridge and The Alan Turing Institute in the United Kingdom.

Strengthening Europe's Leadership in AI through Research Excellence

ELLIS PhD / Postdoctoral Program

Thirdly, ELLIS Alicante actively participates in the ELLIS PhD and Postdoctoral Program. In 2025, ELLIS Alicante was the primary institution for eight students enrolled in the program and the second. The ELLIS PhD Program aims to foster synergies among its students and between students and researchers who are members of the ELLIS network. This is an essential part of its objective to build a highly collaborative and interconnected research community in Europe. ary institution for two further students in this prestigious doctoral program. All students in the ELLIS PhD Program are co-supervised and affiliated with both a primary and a secondary institution, where they undertake a research stay of at least six months.

The ELLIS PhD Program aims to foster synergies among its students and between students and researchers who are members of the ELLIS network. This is an essential part of its objective to build a highly collaborative and interconnected research community in Europe.



Within the ELLIS Postdoctoral Program, ELLIS Alicante is the primary institution for one postdoctoral researcher and the secondary institution for another.

Universidad de Alicante

ELLIS Alicante maintains a close collaboration with the Universidad de Alicante (UA) in the fields of Artificial Intelligence research, education and knowledge transfer. This relationship is particularly significant through the Unidad Mixta de Investigación UA-ELLIS Alicante.

The collaboration with the University of Alicante is further strengthened through initiatives such as the National Artificial Intelligence Strategy Chair (ENIA), as well as through doctoral training and the attraction of research talent. Within this framework, ELLIS Alicante's PhD researchers pursue their training in close connection with the University of Alicante and the international ELLIS PhD Program, fostering joint supervision, international mobility, and interdisciplinary scientific training of the highest standard.

ELLIS Alicante also hosts interns from the Universidad de Alicante, contributing to their practical training in an outstanding international, multisectoral and collaborative scientific environment.



Interaction between students and researchers encourages students to consider a scientific research career as a future professional path accessible to talented individuals from the region.

The partnership between the two institutions has received formal recognition through the Premio Implicación Universitaria 2022 awarded by the Consejo Social de la Universidad de Alicante, in recognition of ELLIS Alicante's commitment to and collaborative relationship with the UA, as well as through the Galardón San Alberto Magno 2025 awarded by the Facultad de Ciencias de la UA, which recognises institutions that have made an outstanding contribution to its academic programs and activities.



Universitat d'Alacant
Universidad de Alicante

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.

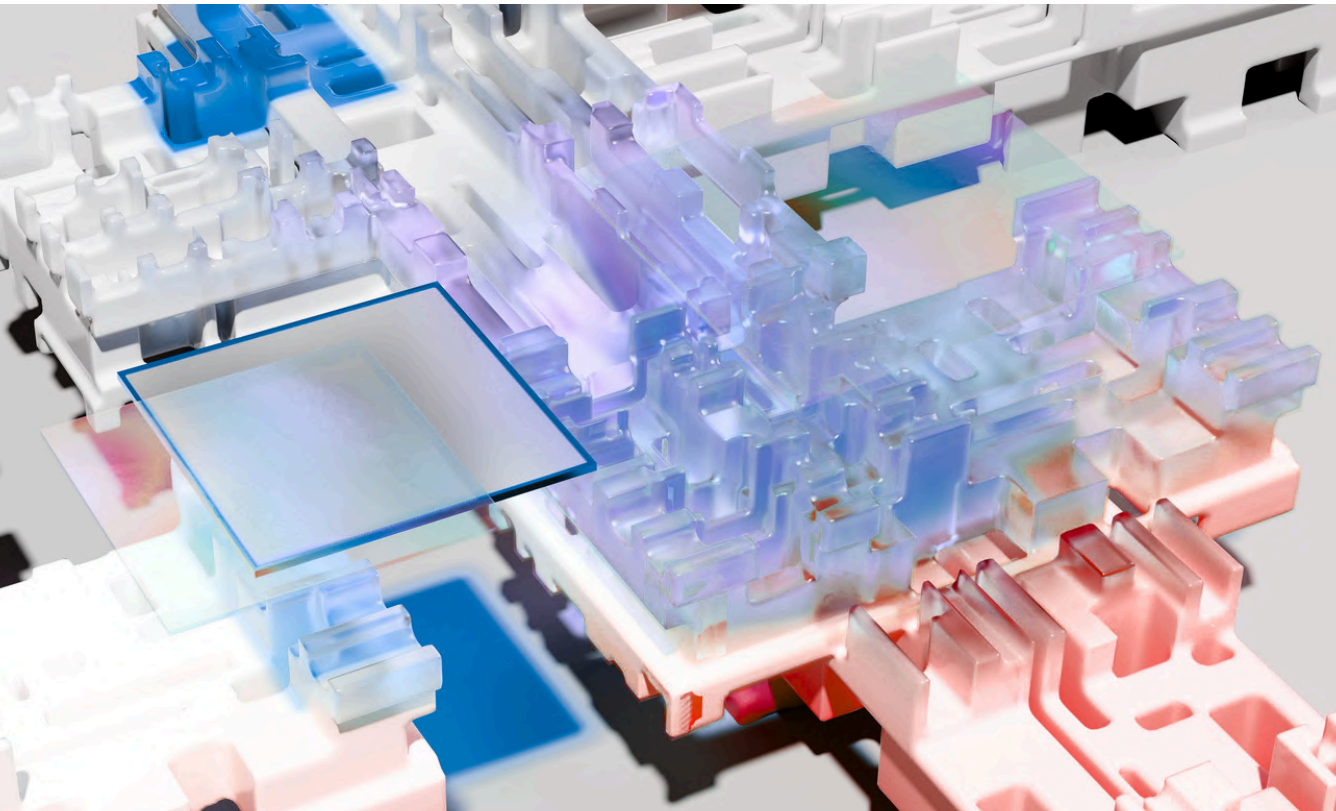
By 2030, AI will measurably influence and impact over 8.5 billion people, across all sectors and various human and terrestrial ecosystems on an unprecedented scale.

