ML Labs

SFI Centre for Research Training in Machine Learning

Centre for Research Training

Science Foundation Ireland For what's next

NEWSLETTER

DECEMBER 2021

HOST INSTITUTIONS

DCU

TU Dublin

UCD Dublin

University College Dublin
An Coláiste Ollscoile, Baile Átha Cliath

FUNDED BY:

Science Foundation Ireland For what's next

MISSION

The mission of the Science Foundation Ireland Centre for Research Training in Machine Learning is to train industry-ready, academically excellent, socially responsible graduates who will lead the current and future transformation of industry, society, and science that machine learning is enabling.

VISION

The Science Foundation Ireland Centre for Research Training in Machine Learning is an internationally connected and globally recognised PhD training centre specialising in Machine Learning and designed to address the urgent demand for machine learning talent. PhD candidates benefit from a world-class, inter-institutional programme in a mature interdisciplinary environment that emphasises cutting-edge research, social responsibility, and industry-relevant and entrepreneurial focus.

CONTENTS

Welcome from the Director 4

The Centre in Numbers 5

Meet the Centre Mascot 6

Impact Highlights 7

Machine Learning & AI in the Classroom 7

Networks & Partnerships 9

Interviews with Industries 9

Innovation in AI at Shutterstock 10

ML Meets Quantum Computing at Equal1 Labs 11

A Partnership with Mastercard 12

International Engagement 13

A Collaboration with the Center for Magnetic Resonance in Biology and Medicine 13

UNEP - United Nations Environment Programme 14

Public Engagement 15

The Dangers of Extreme Weight Cutting for MMA Fighters | RTE Brainstorm 15

PhD & Productivity - YouTube Channel 16

The Bloggers: Mehreen Tahir, Gargi Gupta 17

Cohort-based PhD Programmes in Ireland 18

Living with AI: Should Machines Make Decisions For Us 19

Competitions 21

"Want to Save a Rainforest? Mix AI & Blockchain" 21

An Attack on CNNs 22

Detecting Anomalies in Time Series Data 23

Generate Images from Text 24

Research Articles 25

Other News 29

WELCOME FROM THE CENTRE DIRECTOR

Our mission at the Science Foundation Ireland Centre for Research Training in Machine Learning is to train industry-ready, academically excellent graduates who will lead the current and future transformation of industry, society, and science that machine learning is enabling. It is now a little over two years since we welcomed our first cohort of PhD candidates to our three institutions (UCD, DCU and TU Dublin). Since then we have added two more cohorts to bring the total number of PhD candidates who have joined the centre to almost 80.

The projects that our PhD candidates are working on range from inventing fundamentally new machine learning algorithms, to applying machine learning techniques to new industries, to investigating the impacts machine learning is having on society. Highlights of the impact that this work is having are described throughout this newsletter and include a machine learning and AI training course delivered to over 6,000 secondary school students, the development of new algorithms for quantum-enabled machine learning, and a system that uses machine learning to detect and measure potholes on Irish roads.

2021 was the first year in which our PhD candidates started to complete placements with our industry partners where our PhD candidates augment their academic experience with practical experience of building machine learning solutions. So far we have been delighted to partner with Irish organisations (e.g. PMS, KantanMT, and RCSI), large multinationals (e.g. Eaton, Huawei, Mastercard, Optum, and Shutterstock) and exciting startups (e.g. AimHub, Biosimulytics, BeWell, and The Elephant Hive) on placements. We are also delighted to have ongoing strategic research engagements in place with Colgate-Palmolive, Microsoft, Equal1, and Huawei.

Just like most other organisations the centre has been impacted by the Covid-19 pandemic. Over the last 12 months most of our activities continued to be online, and our PhD candidates, operations staff, and academic supervisors continued to work primarily from home. Through everybody's resilience, however, the wheels have kept on turning and centre activities have continued successfully over the last 12 tricky months, with many of our PhD candidates, staff and supervisors also making significant contributions to the pandemic response.

For us the end of the year coincides with the end of Bootcamp for our 2021 cohort of PhD candidates. Bootcamp involved almost 10 weeks of industry talks, workshops, expert seminars, group activities, social events, and development projects. The end of Bootcamp is full of optimism and excitement for the research work that our PhD candidates are beginning, and we are also looking forward to another year at the centre with optimism and excitement. We hope that you enjoy reading about the work that we have been doing at the centre in the rest of this newsletter, and we look forward to working with you in the new year. Stay in touch, Merry Christmas, and Happy New Year.

2021 ML-LABS STATS

79 Students

26 Workshops

46 Student Publications

19 Keynote Talks

30 Industry Seminars

20 Industry Partnerships

167 Public Engagement Activities

MEET THE CENTRE MASCOT

As part of the ML-Labs Summer School 2021 the cohort was given the challenge to create a mascot that best represents the Centre. **Adam Stapleton, Bashayar Al Mulhaini, Mehreen Tahir, Laoibhse Nifhaolain, Long Mai, Patrick Mc Enroe** are the masterminds that ideated Imeall, the anthropomorphic companion to our fine research centre. Here the winning group answers some questions about Imeall the Octopus, the lovable cephalopod chosen as the winner!

How does an octopus represent the traits of an ML-Labs Researcher?

Octopi are among the most intelligent of all invertebrates (and have even been observed using tools). This represents the need for focus and analytical strength. They are among the most behaviourally diverse of all the invertebrates. This represents the wonderful diversity found in ML-Labs. They can radically change shape and have an ability to camouflage, reflecting the adaptability required of Machine Learning (ML) researchers. They are good at expressing emotions – a key part of the communication skills being developed in our PhD programme. And lastly they have been variously described as chirpy, curious and inquisitive. ML-Labs researchers have been described in similar ways!

Why is it called Imeall?

In ML-Labs, we do cutting edge research and Imeall is an Irish word meaning "edge". Click on this link to hear the correct pronunciation.

The keen eared among us also spotted the connection between the sound of "Imeall" and the sound of "ML" and thought it was too good to be true! Several members of the cohort are also working on edge AI, a discipline within AI that looks at distributed computation to better utilise the resource of individual devices "on the edge" of the system rather than waste traffic communicating back and forth with a central computing resource. Interestingly enough, the brain of an octopus bears this same characteristic – their ganglia (or little brain processing units) are distributed in their tentacles as an extension of their central brain, allowing them to move their tentacles quickly in response to their environment and allowing the main brain to get on with deep octopus thoughts.

What considerations went into the design?

We wanted Imeall to capture the spirit of ML-Labs and the ML-Labs students – smart and inquisitive researchers gladly embarking on the challenge of completing a PhD (and hopefully with a smile on their faces). Octopi don't just have fat heads – they've got the brains to match! We gave Imeall the ML-Labs badge and a shamrock to represent the 3 universities involved in the research centre. As an icon, an octopus could be said to look quite like the neural network diagrams many of us are getting very familiar with. Much of cutting edge ML research makes use of neural networks. Finally with Imeall's eight tentacles we think he would be equally adept at juggling the many tasks at hand facing a budding ML researcher.

