SOPHIA

SPRINGBOARD FOR ARTIFICIAL INTELLIGENCE

EDITION 2023

International AI forum for exchanges
between AI academic and private research players

CONTRIBUTED PAPERS
POSTERS & DEMOS

22 > 24 NOVEMBER 2023

SOPHIA ANTIPOLIS
[French Riviera]

SOPHIA-SUMMIT.COM
#SOPHIASUMMIT

COMMUNAUTÉ D'AGGLOMÉRATION SOPHIA ANTIPOLIS
3iA Côte d'Azur
UNIVERSITÉ CÔTE D'AZUR
SOPHIA ANTIPOLIS
SOPHIA CLUB ENTREPRISES
DÉPARTEMENT DES ALPES-MARITIMES

Author Index

From Neural Networks to Quantum Physics and Back

Stéphane Bressan [*†] [1,2], Mario Gattobigio[‡] [3], Frédéric Hébert[§] [3], Christian Miniatura[¶] [3], Dario Poletti[‖] [4], Remmy Zen[**] [5]

[1] National University of Singapore – 21 Lower Kent Ridge Rd, Singapour 119077, Singapore
[2] CNRS@CREATE – 1 Create Way, Singapour, 138602, Singapore
[3] Institut de Physique de Nice – Université Nice Sophia Antipolis (1965 - 2019), Centre National de la Recherche Scientifique, Université Côte d'Azur – Avenue Joseph VALLOT Parc Valrose 06100 NICE, France
[4] Singapore University of Technology and Design – 8 Somapah Rd, Singapour, 487372, Singapore
[5] Max Planck Institute for the Science of Light – Staudtstraße 2, 91058 Erlangen, Germany

In his 2022 "Annotated History of Modern AI and Deep Learning" (2), Juergen Schmidhuber argues that the statistical mechanics model of ferromagnetism proposed by Ernst Ising and Wilhelm Lenz in the 1920s is the first recurrent neural network architecture. The history may have gone full circle when, in 2017, Giuseppe Carleo and Matthias Troyer, in a seminal Science paper (1), proposed neural-network quantum states, variational quantum states implemented as neural networks approximating the wave functions of many-body quantum systems. Quantum many-body systems are notoriously hard to study because the dimension of their Hilbert space increases exponentially with their size. Neural network quantum states have been shown to effectively and efficiently simulate quantum many-body systems. We devise and present several neural network quantum states architectures (5,4,3) improved with original physics-informed transfer learning techniques. We empirically and comparatively evaluate their scalability, efficiency, and effectiveness for studying quantum many-body systems, particularly identifying quantum critical points. Finding the precise location of quantum critical points is significant to characterise quantum many-body systems at zero temperature.
We present several approaches to finding the quantum critical points of such models as the quantum transverse field and next-nearest neighbour Ising and the Heisenberg XXZ models using neural-network quantum states, analytically constructed innate restricted Boltzmann machines, transfer learning, conjoined networks for finding near-crossing of excited states energies, and unsupervised and supervised learning. We also comparatively investigate the efficiency and effectiveness of neural-network quantum states of different sizes, breadths, and depths.

For example, in (5), we use transfer learning protocols previously devised in (4) to determine the transition from a paramagnetic to a ferromagnetic phase in the transverse field quantum Ising model, where the transition is driven by a ferromagnetic coupling parameter J/—h— (Fig. 1 (a) and (b)). We show that the use of transfer learning techniques between neural networks describing the system for different parameters J/—h— allows a convergence towards a ferromagnetic ordered states where the magnetization is large (b). Starting from a random initial state of the

[*]Speaker
[†]Corresponding author: steph@nus.edu.sg
[‡]Corresponding author: mario.gattobigio@inphyni.cnrs.fr
[§]Corresponding author: frederic.hebert@inphyni.cnrs.fr
[¶]Corresponding author: christian.miniatura@cnrs.fr
[‖]Corresponding author: dario_poletti@sutd.edu.sg
[**]Corresponding author: remmy.zen@mpl.mpg.de

network (a) does not provide such a convergence, thus showing the effectiveness of the transfer protocols. These protocols also improve efficiency by reducing the computation time (4) compared to a random initialisation of the neural networks. The phase transition point is located at the inflection point of the curves in the limit of large sizes N where these curves become steep. Using transfer learning protocols between different sizes as well as between different parameters allows the calculation of the inflection point for large sizes (Fig. 1 (c)) where the results obtained converge towards an expected analytical result J/—h— = 1 for the case presented. This shows that the combination of these different transfer learning protocols provides good scalability for describing large systems by neural network quantum states.

The team developed an open-source library called Mapalus to implement the approaches mentioned above. Mapalus interfaces with TensorFlow and leverages general-purpose graphics processing units. The library is available online at github.com/remmyzen/mapalus.
Drawing on this experience, we discuss how a statistical physics energy-based and spectral interpretation of neural network quantum states may serve as a point of convergence for not only many-body quantum physics and machine learning but also quantum computing and combinatorial optimisation. Exploring this convergence can lead to new insights, methodologies, and applications, ultimately driving advancements in scientific research and practical problem-solving.