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 2025, ELLIS Alicante has continued its membership in the NAIXUS Network.





European Lighthouse of AI for Sustainability

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.



**Funded by
the European Union**

ELLIOT

European Large Open Multi Modal Foundation Models for Robust Generalization on Arbitrary Data Streams

ELLIOT - European Large Open Multi Modal Foundation Models for Robust Generalization on Arbitrary Data Streams

– is a four-year Horizon Europe project with a total budget of €28 million. It brings together 34 leading organizations from 13 countries to develop trustworthy, general-purpose Artificial Intelligence (AI) models.

The project began in July 2025 and is expected to run until June 2029. It is coordinated by the Information Technologies Institute – Centre for Research and Technology Hellas (CERTH-ITI) in Greece.

The project aims to advance a new generation of open-source Multimodal Foundation Models capable of robust generalisation across a wide range of real-world scenarios. These models will be trained to learn from different types of data, including video, text, images, audio, 3D data, sensor streams and satellite data, as well as to adapt effectively to dynamic and previously unseen conditions.

ELLIOT will also support the development of reproducible tools, new standards and benchmark datasets to promote transparent, safe and ethical AI. In addition, the project will address legal, social and regulatory aspects related to Artificial Intelligence, fostering collaboration within the European research community through innovation networks and shared experimentation spaces. ELLIS Alicante contributes research on the ethical aspects, biases and safety of these models.



**Funded by
the European Union**

ELSA

European Lighthouse on Secure and Safe AI

ELSA – European Lighthouse on Secure and Safe AI – is a Horizon Europe project that aims to create a virtual centre of excellence for research into secure, safe and trustworthy Artificial Intelligence in Europe. The initiative seeks to strengthen European leadership in the development of AI technologies aligned with the principles of security, privacy and human oversight. The project runs for 48 months, from September 2022 to August 2026, and has a total budget of approximately €7.4 million.

The project focuses its research activities on key areas related to secure, safe and trustworthy AI, including the technical robustness and security of AI systems,

privacy-preserving techniques and technological infrastructures, as well as human agency and oversight. Through these lines of work, ELSA aims to contribute to the development of AI systems that are more transparent, secure and aligned with European values.

ELSA is coordinated by the CISPA Helmholtz Center for Information Security in Germany and brings together a broad network of European research institutions. The consortium comprises 26 founding members and 41 network members, fostering interdisciplinary collaboration and knowledge exchange in the field of secure, safe and trustworthy AI.