IMPACT HIGHLIGHTS

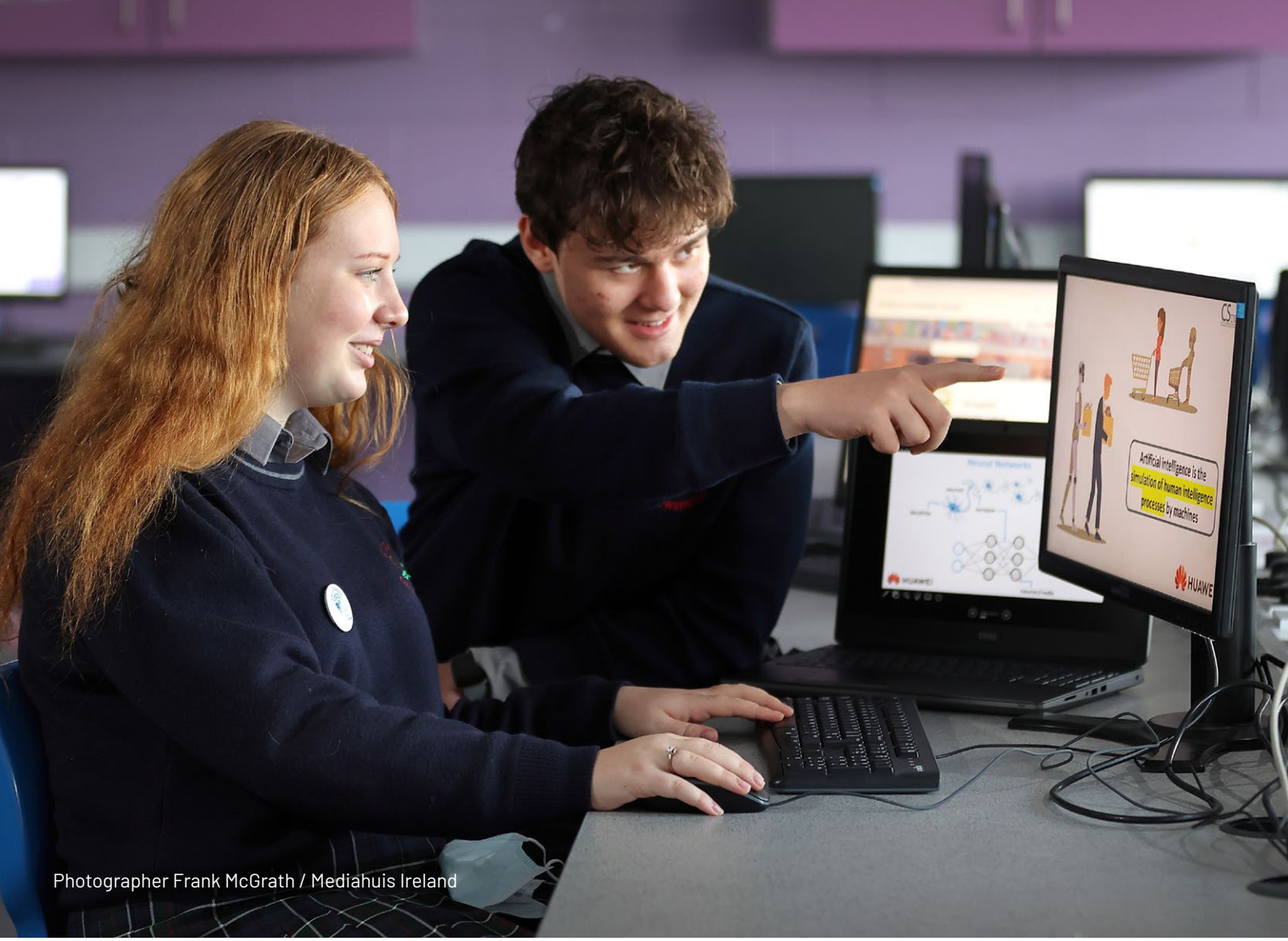
MACHINE LEARNING & AI IN THE CLASSROOM

It is clear that the pandemic has had a dramatic impact on education, which for many has meant an unplanned and rapid move to online and blended learning approaches. In the Summer of 2021 a 'Machine Learning & AI' module was developed by PhD student **Joyce Mahon** at UCD in a collaboration between the SFI CRT in Machine Learning and industry partner Huawei. Joyce is supervised by **Dr. Brett Becker** and **Dr Brian Mac Namee** of the UCD School of Computer Science; and for the duration of this project worked alongside **Dr. Keith Quille** of TU Dublin, and with student volunteers.

'Machine Learning & AI' module was added to the CS_LINC platform developed in 2020 by the CS_INC team in TU Dublin. CS_LINC provides formal computer science curricula through free and easily accessible online modules. As of November 2021, there are over 100 schools registered with approximately 10,000 students enrolled on CS_LINC. 6,973 students from 71 schools and 1 Youthreach Centre are currently registered on the Machine Learning & AI module. Three modules were offered when CS_LINC started in 2020 and 3,000 students from 50 schools enrolled.

The Machine Learning & AI module comprises 8 classes, each 40 minutes in length; and contains lesson plans, presentations, voiceover videos, activities, worksheets and solutions. Advanced supplementary Python code is provided alongside instructions if desired; and any mathematical concepts covered are previously taught to students at lower secondary level. The module is structured as follows: (1) Introduction to Artificial Intelligence, (2) Machine Learning & Data, (3) Data Analysis & Pre-processing, (4) Machine Learning Models 1 - Linear & Logistic Regression, (5) Machine Learning Models 2 - Decision Trees & K-Nearest Neighbor, (6) Neural Networks & Deep Learning, (7) Evaluation & Vision, and (8) Test Your Knowledge.

The module requires minimal setup and has no specific software requirements. It aims to demystify some Machine Learning & AI concepts, reduce misconceptions, and provide a clearer pathway for students who wish to pursue computer science at second and/or third level. In addition, it aims to support 'early school leaver' programs, as well as all students who do not have access to formal computer science education. The module provides a taster for students interested in taking Computer Science for the Leaving Certificate while also giving useful and practical knowledge of how Machine learning & AI are used in the tools and apps that students interact with every day. The module was recently featured in the Irish Independent alongside



Photographer Frank McGrath / Mediahuis Ireland

staff and students of St Conleth's Community College, Newbridge (The article is available at this link).

Joyce will continue her research in the area of Machine Learning & AI in Education, investigating the potential industry possibilities that related technologies are generating; and the subsequent ethical impacts that they have upon society. Current and future pre-University students are growing up in a world where these technologies are commonplace, and it is important that they not only understand how they work but are equipped to make informed judgements about how they will use them. Hence, she will work to expose a wider and more diverse audience to tools relating to areas such as analytics, speech recognition, and natural language processing. This will be accomplished by developing content and creating opportunities for teachers and students to understand these tools and become involved in their design and application.

NETWORKS AND PARTNERSHIPS →

NETWORKS AND PARTNERSHIPS

The SFI Centre for Research Training in Machine Learning is constantly working on establishing new partnerships with private and public entities as a way to provide the best opportunities to its PhD students. International, interdisciplinary and intersectoral training are fostered at the centre through collaboration with industries in different sectors of the Irish economy and through collaborative projects with hospitals, institutions of second level education, and government agencies.

The pandemic has brought with it new challenges in the last twelve month, but despite the difficulties associated with remote working the centre has built new relationships with the outcome of forging a new mind-set to Research, Training, and Innovation.

INTERVIEWS WITH INDUSTRIES

The "Interview with Industries" series is led by students at the Centre and is the outcome of an internal initiative to encourage PhD candidates to propose outreach activities that would enhance their networking and communication capabilities as well as to inform the public on different aspects of research and innovation.