Keywords: neural networks, quantum physics, combinatorial optimisation, quantum computing

Frequency domain complex-valued autoencoders' impact on image reconstruction and generation

Aymene Mohammed Bouayed *† [1,2], Adrian Iacovelli [2], David Naccache [1]

[1] Département d'informatique - ENS Paris – École normale supérieure - Paris – École normale supérieure 45 rue d'Ulm F-75230 Paris Cedex 05, France
[2] Be-Ys Research – Be-Ys Research – 46 Rue du Ressort, 63100 Clermont-Ferrand, France

Deep generative methods utilize deep neural networks to generate data resembling real-world distributions, thereby expanding the size of datasets. In this work, we focus on autoencoder-based generative methods for image reconstruction and generation. Autoencoders are advantageous due to their fast and stable learning compared to other generative methods (1). Also, they are used in a multitude of generative models, such as diffusion models (2). However, they often produce blurry images (1). Consequently, improving their performance will result in improving the performance of the methods employing them.
The blurriness issue arises from the 2 norm being used in the loss function to compare the input image with its reconstruction (1). This comparison only considers the intensity of each pixel individually, neglecting neighborhood information. To address this limitation, various approaches have incorporated neighborhood information into the loss function to enhance image quality. The focal frequency loss (AE-FFL) (3) and the deblurring loss proposed in (4) utilize frequency domain-based regularization terms in addition to the 2 norm in the spatial domain. By considering global relationships and capturing frequencies difficult to represent in the spatial domain, these methods achieve sharper image generation than standard spatial domain autoencoders (AE).

In our approach, we quantify the impact of frequency domain complex-valued autoencoders on image reconstruction and generation. Since each frequency in the frequency domain is a complex number calculated over the entire image, encapsulating global relations, our work transforms spatial domain images to the frequency domain via the Fourier transform, then employs a complex-valued autoencoder (FAE) to reconstruct the latter image. As a loss function, we utilize an equivalent of the 2 norm in the frequency domain, inferred per Parseval's theorem.
Compared to an AE and an AE-FFL on the MNIST dataset, FAE achieves competitive reconstruction performance up to a latent dimension of 32 (Table 1). However, for image generation using Gaussian mixture model latent space sampling, AE still outperforms FAE. These findings are supported by the analysis of magnitude differences and cosine similarity between images and their reconstructions (Figure 1). In our opinion, FAE represent an interesting backbone architecture that we are working on improving to include local relations between pixels and further improve image quality.

Keywords: Frequency domain, Complex, valued neural networks, Autoencoders.

*Speaker
†Corresponding author: Aymene.Bouayed@ens.fr

Deep NLP model to support multidisciplinary consultation meetings - Application to colorectal cancer

Célia D'cruz *† [1], Michel Riveill [1], Frédéric Precioso [1], Jean-Marc Bereder [2]

[1] Equipe INRIA, I3S, MAASAI – Université Côte d'Azur, CNRS, INRIA, I3S, Université Côte d'Azur, CNRS, INRIA, I3S – France
[2] Centre Hospitalier Universitaire de Nice – Centre Hospitalier Universitaire de Nice, Centre Hospitalier Universitaire de Nice – France

Deciding on the best course of action for cancer patients can be challenging even for healthcare professionals. That is why multidisciplinary concertation meetings (RCP) have been mandatory in healthcare establishments since the first Cancer Plan in 2003. At these meetings, doctors meet to discuss their cancer patients: they analyze patient files, discuss the best treatment options in line with national recommendations, and then fill in RCP reports summarizing various patient-related information and indicating treatment proposals.
However, owing to the large volume of patient files to be reviewed in a short period of time, some of the relevant information contained in patients' medical reports is not systematically taken into account.

The aim of our work is to extract critical information from patients' medical records and systematically include it in RCP reports to ensure that they are as complete as possible. We focus exclusively on colon cancer.

In order to automatically populate the fields in the RCP reports, we tested various document classification models to extract the relevant information from each medical report. These models include HiSAN**(1)** (which performs a weighted average of word embeddings to obtain a sentence embedding, then performs a weighted average of sentence embeddings to obtain the document embedding used for classification), SAN**(1)** (which performs classification on the document embedding obtained by simply performing a weighted average of all word embeddings), and also a model**(2)** that performs a simple average of the CLS token embeddings of each text portion. Working on French medical texts, we use Camembert**(3)** as our encoder.

Nevertheless, these models did not perform well on our small medical report dataset. We hypothesized that models like SAN could see an increase of their accuracy by forcing them to focus more on certain relevant words. To do this, we annotated medical reports by locating the words referring to the information we were looking for to fill in RCP fields. This enabled us to create a model combining the SAN classification model with a NER-like model. This new model not only significantly improves classification performance compared with SAN, but also enables to locate relevant words within a document. However, accurately annotating the text of all documents can be a laborious task: we show that the performance improvement remains present - albeit less so - by choosing to annotate the text of only a subset of documents.