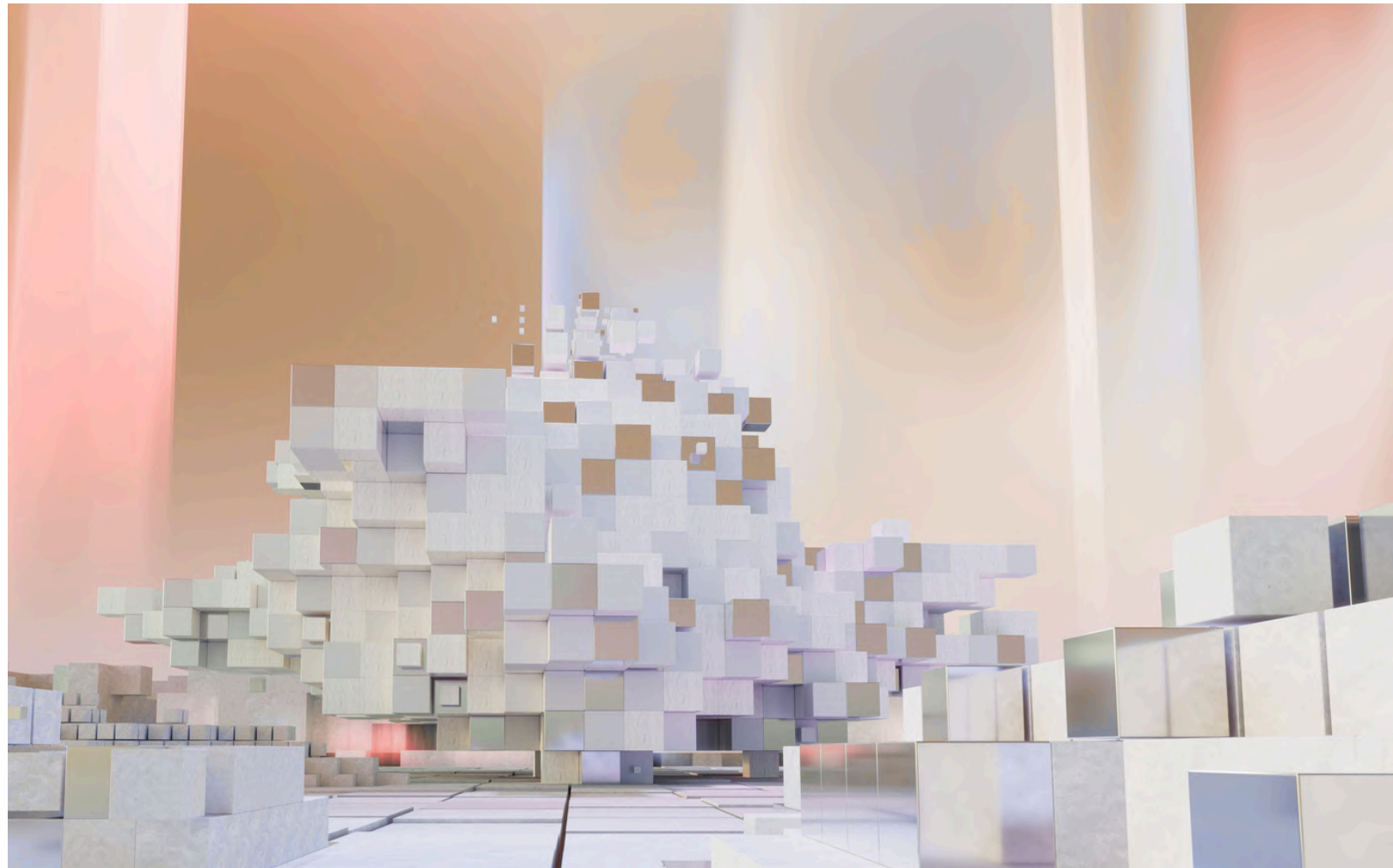
ELLIS Alicante is part of the ELSA network as an associate member, contributing to the scientific exchange and international cooperation activities promoted by the project. As an associate member, ELLIS Alicante does not receive direct funding under this initiative.



**Funded by
the European Union**

Collaborating organizations

During 2025, ELLIS Alicante maintained existing collaborations and established new partnerships and alliances with researchers affiliated with the following academic, scientific and corporate organizations.