Episode.1 - Average Day of a ML Engineer | John Glover, Aylien

Episode.2 - Automated Machine learning Models in our industries | James Conway, Mastercard

Episode.3 - Pearl's of wisdom from the director of AI at Shutterstock | Alessandra Sala, Shutterstock

Episode.4 - Working at the Intersection of Robotics and Artificial Intelligence | Niamh Donnelly, Akara Robotics

Episode.5 - Overcoming Phobia for Math, Journey to Machine Learning | Christophe Guéret, Accenture

INNOVATION IN AI AT SHUTTERSTOCK

"Shutterstock had the pleasure to host one of the PhD students from the SFI Centre for Research Training in Machine Learning which is a unique program to train industry-ready and academically excellent graduates who will lead the current and future innovation in AI.

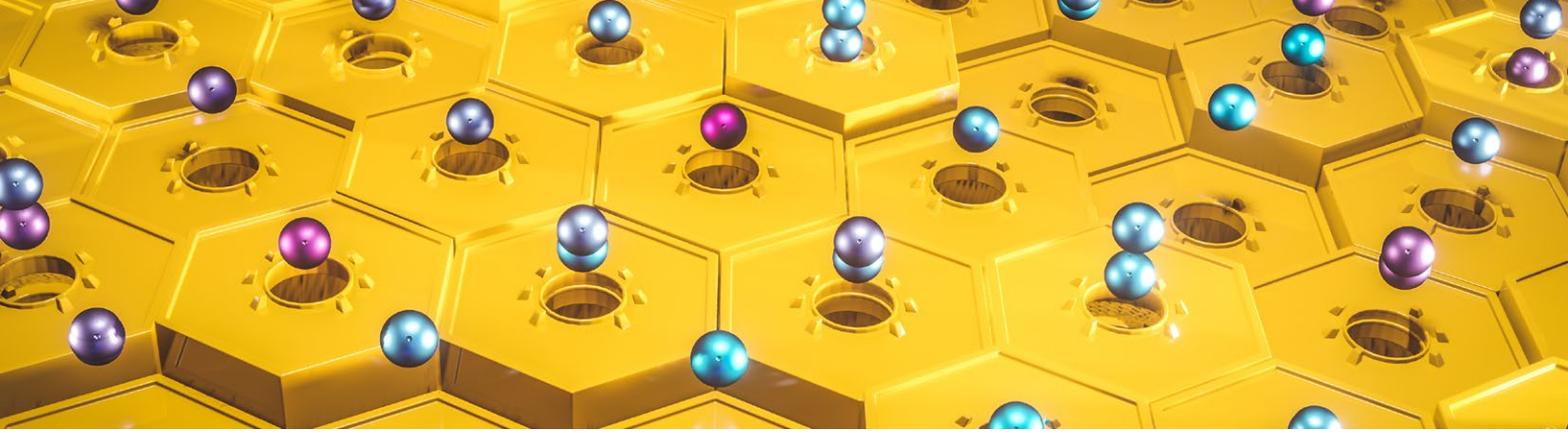
Na Li was the successful student among several other excellent candidates who applied and were interviewed for the placement. We hired Na Li to complete her placement at Shutterstock and to offer her a practical experience in building machine learning solutions within an industrial context.

Na Li worked on a unique idea which has broad applicability to any machine learning project. When working on machine learning products a recurring issue is to build strong quality assurance (QA), quality control and testing capabilities which control the quality of the proposed solution, allow fast development of new ideas, and support improvements of the actual machine learning models. Specifically, the project that Na Li has supported during her 4 months long placement at Shutterstock, addressed the problem of curating sensible benchmark datasets to test the quality of the deployed machine learning models. Quality tests are extremely time consuming but necessary to have a robust development cycle of machine learning models. Visual inspection tools are key to create order of magnitude faster approaches than usual code-based tests.

To this end, Na Li worked on a project to develop an annotation tool for creating specialized datasets for testing internal machine learning projects via a web application. The tool allows how different versions of a machine learning model performs on specific datasets to be inspected in a visual and flexible way. The tool also allows users to compare different versions of machine learning models and to track their changes as new improvements are made. This tool has been proven to speed up the task of curating benchmark datasets and to enable fast inspections of model changes in terms of quality and consistency.

Placements like this one ensure that PhD candidates at the SFI CRT in ML experience different practices and cultures that sharpen their critical thinking and inspire their desire to innovate on cutting-edge technology which also accounts for social responsible and ethical aspects of this fast growing technology space."

Alessandra Sala, director of AI and Data Science at Shutterstock



ML MEETS QUANTUM COMPUTING AT EQUAL1 LABS

Quantum Computing has shown a lot of promise for many domains as it exhibits properties like entanglement and superposition that do not exist in classical scenarios. One area that is expected to profit greatly from this is machine learning. At the SFI CRT in ML, we have partnered with Equal1 labs, a quantum computing start-up, to explore how to leverage the potential of quantum computing for machine learning. This research is led by **Dr Simon Caton** and the PhD student **Patrick Selig**.

When thinking about machine learning using a quantum computer, a lot of the basics stay the same. We still have training data, we still have to train a model, there are parameters to optimise, and we can still derive a loss function (or similar), and evaluate it using a test set. However, things very quickly get very different after this point. The instruction set for quantum programs is fundamentally different, as is the manner with which we represent a quantum program. There are challenges transforming data between a classical and quantum representation for model input and output. Unlike machine learning in classical scenarios, there is a general lack of understanding and guidance concerning the "right" way(s) to design and train quantum machine learning models. If that wasn't enough, modern day quantum computers are still quite error prone. Thus, our work has been looking to derive guiding principles for how to design and train machine learning models using noisy and intermediate scale quantum (NISQ) computers.

In our work, we have explored classification problems described using only a continuous feature vector. Categorical and discrete data are still hard to represent well in a quantum program and having a numerical response (e.g. for regression) also adds complexity in transforming model output from a quantum to classical representation. We build quantum machine learning models using variational principles, aptly referred to as variational quantum machine learning, as there is some evidence that models of this type can withstand certain amounts of error in (quantum) calculations. In this setting, model parameters are classical while the computation of the prediction based on the input, i.e. training process, is performed on the quantum computer. The resulting machine learning model is a parameterised quantum program expressed as a quantum circuit over a number of quantum bits (or qubits). In training, we seek to optimise these parameters.

To design a variational quantum machine learning program for classification, there are two key considerations: how many qubits to use (program width) and "long" (or deep) the program is, i.e. how many quantum operations (or gates) occur along the program's critical path. Both these considerations affect the expressivity and complexity of the function the program represents. While this may, on the surface, appear like a simple 2 dimensional optimisation problem, the design space grows fast, very fast, especially when quantum entanglement operations are leveraged.

To date, we have evaluated over 6500 unique quantum circuits (each representing a candidate machine learning model) using 7 "easy" machine learning datasets. We find that in general shallow (low depth) wide (more qubits) circuit topologies tend to outperform deeper ones. We also explored the implications and effects of different notions of noise (causes of computation error) and have evaluated circuit topologies that are more / less robust to noise for classification machine learning tasks. Based on our findings we have generated a set of guidelines for designing circuit topologies that show near-term promise for the realisation of quantum machine learning algorithms using gate-based NISQ quantum computer.

A PARTNERSHIP WITH MASTERCARD

To answer pressing market needs, the SFI Centre for Research Training in Machine Learning has partnered with Mastercard to prepare a new generation of PhD students with the mix of skills needed to thrive in the modern workforce. Mastercard, through the internship program will host PhD students for four months to work with a team on challenging topics. This collaboration will push the boundaries of industrial Machine Learning applications and the immersive learning experience will give to the students a flavour of working on challenging ideas in an industry environment.