*Speaker
†Corresponding author: celia.dcruz@inria.fr

References:

(1) Gao, S. et al. Classifying cancer pathology reports with hierarchical self-attention networks. Artificial Intelligence in Medicine 101, 101726 (2019).

(2) Mulyar, A., Schumacher, E., Rouhizadeh, M. & Dredze, M. Phenotyping of Clinical Notes with Improved Document Classification Models Using Contextualized Neural Language Models. Preprint at http://arxiv.org/abs/1910.13664 (2020).
(3) Martin, L. et al. CamemBERT: a Tasty French Language Model. in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics 7203–7219 (2020). doi:10.18653/v1/2020.acl-main.645.

Keywords: medical text, deep learning, natural language processing, document classification, information retrieval system

Improving Efficiency and Quality of NeRFs Using Constraints and Priors

Akash Malhotra * [1]

[1] Amadeus – Amadeus, University paris saclay, Laboratoire Interdisciplinaire des Sciences du Numérique (LISN) – Biot, France

3D reconstruction from images, a pivotal technology in computer vision, aims to reproduce real-world objects and scenes in three dimensions using a set of 2D images. This technique has significant applications in various domains such as virtual reality, autonomous driving, and industrial design. However, capturing fine-grained details and accurately representing intricate structures remain challenging. Neural Radiance Fields (NeRF) (4), a recently emerged method, offers a novel approach to this problem by using deep learning to model volumetric scene functions. It uses a fully connected network to predict the volume density and view-dependent color, thereby mapping 5D coordinates (3D spatial and 2D viewing angle) into colors and densities. Despite its innovative approach, traditional NeRF architectures suffer from slow rendering speeds and limited representation capabilities for complex scenes. Despite the different works (3) that attempted to improve on NeRF (4), some limitations are still not addressed such as the need for a high number of posed images and the lack of pretraining.
Our work addresses these limitations by introducing constraints in the form of scene graphs and geometric priors into the NeRF framework. Scene graphs provide a structured and symbolic representation of the scene's entities and their interrelationships. By incorporating this rich, and contextual information into the NeRF model, we can capture complex spatial arrangements and can enhance the accuracy of the generated 3D reconstruction. There have been some attempts at using scene graphs for 3D reconstructions like by Dhamo et al. (2) and by Zhai et al. (5). But to our knowledge, there have been no attempts at using scene graphs to provide constraints for NeRF framework.

Furthermore, we leverage geometric priors to introduce specialized architectures within our enhanced NeRF model, inspired by geometric deep learning(1). These priors, which incorporate knowledge about the data, allow the model to better capture fine-grained details and accurately represent intricate geometries through inductive bias. By integrating such prior information, we reduce the reliance on training data, improve model efficiency, and boost the quality of 3D reconstruction.

Our proposed approach broadens the applicability of NeRF to complex, realworld scenes while maintaining a high level of detail. This research will facilitate advancements in the domains that rely on high-quality 3D reconstructions. It will also encourage further exploration into the integration of knowledge-driven constraints into deep learning-based 3D modeling frameworks.

References

(1) Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. IEEE Signal Processing Magazine, 34(4):18–42, 2017.

*Speaker

(2) Helisa Dhamo, Fabian Manhardt, Nassir Navab, and Federico Tombari. Graph-to-3d: End-to-end generation and manipulation of 3d scenes using scene graphs. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 16352–16361, 2021.

(3) Kyle Gao, Yina Gao, Hongjie He, Denning Lu, Linlin Xu, and Jonathan Li. Nerf: Neural radiance field in 3d vision, a comprehensive review. arXiv preprint arXiv:2210.00379, 2022.

(4) Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 65(1):99–106, 2021.

(5) Guangyao Zhai, Evin Pinar ÅN Ornek, Shun-Cheng Wu, Yan Di, Federico Tombari, Nassir Navab, and Benjamin Busam. Commonscenes: Generating commonsense 3d indoor scenes with scene graphs. arXiv preprint arXiv:2305.16283, 2023.

Keywords: 3D reconstruction, NeRF, Geometric deep learning, Scene Graphs

Hybrid use of operations research and deep learning in the pump scheduling problem

Sophie Demassey , Valentina Sessa , Amirhossein Tavakoli *† [1,2]

[1] Center for Applied Mathematics – MINES ParisTech - École nationale supérieure des mines de Paris – France

[2] Université Côte d'Azur – 3iA Côte d'Azur – Parc Valrose, 28, avenue Valrose 06108 Nice Cedex 2, France

The pump scheduling problem is to plan the pumping operations to minimize the electricity cost within a fixed horizon. The scheduled operations must satisfy the electricity demand while taking into account physical constraints. The pump scheduling problem in water network distribution suffers from integrality constraints (pump status) and nonconvexities (the nonlinear equality relationship between pressure-flow in pipes and pumps). In addition to these nonconvexities, the water stored in tanks (i.e., the states of the problem) links each time step to the other one. As a result, the problem is cast into a large-scale nonconvex Mixed-Integer Nonlinear Programming (MINLP) model(1). Traditional optimization solvers struggle with this, often failing to find converging optimality gaps or feasible solutions in a reasonable amount of time. In the global optimization context, several methods have been proposed to handle these difficulties by providing a tighter polyhedral relaxation via bound tightening and generation of linear inequalities (i.e., cutting planes)(2). In this work, as an alternative, we investigate the hybrid use of temporal/spatial decomposition of the MINLP problem and deep learning to find good feasible solutions with a small computational cost.