Scientific Collaborations

Scientific Advisory Board

Alan Turing Institute
Consejo Superior de Investigaciones Científicas
Cornell University
European Commission
Fondazione Bruno Kessler
Max Planck Institute
MIT Connection Science
Universitat Pompeu Fabra

Secondary ELLIS Unit

Czech Institute of Informatics, Robotics, and Cybernetics
ETH Zurich
Bruno Kessler Foundation (FBK)
Universidad Johannes Kepler Linz
The Max Planck Institute For Software Systems
University of Tübingen
University of Sussex

Visiting Researches Home Institutions

University of Amsterdam
University of Oulu
Université Libre de Bruxelles
The University of Osaka
Utrecht University
University of Stuttgart
University of Dundee
Universidad de Murcia
Tampere University
Assessment, Policy Development and Evaluation Unit of
Public Health Seattle King County
Czech Institute of Informatics, Robotics, and Cybernetics

Interns Home Institutions

Universidad de Alicante
Universitat Rovira i Virgili
Universidad Alfonso X el Sabio
Universitat Autònoma de Barcelona

Institutional Relationships

Collaboration Agreements

Universidad Miguel Hernández
Universidad CEU San Pablo
Universidad de Alicante
Universidad internacional de Valencia
Universidad Rovira i Virgili
Centro de formación profesional Alfonso X el Sabio
Universidad Autónoma de Barcelona
Centro superior de idiomas de la Universidad de Alicante
Universita degli Studi di Trento
Fundación Gestamp
Fundación Centro Internacional De Investigación En Inteligencia Artificial (Nair Center)
Fundación Instituto Hermes Para El Empoderamiento De La Ciudadanía
Consejo de Fundaciones por la Ciencia del FECYT
Fundación Banco Sabadell
CaixaBank S.A
Nippon Sanso
Angela Impact Economy

Memberships

Fundaciones por el Clima
Registro de Entidades Innovadoras en la Red Privada de Blockchain APTE
Plataforma Tecnológica Española de Tecnologías Disruptivas
Distrito Digital Comunitat Valenciana
NAiXUS



Research

ELIAS Consortium Members

University of Trento (UNITN)
University of Amsterdam (UvA)
Polytechnic University of Milan (POLIMI)
University of Modena and Reggio Emilia (UMORE)
Polytechnic Institute of Paris (IPP)
University of Tübingen (UT)
Fondazione Bruno Kessler (FBK)
Centre of Research & Technology Hellas (CERTH)
University of Valencia (UVEG)
Hasso Plattner Institute (HPI)
University of Milan (UNIMI)
University of Copenhagen (UCPH)
Jožef Stefan Institute (JSI)
Czech Technical University (CTU)
University of Toulouse (ANITI)
Max-Planck Institute (MPG)
Umeå University (UMU)
Istituto Italiano di Tecnologia (IIT)
Politehnica Bucureşti (UNSTPB)

IDEAS NCBR (IDEAS)
Bitdefender (BD)
IBM (IBM)
INRIA (INRIA)
ROBERT BOSCH GMBH (RB)
ROBERT BOSCH KFT (RBHU)
Prometeia (PRM)
Toulouse School of Economics (TSE)

ELIAS Affiliated entities

Institut Mines-Télécom (IMT-TP)
École Polytechnique (EP)
Università di Genova (UNIGE)

ELIAS Associated partners

Idiap Research Institute (IDIAP)
Eidgenössische Technische Hochschule Zürich (ETH)
Aalto University (AALTO)
University of Manchester (UMAN)

Research

ELLIOT Consortium Members

Centre for Research and Technology Hellas - ITI (Information Technologies Institute (CERTH))
University of Tübingen (UT)
Forschungszentrum Jülich (FZJ)
University of Amsterdam (UvA)
Eindhoven University of Technology (TU/e)
University of Trento (UNITN)
Computer Vision Center (CVC)
Jožef Stefan Institute (JSI)
LOBA
Barcelona Supercomputing Center (BSC)
IT Center for Science (CSC)
CINECA
Ludwig-Maximilians-Universität München (LMU)
University of Valencia (UVEG)
University of Modena and Reggio Emilia (UNIMORE)
Aalto University (AALTO)
ELLIS Institute Tübingen (ELLIST)