Steven Flinter, Vice-President of AI at Mastercard Labs in Dublin, in an exclusive interview with Silicon Republic explains that the PhD Programme in Machine Learning *".... is going to give us access to top-grade PhD candidates,.... Mastercard can support the training of those PhD students to help them get to a point where they can translate their training and their education into industry once they've graduated."*

The full interview is available on the SiliconRepublic YouTube Channel.

INTERNATIONAL ENGAGEMENT

A COLLABORATION WITH THE CENTER FOR MAGNETIC RESONANCE IN BIOLOGY AND MEDICINE (CRMBM)

In June 2021, **Misgina Tsighe Hagos** and **Niamh Belton**, under the supervision of Prof. Kathleen Curran, were granted Ulysses funding by the Irish Research Council (IRC). Ulysses funding aims to promote research collaborations between Ireland and France-based researchers.

IRISH RESEARCH COUNCIL
An Chomhairle um Thaighde in Éirinn

"We are collaborating with the Center for Magnetic Resonance in Biology and Medicine (CRMBM) research group, based in Aix-Marseille, who are experts in musculoskeletal medical imaging. They have acquired datasets of Magnetic Resonance Imaging (MRI) and Diffusion Tensor Imaging (DTI). Measurements of the muscle architecture can be more accurately calculated in 3D space from DTI. These measurements can provide critical information on injury risk, injury recovery process, and muscle disorders. This presents an opportunity to develop novel machine learning models for automating muscle architecture analysis.

We plan to work with an MRI dataset of marathon runners to assess changes to their muscle architecture. The dataset contains MRIs of the runners before and after the marathon. This is a small dataset and therefore, we will leverage techniques such as data augmentation, transfer learning, and few-shot learning.

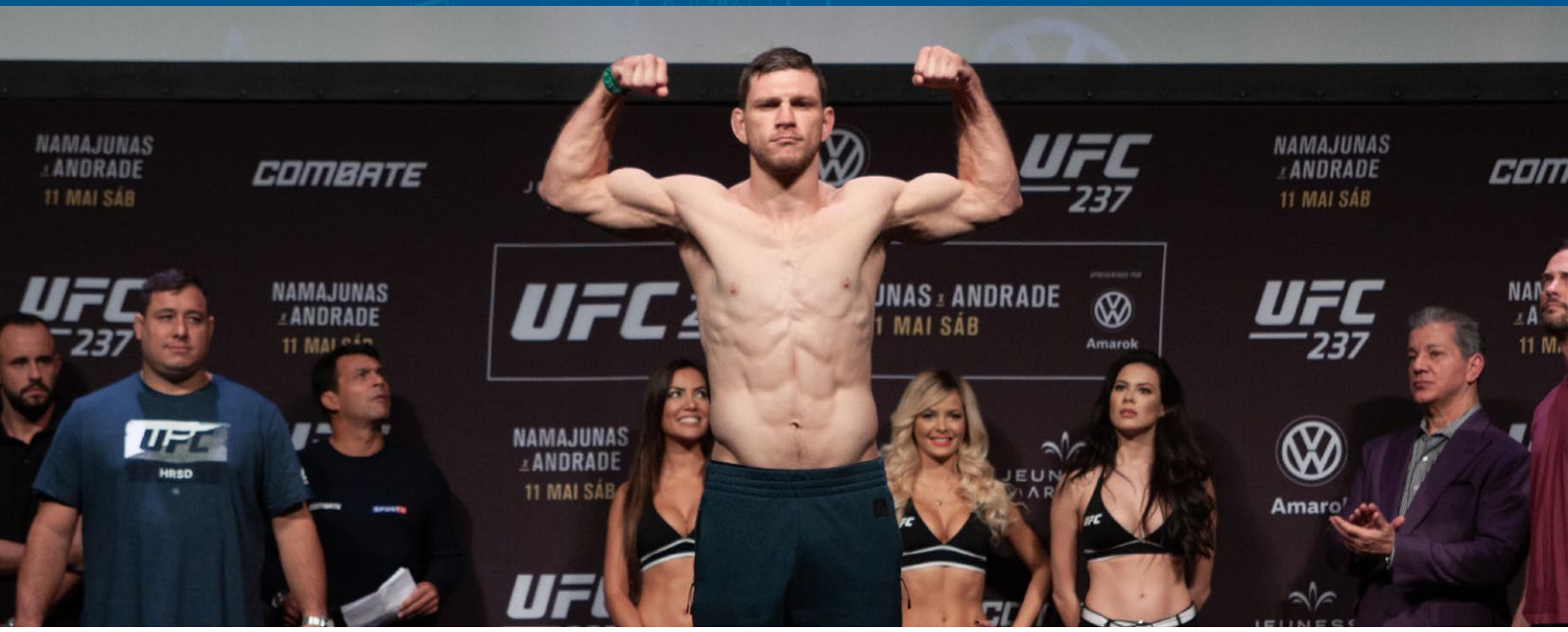
We had the pleasure of hosting Prof. David Bendahan, CNRS research director and leader of the "Magnetic Resonance MSK" group at CRMBM, on a research visit for three days where we exchanged research ideas. He also gave an interesting seminar on 'Segmentation of individual muscles in MR images: How could we handle pathological changes?' to ML-Labs. We are looking forward to visiting Aix-Marseille university in 2022."

UNEP - UNITED NATIONS ENVIRONMENT PROGRAMME

Tlamelo Makati was invited to join the United Nations Environment Programme by the Resilience to Disasters and Conflicts unit. The four month internship is open to individuals with technical backgrounds that are passionate about AI, Robotics, IoT and Drones and are keen on understanding how these technologies have been effectively employed in real case studies around the world to reduce disaster risks. Tlamelo focused on real life situations where AI & Robotics helped to manage risks while four other interns examined IoT and Drones applications. The internship also provided the opportunity to organise and host talks for the Modern Technologies for Disaster Management Webinar Series. The talks were presented by leading experts from NASA and the Centre for Robot-Assisted Search & Rescue (CRASAR) who explained to the general public opportunities and challenges associated with these technologies and their use in disaster management.

PUBLIC ENGAGEMENT →

PUBLIC ENGAGEMENT



THE DANGERS OF EXTREME WEIGHT CUTTING FOR MMA FIGHTERS | RTÉ BRAINSTORM

Mark Germaine shares his interest in machine learning applications in sports and physical performance on RTE Brainstorm. In a series of articles and videos he explains fundamental concepts on how physical activities impact on the human body. In this recent RTE Brainstorm Video tells us about the dangerous effect of rapid weight loss and how this has led to some researchers calling for rapid weight loss to be banned from combat sports.

"Combat sports are typically divided into "weight" (body mass) divisions to ensure athletes of similar size compete against each other in an attempt to ensure fair competition. In professional combat sports, athletes weigh in for the official bout roughly 36 hours prior to the fight. As such, weight regulation, and rapid weight loss in particular, has become a key component of the culture within combat sports such as mixed martial arts (MMA) and boxing.

Whilst weight regulation is not unique to combat sport, the rapid weight loss methods typically deployed in the final 5-10 days somewhat are. During this period, athletes engage in a number of practices such as

Water Loading, Excessive Exercise, Hot Baths, Saunas, and fluid and food restriction in order to manipulate body water (dehydrate) to "make weight". However, whilst dehydration is not uncommon in other sports, the magnitude of dehydration in combination with the duration the athlete is dehydrated places combat sport athletes at greater risk. For example, Brazilian MMA athletes report losing ~13% of body mass in the final week before weigh in, and a recent case study in Liverpool observed 9.3% of body mass loss through dehydration in the final 24 hours alone.