Every day, operators have to find an (near-)optimal planning of pumps for a given input, which is the forecast demand and tariff profile over the very next day. Hence, a large amount of data is generated by repeatedly solving the decision-making problem for every day along the years. The data correspond to a collection of tariff and demand profiles along with the corresponding optimal solution.Therefore, it is reasonable to exploit this large quantity of data and devise a learning algorithm to map these inputs (demand and tariff profile) to the optimal decision (pump status) according to historical data for a specific network. However, the direct mapping from input to optimal pump status might not lead to a feasible solution of the MINLP model due to a high number of binary decision variables and the complexity of the problem. To address this issue, instead of directly predicting the decision variables (a set of binary sequences), we predict the optimal states of the problem (the stored water in the tanks), which usually have a smaller dimension. Moreover, the states are continuous, and their values follow a fairly regular patterns capturing the underlying temporal dependencies. Subsequently, to retrieve the pumps status from a predicted state, a post-processing can be implemented. Note that, for given states, there is at most one feasible pumps configurations.

Moreover, there is no guarantee that the predicted states yield to a feasible solution of the MINLP problem. To recover feasibility, we utilize a decomposition algorithm, analogous to Penalty Alternating Direction Methods (PADM) (3). The algorithm breaks down the large-scale problem into smaller, independently handled subproblems, and penalizes any mismatch in coupling constraints. The efficiency of this method significantly depends on the initial states (predicted tanks levels), emphasizing the importance of an accurate prediction.

*Speaker

†Corresponding author: amirhossein.tavakoli@mines-paristech.fr

Our work contributes to the field by: (i) introducing a deep learning architecture to predict optimal states for given demand and tariff profiles; (ii) restoring feasible solutions via the decomposition algorithm, penalizing mismatched coupling constraints; and (iii) capturing model epistemic uncertainty through Bayesian inference approximation via Monte Carlo dropout layers(4). This not only provides credibility intervals for predicted states, but also offers multiple starting points when the optimal predicted states do not yield feasible solutions. The preliminary results suggest the eminent role of the predicted states over the decomposition efficiency.

(1)Bonvin, G., Demassey, S., & Lodi, A. (2021). Pump scheduling in drinking water distribution networks with an LP/NLP-based B&B. Optimization and Engineering, 1-39. 2

(2)Tavakoli, A., Demassey, S., & Sessa, V. (2023, February). Strengthening Mathematical Formulation for Global Optimization of the Operational Water Network Distribution. In 24ème édition du congrès annuel de la Société Française de Recherche Opérationnelle et d'Aide à la Décision ROADEF

(3)Geißler, B., Morsi, A., Schewe, L., & Schmidt, M. (2018). Solving highly detailed gas transport MINLPs: block separability and penalty alternating direction methods. IJOC, 30(2), 309-323

(4)Gal, Y., & Ghahramani, Z. (2016, June). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In ICML (pp. 1050-1059)

Keywords: pump scheduling decision making problems, mixed integer nonlinear program, deep learning, penalty alternating direction methods

Deep NLP model to support multidisciplinary consultation meetings - Application to colorectal cancer

Célia D'cruz *† [1], Michel Riveill [1], Frédéric Precioso [1], Jean-Marc Bereder [2]

[1] Equipe INRIA, I3S, MAASAI – Université Côte d'Azur, CNRS, INRIA, I3S, Université Côte d'Azur, CNRS, INRIA, I3S – France
[2] Centre Hospitalier Universitaire de Nice – Centre Hospitalier Universitaire de Nice – France

Deciding on the best course of action for cancer patients can be challenging even for healthcare professionals. That is why multidisciplinary concertation meetings (RCP) have been mandatory in healthcare establishments since the first Cancer Plan in 2003. At these meetings, several doctors from different disciplines (surgeons, radiologists, etc.) meet regularly to discuss their cancer patients. During these meetings, doctors analyze patient files, discuss the best treatment options in line with national recommendations, and then fill in RCP reports summarizing various patient-related information and indicating treatment proposals.
However, owing to the large volume of patient files to be reviewed in a short period of time, some of the relevant information contained in patients' medical reports is not systematically taken into account. Designing an information retrieval system would therefore be valuable.

The aim of our work is to extract critical information from patients' medical records and systematically include it in RCP reports to ensure that they are as complete as possible. We focus exclusively on colon cancer.

Medical reports are a few pages long and follow a structure (a pre-determined sequence of sections). The text of each section is freely written. One of the main types of medical reports is follow-up letters with sections describing the patient's comorbidities, cancer history, previous treatments, biological parameters, etc. Another important type of report is the anatomopathological one describing the characteristics of the tumor. Other types of reports include surgical reports, scanner results, etc.