Czech Technical University in Prague (CTU)
CISPA Helmholtz Center for Information Security (CISPA)
KUL Centre for IT & IP Law (CiTIP)
Voxist
Valeo
RoboTwin
Openchip
Indra Deimos
Generalitat de Catalunya: Departament de Territori, Habitatge i Transformació Ecològica (GENCAT)
VRT
École Polytechnique Fédérale de Lausanne (EPFL)
ETH Zurich
Prior Labs
NXAI
Letxbe AI
Goethe University Frankfurt

Scientific Publications

Universitat d'Alacant, Spain

École Polytechnique Fédérale de Lausanne (EPFL),
Switzerland

Department of Computer Science and Artificial Intelligence,
University of Sussex, United Kingdom

Department of Informatics, Basque Center for Applied
Mathematics, Spain

University of Amsterdam, Netherlands

Johannes Kepler University Linz, Austria

The University of Osaka, Japan

University of Notre Dame, Notre Dame, United States

Weizenbaum Institute/Universität der Künste Berlin, Germany

Université de Mons (UMONS), ISIA Lab, Belgium

Université Libre de Bruxelles (ULB), IRIDIA Lab, Belgium

University of Tübingen, Germany

ELLIS Unit, LIT AI Lab, Institute for Machine Learning, JKU Linz,
Austria

NXAI GmbH, Linz, Austria

Universidad Alfonso X el Sabio, Spain

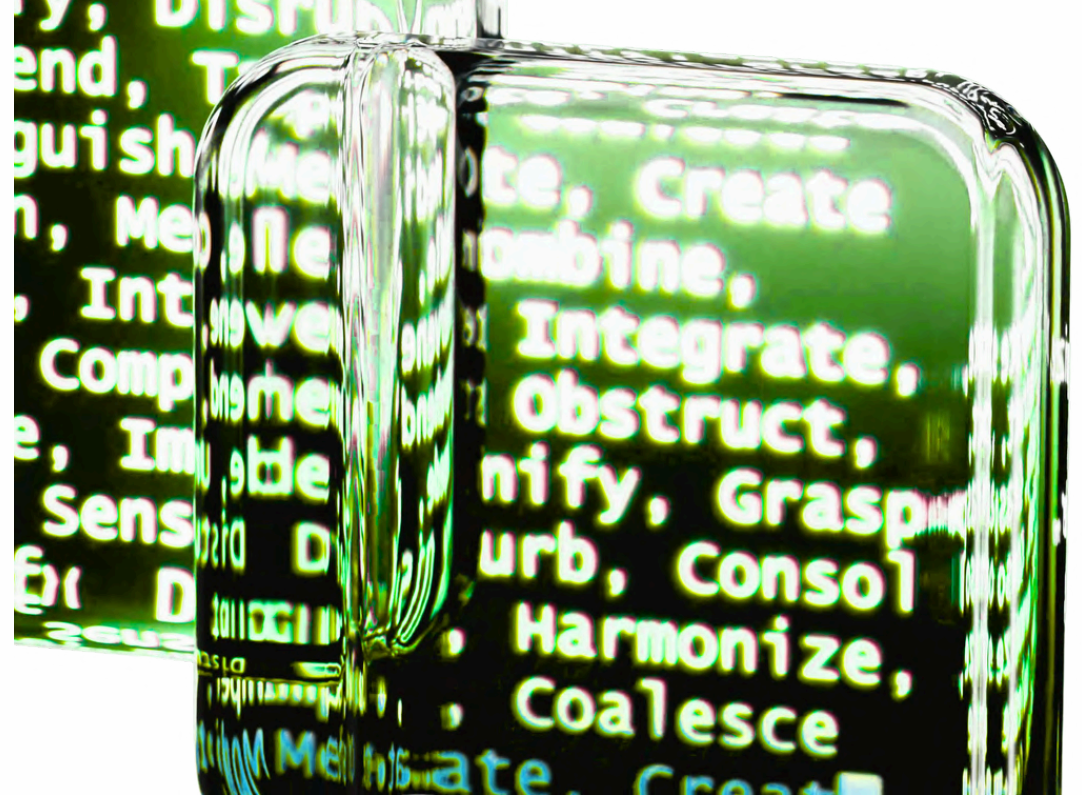
University of Stuttgart, Germany

Universitat Autònoma de Barcelona, Spain

NEC Laboratories Europe, Germany

Max Planck Institute for Software Systems, Germany

CTU in Prague, Czech Republic



School of Health Science, The University of Nottingham, United
Kingdom

School of Psychology, Public Health and Social Care, University
of Derby, United Kingdom

School of Computer Science & School of Medicine, The
University of Nottingham, United Kingdom

Polytechnic University of Turin, Italy

Fondazione Bruno Kessler, Italy

University of Trento, Italy

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.