As a result of such rapid weight loss, rises in creatinine have been recorded which indicate acute kidney injury in addition to tripling of stress hormone concentrations (cortisol) and testosterone plummeting. Further to this, plasma sodium concentration is elevated to levels indicative of hypernatremia, which in extreme instances can be fatal. Elevated sodium concentration is likely due to the reduced blood plasma volume observed, as it is lost during the dehydration process. In addition to hypernatremia, this also leads to low blood pressure and is mainly why you may see some athletes struggle to stand on the scales in some instances. In the most extreme circumstances this can lead to death, as has been observed perhaps most famously in 1997 in wrestling.

This has led to some researchers calling for rapid weight loss to be banned from combat sport on the premise that it violates the World Anti-Doping Agency three criteria:

1. Potential to enhance sports performance
2. Potential to risk athletes' health
3. Violates the spirit of the sport

While other areas of research recognise that it is unlikely that banning would have the desired effect and instead aim to explore methods to make the process safer for athletes to engage in. "

More articles from Mark are available on RTE Brainstorm:
What does it take to produce a successful GAA intercounty player?
How does cycling 3,483km in 21 days affect your body?

PHD & PRODUCTIVITY - YOUTUBE CHANNEL

Ciara Feely started her YouTube Channel "PhD and Productivity" in 2019, a few months into 1st year of her PhD. In the past two years it has grown to over 16,000 subscribers. In total the channel has received over half a million views and accumulated 50,000 hours of watch time. Ciara has shared her motivations for becoming a YouTuber with us.

"A few months into my PhD I quickly learnt how challenging a role it can be in terms of productivity, as you are essentially your own boss and the project

manager of a 4-year project. While there is an expected number of hours completed per week, there is flexibility in the specific days and times students work, and it can be challenging to know which work tasks should be prioritised etc.

I read a number of different productivity books to learn about the difficulties of knowledge workers (such as PhD students) working in a distracted world, and the potential solutions, which I found to be very helpful.

I found that a lot of other students were struggling with this, and that there was no one else on YouTube talking about how to plan and organise yourself as a PhD student, and what the best way to work was, so I decided to fill the gap.

I started my YouTube channel "PhD and Productivity" a few months into 1st year of my PhD and in the past two years it has grown to over 16,000 subscribers. In total the channel has received over half a million views and accumulated 50,000 hours of watch time.

I have found the channel to be a great way to connect with other PhD students, to share my research with a wider audience, and to learn about the research that students in universities worldwide are doing.

Ciara's YouTube Channel "PhD & Productivity"

DISCOVER OUR TECH BLOGGERS

Mehreen Tahir wrote her first article on Code Project and since then she has posted more than 20 tech columns covering topics related to Neural Networks, Big Data, keras, and more. Code Project is a community for tech professionals, software engineers, and IT pros with more than 14 million members. They have around 2 Million expert contributors out of which only around 20 get awarded each year. Well done to Mehreen for being on the top one !

Gargi Gupta has only recently started her blog on the medium open platform where she shares her passion for AI and related topics. Do not miss her latest article on a gentle introduction to Explainable Artificial Intelligence (XAI).

COHORT-BASED PHD PROGRAMMES IN IRELAND

In June the SFI Centre for Research Training in Machine Learning hosted a panel discussion titled "Cohort-Based PhD Programmes in Ireland Reflections, Learning & the Future". The event was sponsored by the National Forum for the Enhancement of Teaching and Learning in Higher Education.

The motivation to organise the event goes back to 2018 when SFI launched two calls: The Centres for Research Training programme (CRT), and the Centres for Doctoral Training programme (CDT). Under these two calls 13 cohort-based PhD training programmes were established in 2019 (6 CRTs and 7 CDTs) for a total investment of over €139 million euro in 800 PhD students trained in the skills of the future.

The CRT programmes are designed based on the model of cohort-based PhD training, which is a new educational approach in Irish third level institutions. The aim of the panel discussion was to share knowledge, experience and lessons learned in designing and running these programmes. The ultimate goal was to create a community of practise and to map options for this model into the future.

The panel included directors, managers and students of 5 out of the 13 centres. It became clear that the training programmes designed by the CRTs have common features and include activities like group projects, peer learning, taught lectures, student-led symposia, industry placements, and core modules for technical and transferable skills. The students joining a CRT programme are from all different backgrounds and work on interdisciplinary research projects. The multidisciplinary background of the students necessitated combining the students' existing knowledge with a solid technical foundation. Some of the Centres addressed this challenge with an intensive cohort based training during which the students worked and, in one case, lived together for a period of several weeks. The effectiveness of the training was in all cases measured through surveys and student

performance that helped to improve the structure of the programme for the following cohort.

The CRTs met similar challenges. Some of these included recruiting high calibre students with a strong motivation to persevere in their PhD studies, efficiently managing the interaction between stakeholders, and developing a learning community and a culture that students found very supportive and nurturing.

The pandemic added new challenges. The programmes had to adapt to remote learning and full-time training was replaced by part-time training to accommodate different time zones and mitigate the difficulties associated with online teaching. Stress, anxiety and mental health of the students became more pressing problems that centres addressed using the local professional support available in their host universities.

Many questions were addressed during the discussion including supervisors' experience in handling remote supervision, lessons learned during the pandemic and their value to traditional face-to-face teaching, students' experience before and during the pandemic, and tools to use to compare the effectiveness of the cohort-based PhD model compared to a standard PhD model.

LIVING WITH AI: SHOULD MACHINES MAKE DECISIONS FOR US?

How can algorithms be trained to understand how to make decisions with the same judgment used by people when human judgment is not always objective? How can we ensure that AI is developed in such a way that it does not amplify current social injustice? How can unintentional bias be prevented from being encoded in algorithms?

The esteemed panellists Dr Margaret Mitchell, Prof Eugenia Siapera, Dr Susan Leavy , and Prof Vincent Conitzer helped to unpack these big questions and more in a discussion moderated by the Irish Times journalist Marie Boran during the panel discussion entitled "Living with AI: Should machines make decisions for us?". The event was part of the annual Summer School at the SFI CRT in ML. The recording of discussion is available on the ML-Labs YouTube channel.

Marie Boran is a technology journalist. She writes for the Irish TImes. In 2018 she was nominated Irish Science Writer of the Year, 2018 and in 2016 Technology Reporter of the Year. She also lectures in DCU journalism; science communication; science, technology & society. Marie's current doctoral project focuses on online science audiences/ public narratives around emerging technologies. She has a keen interest in social network analysis, computational social science

Dr Margaret Mitchell is an industrial researcher, works on algorithmic bias and fairness and machine learning. Margaret received her PhD in Scotland where she studied natural language processing primarily, looking at the generation of referring expressions to visible objects mixing cognitive science, statistics and traditional language generation. From there she did a postdoc at Johns Hopkins where she looked at basie and modeling and information extraction sort of more traditional NLP tasks. She increasingly started working in computer vision as well, sort of mixing natural language processing and computer vision work to do things like generation of sentences and descriptions of images. When image captions started to be popular she got interested in the needs of blind people and how image captions can help them. With this mindset she started working on an application for blind people called "Seeing AI". This research approach of looking at what good an application can do brought her to Google where she worked for several years. She is now managing her own company and she is working on a lot of different projects.