RCP reports are structured documents mostly in key-value format where values are to be selected among a list of predefined options. The first part of it indicates critical patient information such as the characteristics of the cancer, previous cancer treatments, patient's functional capacity, comorbidity, etc. that is found in the patient records. The last part of those reports is the doctors' treatment proposal.

In order to automatically populate the fields in the RCP reports, we tested various document classification models to extract the relevant information from each medical report. These models include HiSAN**(1)** (which performs a weighted average of word embeddings to obtain a sentence embedding, then performs a weighted average of sentence embeddings to obtain the document embedding used for classification), SAN**(1)** (which performs classification on the document em-

*Speaker
†Corresponding author: celia.dcruz@inria.fr

bedding obtained by simply performing a weighted average of all word embeddings), and also a model**(2)** that performs a simple average of the CLS token embeddings of a Transformer Encoder on each text portion. Working on French medical texts, we use Camembert**(3)** as our encoder.

Nevertheless, these models did not perform well on our small medical report dataset. We hypothesized that models like SAN could see an increase of their accuracy by forcing them to focus more on certain relevant words. To do this, we annotated medical reports by locating the words referring to the information we were looking for to fill in RCP fields. This enabled us to create a model combining the SAN classification model with a NER-like model (Named Entity Recognition). This new model not only significantly improves classification performance compared with SAN, but also enables to locate relevant words within a document. However, accurately annotating the text of all documents can be a laborious task: we show that the performance improvement remains present - albeit less so - by choosing to annotate the text of only a subset of documents.

We are also currently working on a dictionary-based approach, where we have compiled a list of terms that usually refer to a RCP field. However, dictionary terms taken out of context are not necessarily relevant when found in reports. We aim to create a model that both selects the right terms within a text and outputs the RCP field classification.

References:

(1) Gao, S. et al. Classifying cancer pathology reports with hierarchical self-attention networks. Artificial Intelligence in Medicine 101, 101726 (2019).

(2) Mulyar, A., Schumacher, E., Rouhizadeh, M. & Dredze, M. Phenotyping of Clinical Notes with Improved Document Classification Models Using Contextualized Neural Language Models. Preprint at http://arxiv.org/abs/1910.13664 (2020).
(3) Martin, L. et al. CamemBERT: a Tasty French Language Model. in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics 7203–7219 (2020). doi:10.18653/v1/2020.acl-main.645.

Keywords: medical text, deep learning, natural language processing, document classification, information retrieval system

RETROcode: Leveraging a Code Database for Improved Natural Language to Code Generation

Sylvain Marty * , Nathanaël Beau * † [1]

[1] Université de Paris Cité – Paris Cité – Université de Paris Cité, LLF, CNRS, 75013 Paris, France, France

Code generation is the task of automatically creating computer programs from natural language, generating potentially previously unseen code. It has a wide range of applications, from creating code snippets for developers to generating complete software applications. In recent years, the increasing availability of large amounts of code and natural language data has facilitated the development of powerful neural network models that can perform code generation with high accuracy.
One challenge in working with large amounts of natural language and code data is the lack of aligned examples, which require human expertise to annotate. To address this issue, one approach is to use large pre-trained models that have been trained on a large volume of code and/or natural language data, and then fine-tune them on the available annotated data (Li et al., 2022). The use of large models with a high number of parameters can provide computational benefits during training and inference, as well as improved memorization of the training data. However, training these models can be computationally expensive, and the large number of parameters may lead to overfitting on the training data (Bender et al., 2021).

An alternative approach for translating natural language to code is code retrieval, which involves searching for and retrieving an appropriate code snippet from a code database (Gu et al., 2021). However, these methods are becoming less commonly used as it is now possible to use pre-trained models that are trained on the entire code database and generate personalized code responses to a given query.

Methods for natural language generation often involve the use of generative models that are trained to associate text with data in a database. These solutions have two main advantages: they allow for the separation of world knowledge from language learning, and they enable the use of smaller model sizes. For instance, the KNN-Based Composite Memory system (Fan et al., 2021) assists a conversational agent by providing access to information from similar discussions and by supplying relevant knowledge from various sources based on the input user prompt. Another example is RETRO architecture (Borgeaud et al., 2022), which provides information to a language model as decoding goes using sentences similar to what was generated. In both examples, queries are made to a database by comparing the embeddings of the input or output with those in the database to obtain the nearest neighbors, and the resulting information is provided to the encoder or decoder, respectively.

We introduce RETROcode, a transformer-based architecture combining information from a natural language utterance and codes from a database similar to what is being generated. Our

*Speaker
†Corresponding author: n.beau@groupeonepoint.com

approach aims to take advantage of the large amounts of code available while minimizing the number of model parameters.

In this presentation, we showcase a novel Transformer-based architecture adept at synthesizing code from natural language description with analogous code drawn from a database. We delve into two distinct methodologies for integrating this data within the decoder and analyze the influence of various integral components on the overall performance of the system. Our research results underscore the cutting-edge performance in the realm of code generation. In particular, this discussion:

• Unveils the innovative Transformer-based architecture that marries natural language inputs with corresponding code from a database. We put forth two prospective methods for amalgamating this data.