Law 6/2025, of 30 May, on the Budget of the Generalitat for the 2025 financial year (DOGV No. 10120, of 31 May 2025) and within budget section G0111 (Ministry of Innovation, Industry, Trade and Tourism), Service G011102 (Regional Secretariat for Innovation), Managing Centre G01110201 (Directorate General for Innovation), budget subprogram 121H00 – Advancement of the Digital Society, grant line S1600, endowed with an amount of €370,000 and intended for the grant “Support for the operational expenses of the ELLIS Alicante Node”, and grant line S1233 with an amount of €10,000, intended for the grant “Support for excellence research in Artificial Intelligence at the ELLIS Node”, during the 2025 financial year.

The funding obligations associated with the grant provided under the aforementioned budget line were governed by a resolution signed on 11 November 2025 between the Administration of the Generalitat Valenciana, through the Ministry of Innovation, Industry, Trade and Tourism, and ELLIS Alicante, entitled “Resolution of the Minister of Innovation, Industry, Trade and Tourism and the Fundació de la Comunitat Valenciana Unidad ELLIS Alicante regulating and granting a direct subsidy to the Fundació de la Comunitat Valenciana Unidad ELLIS Alicante to support excellence research in Artificial Intelligence, and the operational expenses of ELLIS Alicante during the 2025 financial year.



EUROPEAN UNION
European Regional Development Fund

Private Institutions

To help the ELLIS Alicante fulfil its mission, companies and other partner institutions can contribute and donate through a range of channels aligned with the ELLIS Europe sponsorship program.

We are immensely grateful for the generous contributions made by our donors and sponsors. Their support for ELLIS Alicante's mission, vision and activities is instrumental to the success of our organization.