Eugenia Siapera is head of the School of Information and Communication Studies in University College Dublin. Her research interests include digital politics data and equality online hate speech and AI in social justice has a background in social science with degrees in psychology and communications and political science. Her research interest is in the area of social justice especially race, racism, misogyny and these kinds of manifestations of inequality and injustice across society. An aspect of her work is the combination of a social scientific perspective with a technological computer science perspective. She recently worked on a research project with computer scientists to develop a new approach to content moderation using neural networks for identifying race contents and social media.

Prof Vincent Conitzer is a professor at the University of Oxford on Duke University, and he is head of the technical engagement at the Institute of ethics in Oxford. His research is on automated moral decision and and professional conduct your works on the problem with automated moral decision making and is interested in ethical societal and policy aspects of AI.

Dr Susan Leavy is assistant professor at University College of Dublin in the School of Information and Communication Studies. Her research interests include AI text mining and she's currently working on cross disciplinary approaches to mitigating bias and machine learning algorithms.She also looks at complexity associated with defining policies for dealing with bias and AI. Susan's primary background is English and philosophy and that led her to artificial intelligence back in the 90s. She combined machine learning and feminist studies in her PhD using natural language processing.

COMPETITIONS →

COMPETITIONS

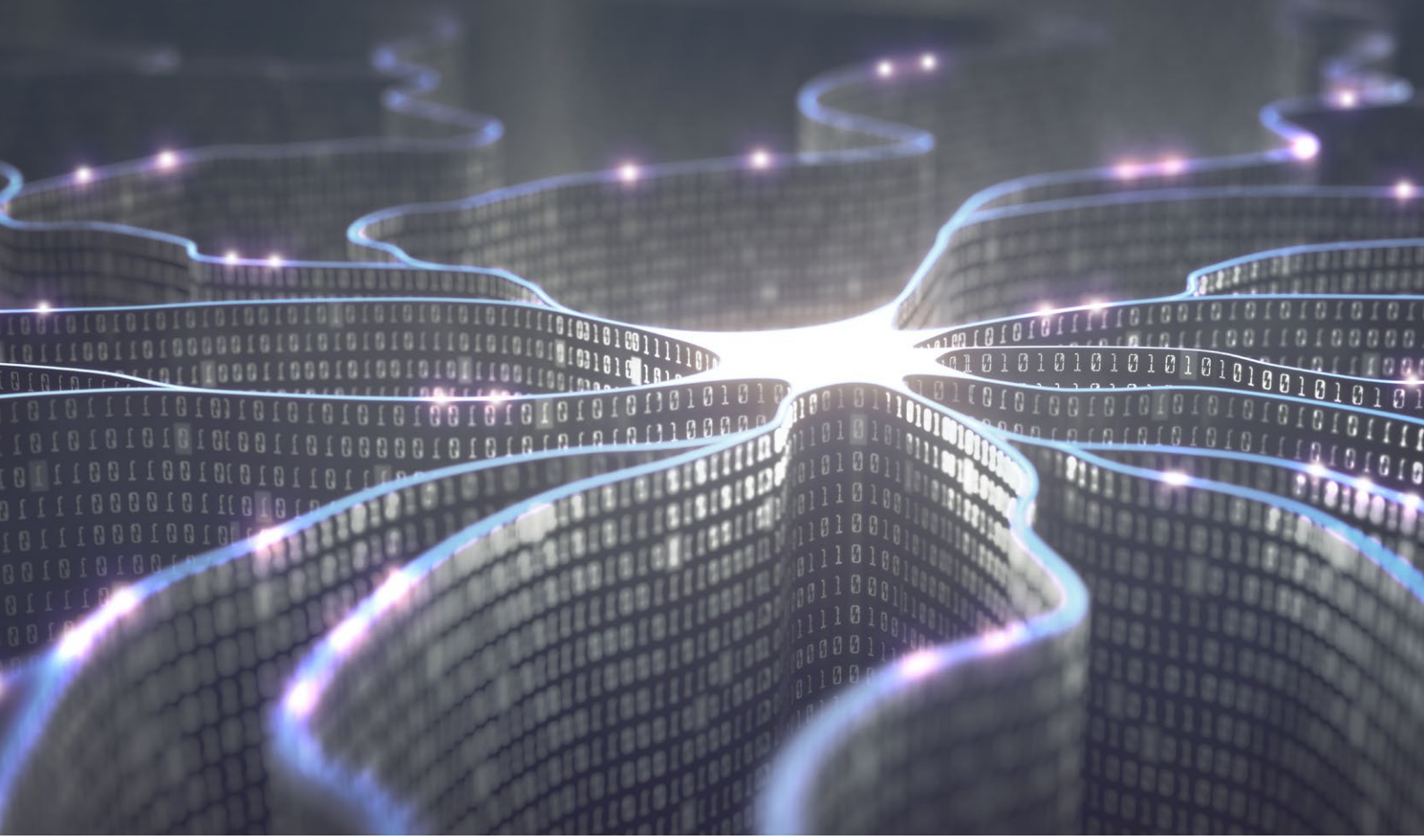
"WANT TO SAVE A RAINFOREST? MIX AI & BLOCKCHAIN"

Adam Stapleton took second place in Ireland's recent FameLab science communication competition. Famelab is a global science communication competition that challenges budding science communicators to give a compelling talk on a topic in STEM in only 3 minutes. The competition is run by the British Council of Ireland with funding from Science Foundation Ireland (SFI). This was the ninth and final year of FameLab, and has taken place entirely virtually due to the pandemic.

"I entered FameLab because it seemed like it would be enjoyable and that I would learn a lot from the process which would be transferable to my own research work. It presented a really good opportunity for me to challenge myself and learn more about scientific communication in a very practical way, as well as the chance to connect with a host of passionate people that share the same affinity for the sciences as I do"

I gave my talk on an application of ML/AI and blockchain in rainforest conservation - a project called GainForest which I really admire. I think it's a great example of how these technologies can be applied for the good of all and achieve things that wouldn't be possible otherwise."

Adam's winning talk is available on Youtube "Want to save a rainforest? Mix AI & Blockchain" and if you want to learn more about Adam's passion for science do not miss his interview with Silicon Republic.



AN ATTACK ON CNNS?

Rian Dolphin, William Blanzeisky, and Jack Nicholls finished in 2nd place at the MUSKETEERS 2nd Hackathon, 2021.

Musketeer is a group funded by the EU Horizon 2020 programme who aim to create a validated, federated, privacy-preserving machine learning platform. They hosted a hackathon in late September 2021 inviting students to develop attacks capable of penetrating their defences in a federated learning environment. Three students from the SFI CRT in ML (Jack Nicholls, Rian Dolphin, and William Blanzeisky) competed in this hackathon and finished in second place.

This Hackathon was a great learning experience on attacking and defending models, and federated ML. Our aim was to attack a federated machine learning model to damage the output. Federated machine learning is a method of using a network of computers to host their own data and run a machine learning model and then submit results to a master node which aggregates the scores. The data remains on the host computers and is not visible to the network. We implemented an attack on the weights of a Convolutional Neural Network (CNN) model which was attempting to classify handwritten numbers on the popular dataset MNIST. Our attacks attempted to skew the overall aggregated output across ten computers by misclassifying certain digits (1s and 7s) in an attempt to trick the CNN.