• Inspects how vital elements of the architecture sway the system's performance. We consider the dimensions of the database (db), the preprocessing approaches applied to db samples, the size of the model, the mechanism used to blend natural language and db information, and the training duration when juxtaposed with larger models.

References

- Competition-level code generation with alphacode. https://doi.org/10.48550/arXiv.2203.07814

- ACM Conference on Fairness, Accountability, and Transparency, Virtual Event / Toronto, Canada, March 3-10, 2021 https://doi.org/10.1145/3442188.3445922

- Deep code retrieval based on semantic dependency learning. https://doi.org/10.1016/j.neunet.2021.04.019

- Augmenting transformers with knn-based composite memory for dialog. https://doi.org/10.1162/tacl.a.00356

- International Conference on Machine Learning. https://proceedings.mlr.press/v162/borgeaud22a.html

Keywords: Code Generation, kNN neural network, database retrieval, transformer based architecture

Augmented SPDNet: Second-Order Neural Network for Motor Imagery–Based BCI

Igor Carrara * [1,2], Bruno Aristimunha * [3,4], Marie-Costance Corsi [5], Raphael Yokoingawa De Camargo [3], Sylvain Chevallier [4], Theodore Papadopoulo [1,2]

[1] Inria centre at Université Côte d'Azur (INRIA - CRONOS team) – L'Institut National de Recherche en Informatique et e n Automatique (INRIA) – Inria Sophia Antipolis – INRIA – Méditerranée 2004, route des Lucioles – BP 93 06902 Sophia Antipolis Cedex, France, France
[2] Université Côte dÁzur – Université Côte d'Azur (UCA) – Parc Valrose, 28, avenue Valrose 06108 Nice Cedex 2, France
[3] Universidade Federal do ABC = Federal University of ABC = Université Fédérale de l'ABC [Brazil] – Av. dos Estados, 5001 - Bangú, Santo André - SP, 09210-580, Brésil, Brazil
[4] Université Paris-Saclay – Université Paris-Saclay,Sorbonne Universités – Bâtiment Bréguet, 3 Rue Joliot Curie 2e ét, 91190 Gif-sur-Yvette, France
[5] Algorithms, models and methods for images and signals of the human brain – Sorbonne Universite, Inria de Paris, Institut du Cerveau = Paris Brain Institute – Hôpital de la Pitié-Salpêtrière Batiment ICM 47 boulevard de l'hopital 75013 Paris, France

A Brain-Computer Interface (BCI) establishes a direct communication between the brain and an external device without requiring muscle control. Electroencephalography (EEG) is commonly employed in BCI systems, relying on electrodes to measure difference potentials at the scalp surface over time.
Deep learning (DL) techniques have proven promising for Motor Imagery (MI) based BCI. Traditionally, DL has been applied on raw EEG data, jointly processing the spatial and temporal features. However, DL models often exhibit sub-optimal performance when compared to machine learning based on Riemannian geometry in the most common evaluation (1). This approach involves extracting Symmetric Positive Definite (SPD) matrices from EEG signals and treating them in their appropriate geometrical space, which is a differentiable manifold with a natural Riemann structure.

In this research, we propose a Deep Learning architecture that explores alternative features capturing additional types of information and enhance the discrimination of mental states among subjects. In particular, we test various estimators like the Covariance, Instantaneous and Imaginary Coherence combining them with the same methodology develop in the Augmented Covariance Method (ACM) (2). Of these features, Functional Connectivity (FC) – represented by Instantaneous and Imaginary Coherence – incorporates complementary features that reflect interactions between different brain regions. FC is modelled as an SPD matrix, considers global connectivity patterns instead of relying solely on local measurements, and leads to performance improvements (3).
Another effective method is the Augmented Covariance Method (ACM), which extracts SPD features from EEG data containing both spatial and temporal information, leading to solid and robust performance. To classify these different SPD representations, we employ the SPDNet architecture (4), which is specifically designed to handle SPD matrices.

*Speaker

To validate our approach, we rely on the MOABB (5) package and focus on Within-Session evaluation, conducting tests across multiple tasks and datasets. The results of our research demonstrate that the augmented approach applied on different estimator and classified using SPDNet outperforms the current state-of-the-art DL architecture in BCI decoding. However, there is still room for improvement, as this method obtain sub-optimal performances compared to the ACM methodology combined with a Support Vector Machine (SVM).

Keywords: Brain, Computer Interfaces, Electroencephalogram, Functional connectivity, SPD manifold, Riemannian optimization, Neural network, Motor Imagery, SPDNet..

Estimating the energy consumed by machine learning.