**SUPPORT RESPONSIBLE, HUMAN-CENTRIC AI RESEARCH
FOR SOCIAL GOOD.**



SUPPORT US

GOLD



SILVER



FRIENDS OF ELLIS ALICANTE



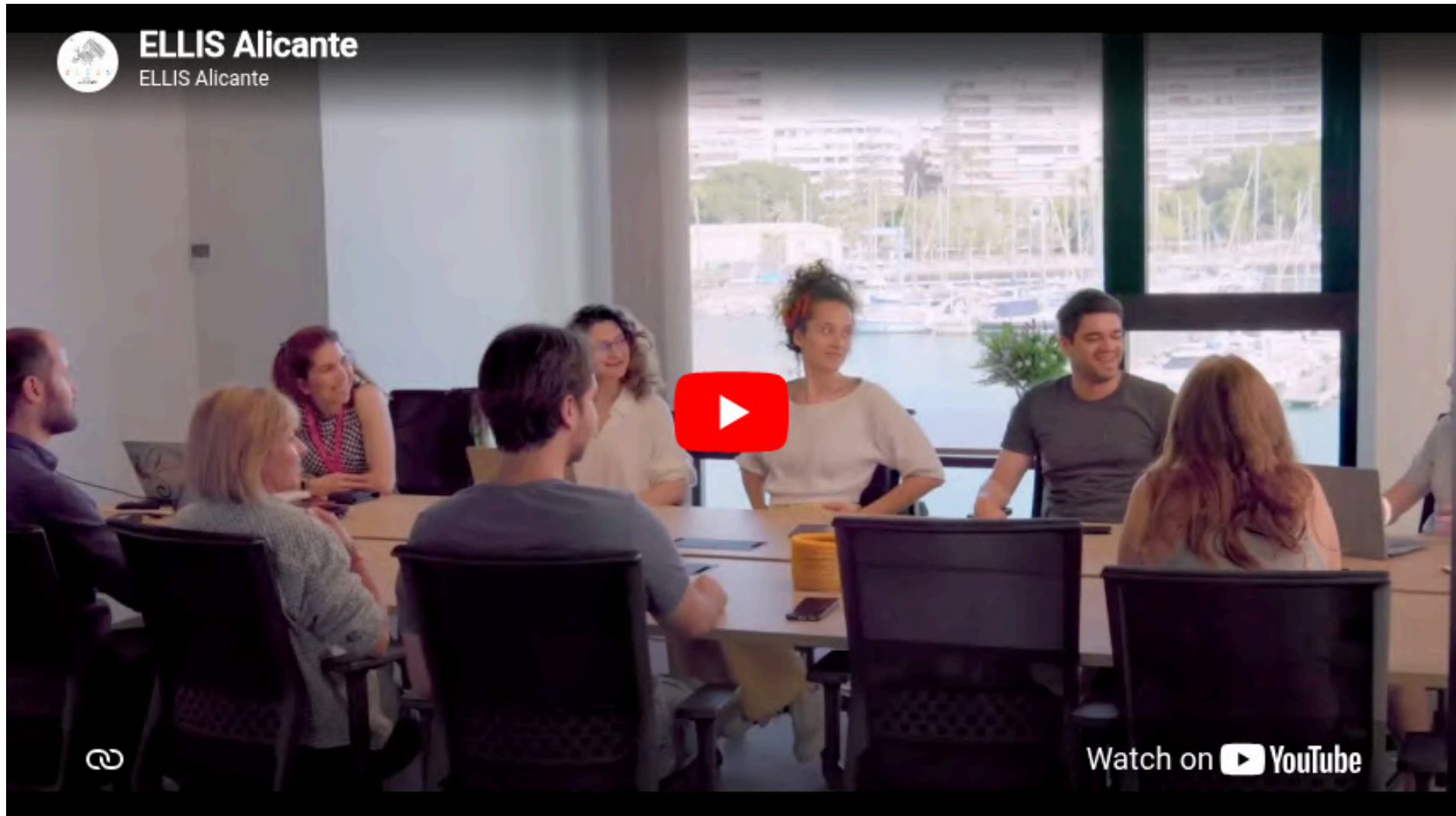
MICRO-FINANCING



Partner with us

Level	Description and benefits
PLATINUM PARTNERSHIP €500K first year, €200K following years	HIGHEST LEVEL OF STRATEGIC ENGAGEMENT (MINIMUM 2 YEARS) Comprehensive partnership with maximum visibility and influence in shaping research directions. Includes a board seat option, unlimited site visits, logo in all communications, newsletter announcements and priority access to training programs
STRATEGIC RESEARCH PARTNERSHIPS (GOLD LEVEL) €100K–€500K+ annually	MULTI-YEAR COMMITMENTS SUPPORTING RESEARCH PROGRAMS Funding for research teams working on defined focus areas—from fair AI systems to privacy-preserving methods to transparency tools. Includes regular briefings, advisory participation, and prominent recognition.
PHD EXCELLENCE FELLOWSHIPS (SILVER LEVEL) €200K over 4 years	CULTIVATING THE NEXT GENERATION OF AI LEADERS Fund doctoral researchers conducting cutting-edge research while being trained by world-class scientists. Includes regular briefings, advisory participation, and prominent recognition
PROJECT-SPECIFIC FUNDING (SILVER/GOLD LEVEL) ≥€100K	TARGETED SUPPORT FOR SPECIFIC RESEARCH QUESTIONS Fund investigations of particular challenges with defined objectives and deliverables.
BRONZE ≥€25,000	FUND SUBSTANTIAL RESEARCH PROGRAMS All the benefits available to Friends of ELLIS Alicante, plus a prominent profile on our dedicated sponsors page, direct engagement with researchers, the opportunity to co-sponsor doctoral research and special recognition at events.
FRIENDS OF ELLIS ALICANTE €5,000 - €25,000	ENABLE SIGNIFICANT RESEARCH CONTRIBUTIONS All the benefits of Micro-funding, plus prominent recognition on our website, quarterly research briefings, the opportunity to visit our premises once a year and early access to publications.
MICRO-FINANCE <€5,000	HELP FUND ONGOING RESEARCH PROJECTS Name listed on our website, recognition in our annual report, six-monthly research updates, a donation certificate for tax-relief purposes and invitations to events.

Learn more about ELLIS Alicante



Thank you for making
ELLIS Alicante
possible



UNIT
ALICANTE

Fundación de la Comunitat Valenciana
Unidad de ELLIS Alicante
Distrito Digital 5 - Edificio A
Muelle de Poniente, 5
Puerto de Alicante
03001 Alicante - España

+34 965 641 175
www.ellisalicante.org
info@ellisalicante.org