DETECTING ANOMALIES IN TIME SERIES DATA

Anastasia Natsiou, Sagar Saxena & **Qingyuan Wang** ensured the 3rd place at the Huawei European University Challenge Hackathon 2020. Huawei Ireland, one of the leading tech firms in AI research, organises an annual University Challenge for postgraduate students to compete on a technical solution using Machine Learning techniques.

"This competition has been for us a learning opportunity to understand and apply machine learning methodologies to solve a complex industrial problem. This challenge has not only helped us to hone our technical prowess but to perform as a team and learn from each other as well.

The Huawei Time Series Anomaly detection competition aims at identifying anomalies in a dataset provided by Huawei Ireland. This dataset included labeled sensor-data recorded at specific time intervals for anomalous behavior. Time series analysis is an active area of research in machine learning, and an area of particular focus at the SFI CRT in ML. But anomaly detection in time-series has been one of the most challenging aspects. This may be due to imbalances in the dataset-distribution formed by labeling anomalous data, which are very small compared to normal data. We developed an algorithm that helped in predictive maintenance by predicting anomalous behavior depending on KPIs provided.

The competition has taught us a myriad of different aspects that we were effectively unaware of. Such as learning about anomalous behavior in temporal data and methodologies proposed in the literature to be effective, working in a team, sharing knowledge through brainstorming techniques, to learn from each other, managing work remotely, and meeting deadlines for predefined milestones. Thus, the competition for us was an implicit opportunity to get an insight into this domain-application of ML while learning how to maintain better software engineering practices for code-readability".

GENERATING IMAGES FROM TEXT

In July, **Phúc Le Khac Hong** won the first prize in an online community event with almost 800 participants hosted by HuggingFace. In this event co-organized with the JAX/Flax and Google Cloud teams, compute-intensive NLP, Computer Vision, and Speech projects were made accessible to a wider audience of engineers and researchers by providing free TPUv3s, with tutorials and guides on how to use, train and develop large models using Google's JAX/Flax framework and HuggingFace's Transformers library.

Phúc is a member of DALLI-E Mini team, which trains an open-source a text-to-image generation model, similar to OpenAI's DALL-E.

Unlike the original version, DALLI-E Mini is 27x smaller and only uses publicly open paired image-caption datasets. To make it work, we also leveraged publicly available pre-trained text encoder and

image decoder models from the Transformers' Hub. Despite the time and computation constraints, the resulting model can still output very interesting results (if you squeeze your eyes enough!). What the model lacks in realness, it makes up for in abstractness.

The demo is accessible online on the HuggingFace Space and more information about the model and data set are available on the WandB platform at this link.

RESEARCH ARTICLES →

RESEARCH ARTICLES

WHAT HAS A TRAINED CONVOLUTIONAL NEURAL NETWORK REALLY LEARNED? UNDERSTANDING DEEP REPRESENTATIONS IN COMPUTER VISION THROUGH KNOWLEDGE GRAPH DISTILLATION AND ANALYSIS

By Alessandra Mileo

The problem of understanding and explaining the decision making process of neural networks has been explored by researchers since the 1990s when the models were still shallow (models with only a few hidden layers) and consisted of only fully connected layers. Even with these simplifications compared to Deep Neural Networks (DNNs), providing meaningful explanations for such models proved to be a challenging task. And things did not get easier as more complex architectures such as Convolutional Neural Networks (CNNs) have succeeded in efficiently solving hard problems in Computer Vision such as image classification and object detection.

But what does it really mean to understand (and therefore explain) a CNN? Saliency maps and Concept Activation Maps have been used quite broadly to interpret visually the results of a CNN. But they require human interpretation which can be misleading especially when the CNN makes a mistake, thus introducing potential bias (including human bias) and error.

To overcome these issues, we are looking into understanding neural representations (e.g. trained CNN) by converting them into something we humans can understand and relate to: Knowledge Graphs. Graphs are very well-studied mathematical structures, and no matter how complex they can be, you can still zoom into a knowledge graph and understand how pieces of information are semantically related to each other. In other words, they are inherently explainable!

Our research team has defined the notion of co-activation graph, built from a trained CNN, where nodes are neurons, and edges are statistical correlations of their co-activation values.

Once we generate a co-activation graph we use graph analysis to understand what this graph tells about the neural representations, and we discover many interesting things.

The detection of communities in co-activation graphs can reveal groups of classes that are more similar from the CNN point of view. In fact, we found a strong correlation between the overlaps of communities in co-activation graphs, and mistakes in the neural networks. This means we can find pairs of classes that are likely to be the origin of most mistakes in the model (see figure 1 on MNIST dataset).

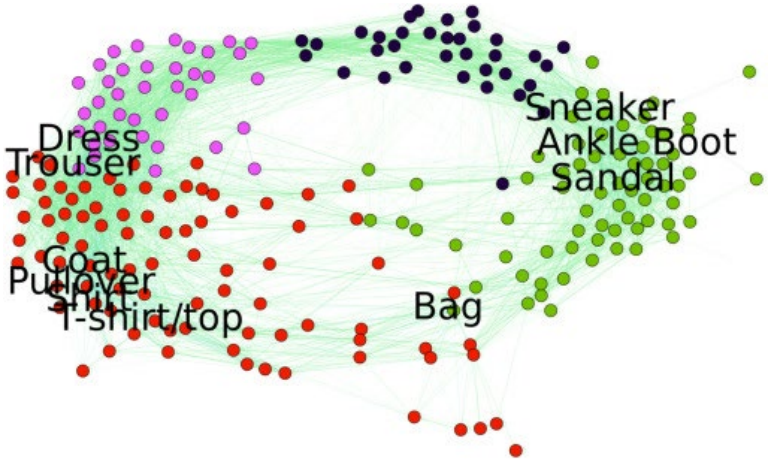


Fig 1 - Community Analysis on co-activation graph generated from MNIST

But the graph can also tell us more about the taxonomy learned by the network as we zoom in and out looking for bigger or smaller communities, or the semantic centrality of certain nodes via node centrality analysis.

Although this helps understanding the model "as a whole", we also started looking into how to generate specific explanations, e.g. for misclassification of a specific image.

To tackle this problem, we have extended our graph by linking instances/images, classes for those images (predicted or labeled) and features of those classes available from external knowledge bases, thus ending up with a much richer graph.

Applying link prediction to such a graph, we were able to identify property values a trained CNN would have likely associated with a specific input image, and we used such prediction to generate factual and counterfactual explanations of mistakes.

An example is illustrated in Fig.2. On a misclassification of a bird on the CUB-200-2011 dataset.

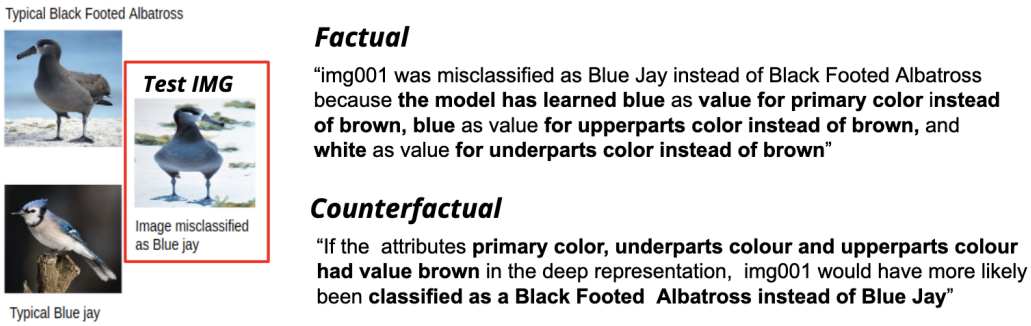


Fig. 2 - Example of local explanations generated in terms of properties through link prediction

A lot of interesting directions for investigation have started from here. For example, we are looking into identification of bias in a dataset when certain properties of a specific class are not learned from the CNN because there are missing instances in the data.