Charlotte Rodriguez *† [1,2], Laura Degioanni [2], Laetitia Kameni [2], Richard Vidal [2], Giovanni Neglia [3]

[1] Inria Sophia Antipolis - Méditerranée – Institut National de Recherche en Informatique et en Automatique – 2004 route des Lucioles BP 93 06902 Sophia Antipolis, France
[2] Accenture Labs [Sophia Antipolis] – Accenture Labs [Sophia Antipolis] – France
[3] Inria Sophia Antipolis - Méditerranée – Institut National de Recherche en Informatique et en Automatique – 2004 route des Lucioles BP 93 06902 Sophia Antipolis, France

Machine Learning (ML) algorithms may consume a significant amount of energy, a well-known fact that has garnered significant attention since (1) raised concerns about the environmental impact of training large natural language processing (NLP) models. Several methods, such as that used by (1), have been proposed to measure, estimate and even forecast the energy consumed by machine learning, and especially by the training of a given ML model. Their relevance, however, is not always universally agreed upon, as demonstrated by the response of Google researchers (2) who were involved in training some of the ML models examined by (1). They assert that the estimations provided by the latter are off by a substantial factor when applied to a company like Google. This specific example, along with others, emphasizes the necessity for further study of these methods.
In light of this, and in line with the research conducted by (3,4), we propose to analyze the energy consumption outputs of five of the most common or popular methods encountered, across a range of ML tasks. The considered machine learning tasks span two domains - computer vision and NLP - and vary in computational complexity, enabling us to assess the methods in diverse contexts. Here, each method is tested in up to three different usage modes. These methods (and modes) differ from each other on several points. Specifically, the underlying calculation can be based on floating-point operations, the duration of the task and utilization rate of the hardware (CPU, RAM, GPU), or on data obtained from internal sensors embedded within these hardware components. While some methods are designed for machine learning, others are applicable to any computing task. They may also require different levels of information, ranging from user experience of running the concerned type of ML algorithm, to concurrent execution with a tool. This notably entails that some methods, but not all, may be used for forecasting the energy consumed.

This experimental protocol, mainly implemented in Python, was executed on commonly available hardware. The methods' outputs are compared with measurements from an external power meter (EPM). For each ML task, we conduct repeated testing of all methods for a specific number of iterations. In each iteration, the methods are tested sequentially but in a randomized order. The protocol's architecture allows for the relatively easy addition of new methods to be tested, and new ML algorithms on which to evaluate them. We conduct qualitative and quantitative observations on the progress and outcomes of these experiments. The former include variations in the robustness of the tested methods under similar conditions in terms of hardware and the ML task being tracked, which may be attributed to differences in required resources.

*Speaker
†Corresponding author: rdz.charlotte@gmail.com

We also compare the output of all methods with that of the EPM's and analyze their evolution across our range of ML algorithms (Fig. 1). In addition, some methods allow for the finer observation of relative energy usage among hardware. We believe that the simplicity of this setup will facilitate further expansion of the current work in the community, and notably enable the analysis of new methods in a broader range of ML contexts.

(1) Strubell et al. 2019. Energy and Policy Considerations for Deep Learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.
(2) Patterson et al. 2021. Carbon emissions and large neural network training. arXiv:2104.10350.
(3) Bannour et al. 2021. Evaluating the carbon footprint of NLP methods: a survey and analysis of existing tools. In Proceedings of the Second Workshop on Simple and Efficient Natural Language Processing.
(4) Jay et al. 2023. An experimental comparison of software-based power meters: focus on cpu and gpu. In CCGrid 2023-23rd IEEE/ACM international symposium on cluster, cloud and internet computing.

Keywords: energy consumption, environmental impact, machine learning, experimental protocol, benchmark, sustainable AI

NLP technologies for building bridge between recruiters and candidates: A new hybrid approach

Chemchem Amine* [1], Sabrine Farah † [2], Loïc Maisonnasse [3], Christian Retore [4]

[1] Pôle Data Driven Intelligence, Atos Montpellier – ATOS Montpellier, immeuble Archimède, Montpellier, France – France
[2] Pôle Data Driven Intelligence, Atos Montpellier – ATOS Montpellier, immeuble Archimède, Montpellier, France, Laboratoire d'informatique, de robotique et de microélectronique de Montpellier (LIRMM) – France
[3] Pôle Data Driven Intelligence, Atos Montpellier – ATOS Montpellier, immeuble Archimède, Montpellier, France – France
[4] LIRMM – Laboratoire d'informatique, de robotique et de microélectronique de Montpellier (LIRMM) – France

In today's competitive job market, efficient matching the right talent to the right roles by recruiters is crucial for successful hiring processes. One of the most significant challenges in management is the automatic and efficient matching of job seekers' skills with the requirements of job openings. This research paper presents a solution that effectively addresses this issue.
Initially, we conducted a syntactic extraction (1) from the European skills repository, employing three algorithms: N-gram, N-gram RI on prefix tree, and regular expressions on raw text (2). However, the results obtained from this initial approach proved inconclusive. It became evident that even with meticulous corpus cleaning techniques, such as removing stopwords, converting to lowercase, and lemmatizing terms, the approach yielded limited results when the skills were expressed differently from those in the repository.

Addressing this, we delved into semantic extraction by transforming the skills and utilizing pretrained models such as Bert (3), which yielded slightly improved results. However, we opted for Camembert (4), the French equivalent of Bert, as FlauBERT (5); another French model, was trained on more formal texts that did not align with our corpus. A comparative study between the two models revealed that Camembert exhibited significantly better performance for our specific use case

We subsequently implemented the Camembert model through two approaches: (1) global embedding, which transforms the job posting into a global semantic vector, and (2) local embedding, which tokenizes the job posting and transforms each token into a semantic vector. The global approach yielded inconsistent results and generated a significant number of false positives, unlike the local approach, which successfully identified the correct skills but only when the token contained a skill.