We have also started to look into alternative formalisms for knowledge distillation such as causal rules and how wider adoption of these approaches in areas such as Medical Image Analysis for clinical diagnostics.

EXPLAINABLE ARTIFICIAL INTELLIGENCE:FROM DATA-CENTRIC NEURO-SYMBOLIC AGGREGATION TO EXPLANATORY HUMAN-CENTERED AI

By Luca Longo

Connectionist paradigms for Artificial Intelligence in general, and machines in particular, are often viewed as the future of applications like medical diagnosis, and autonomous transportation relying on large datasets and large scale computation. Similarly, reasoning and logic have come to be seen as technologies from the past, where an abundance of data and computational power was not available, but still the derivation of logical conclusions from available knowledge, facts, and beliefs was necessary - for example in expert or recommender systems. Today, the situation is different, with the formation of neuro-symbolic systems that exploit advantages of both words. However, one fundamental issue of these systems lies within the realm of explainability. More precisely, these systems might be very accurate, but in application terms, they fail to provide their end-users with descriptions on how they were built, or to produce convincing explanations for their inferences, whether used for prediction, classification, or forecasting.

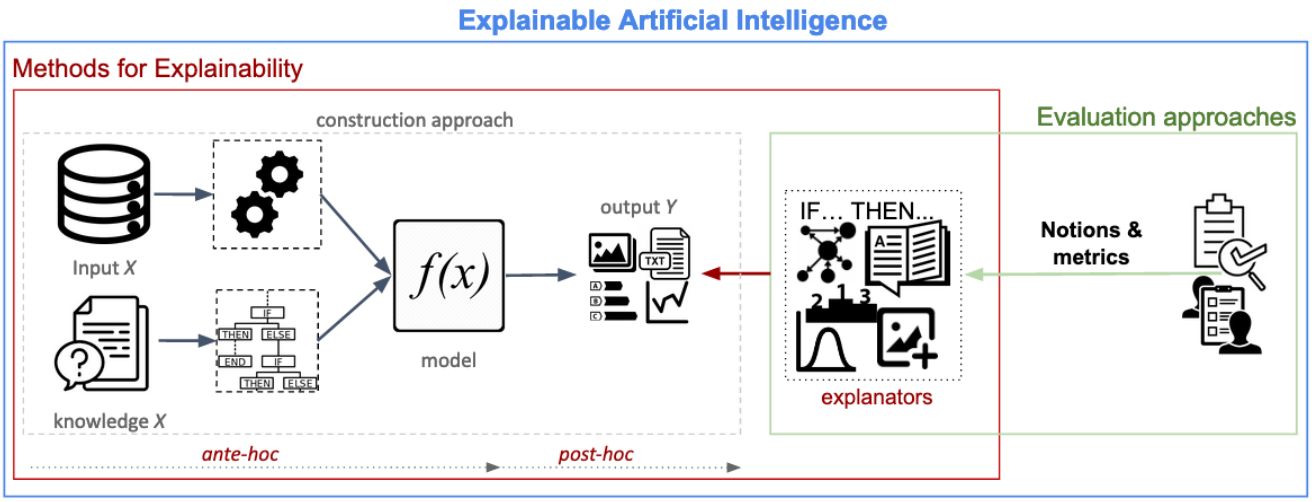


Fig. 1 - The typical XAI pipeline.

In many sensitive applications, such as in medicine, law, and other sectors where the main workers are not computer scientists or engineers, the direct application of these systems, often with underlying complex models, without human oversight, is currently inappropriate. The reasons are not only technical, like their

accuracy, its stability to decisions and susceptibility to attacks, but often arise from sociological concerns, practically settling on the issue of trust. In fact, one of the principal reasons to produce an explanation is to gain the trust of users.

Trust is the main way to enhance the confidence of users with a system as well as their comfort while using and governing it. In turn, trust connects to ethics and the intensity of regulatory activities, as for instance the General Data Protection Regulation in the European Union, leads to many legal and even ethical questions: responsibility for safety, liability for malfunction, and tradeoffs therein must inform decision makers at the highest level. Consequently, to solve the above issues, a field of research has emerged in the last decade: eXplainable Artificial Intelligence (XAI). State-of-the-art XAI-based systems usually follow the pipeline depicted in Figure 1.

Here, neuro-symbolic approaches are usually employed for inductive learning from data, or for exploiting ad-hoc expert-driven knowledge-bases with data. This leads to models that employ mathematical terms and concepts that are usually hard to grasp by lay people. Hence, they are often referred to as 'black-boxes' since their formation, functioning as well as their inferences are hard to explain to lay consumers.

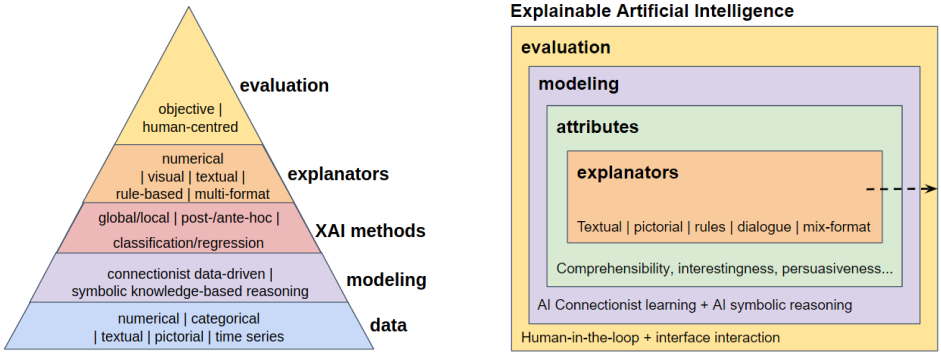


Fig. 2 (left and right parts): XAI systems and their key components.

In a nutshell, Explainable Artificial Intelligence tries to solve these problems, by proposing XAI methods that wrap neuro-symbolic systems with explanators, that in turn are, unfortunately sporadically, evaluated against some metric of explainability by humans (Figure 2, left part). A plethora of ante-hoc and post-hoc methods have been proposed, and these are aimed at producing some output that aims at being explainable. Unfortunately, the construct of `explainability' is ill-defined, multi-dimensional, and context-dependent, largely complicating the production of meaningful explanators. For these reasons, the concept of explainability and the related formation of robust explanators should be treated as the core argument in the field of XAI (Figure 2, right part). In other words, tackling the problem of explainability of systems that manipulate data should start from an analysis of the explanators humans are meant to consume in a given context and within the boundaries of an underlying system. This means scholars should firstly focus on the design and applications of attributes and notions of explainability, and only subsequently build neuro-symbolic models that in turn can be evaluated with a human-in-the-loop approach against actual explainability, by taking advantage of advances in human-computer interaction.

This article is based on the following article which contains a more detailed discussion of the issues raised:: Vilone, G., & Longo, L. (2021). Notions of explainability and evaluation approaches for explainable artificial intelligence. Information Fusion.

OTHER NEWS

TWO BABIES AND A WEDDING

The Centre continues to expand with the addition of new families and two beautifully healthy babies.