Therefore, we locally fine-tuned Camembert using a dataset containing core skills from the reference framework along with stopwords. We then combined it with a filter to extract the job

*Corresponding author: lamine.chemchem@atos.net
†Speaker

domain and calculate the semantic similarity between each stopword token and the domain to retrieve misclassified skills.
The results obtained with this hybrid approach are highly satisfying, achieving an f-score of 87.08%, surpassing all other approaches. This solution will soon be proposed to *pole emploi* and will be easily replicable for all other recruitment structures.

Keywords: Natural langage processing, semantic extraction, transformers models, Bert models

Who is likely to engage with bots on Twitter?

Zakaria Babutsidze * [1,2], Jake Koerner [3], William Rand [4], Ana-Maria Staicu [3]

[1] Observatoire français des conjonctures économiques – Sciences Po – France
[2] SKEMA Business School – SKEMA Business School-UCA – France
[3] Department of Statistics, North Carolina State University – United States
[4] Poole College of Management, North Carolina State University – United States

The proliferation of online social networks (e.g. Facebook, Twitter, etc.) has created significant challenges for modern societies. One of them is the easiness of implementing automated accounts (hereafter bots) who could pedal certain type of content through the network without regard of the veracity of this content. Bots on Twitter, Facebook and YouTube have been explicitly linked to disinformation campaigns (1,2). The fact that automated bots contribute to the spread of fake news, low-credibility and inflammatory content is without a doubt (3,4). The scale of inflicted damage of such activities could reach devastating proportions even for establisheddemocracies (5).
Yet, not much is known if one can characterize the behavior of Twitter users who are susceptible to interacting with bots. Such characterization would allow for potential identification of susceptible users and for a better management of the social platform. When properly employed, such analyses could have far reaching implication not only for the financial bottom-line of electronic social platforms (like Twitter), but also for the democratic processes at large. In this work we try to contribute to filling this gap.

We use a unique data-set comprised of tweeting behavior of users who interacted with a known automated bot, as well as those users who have not interacted with a known bot. The main statistical challenge is accurately modeling processes underlying decision of users to tweet in a given period, as well as ones determining frequency of generated tweets and the length of the generated content. An important point is the comparability of the two types of users - treated, and not treated. This is achieved by matching as described bellow. Matching gives us a pair of accounts (one treated and one non-treated) that is used as a unit of analysis. The discrepancies between treated and non-treated accounts is modeled using two latent processes that are assumed to have smooth moment functions. We characterize representative behavior of two groups based on their activity on social network, as well as their tweeting style.

Following such characterization, we develop a new supervised machine learning tool to classify a new unlabeled account into either as susceptible to interact with the bot, or not susceptive to interact with an automated agent. The classifier is based on a dimensionality reduction technique of Functional Principal Component Analysis and joint multiple Likelihood ratio testing. We show that the new classifier is sufficiently accurate in distinguishing susceptible accounts from non-susceptible ones few months before the potential interaction. This gives policy-makers an opportunity to intervene locally to counter-act potential damaging actions from bots. One advantage of the classifier is that it can be extended to include data on the content generated by the accounts. Such information is likely to increase the performance of the machine learning

*Speaker

tool even further.

References:

(1) Golovchenko, et al. (2020) Cross-platform state propaganda: Russian trolls on twitter and youtube during the 2016 US presidential election. The InternationalJournal of Press/Politics, 25(3):357–389.

(2) Ferrara. (2017) Disinformation and social bot operations in the run up to the 2017 french presidential election. CoRR, abs/1707.00086.

(3) Lazer, et al. (2018) The science of fake news. Science, 359(6380):1094–1096.

(4) Stella, et al. (2018) Bots increase exposure to negative and inflammatory content in online social systems. Proceedings of the National Academy of Sciences, 115(49):12435–12440.
(5) Kramer, et al. (2014) Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences, 111(24):8788–8790.

Keywords: Social networks, Classification, Dimensionality reduction, Functional Principal Component Analysis

Public disorder image classification using Deep Learning in MLOps

Romain Bardon * [1], Yannick Lecroart

[1] Romain Bardon – ATOS Montpellier, immeuble Archimède, Montpellier, France – France

We present a Deep Learning model integrated within an MLOps platform to classify public disorder images (1) (2). The core objective of this work was to improve users experience by suggesting relevant disorder categories from the user's pictures. Before the implementation of this AI service, citizens had to navigate in the existing mobile application through a cumbersome selection process, choosing from more than 35 categories to report incidents. This led to confusion, early termination of the process and a long scrolling process. With the integration of the AI as a web service, citizens now have the convenience of selecting from three automatically detected categories while still using the same application they were used to. This streamlined process secures the quality of the information provided by the user and improved citizen satisfaction. It therefore improves the city's technical services ability to swiftly assign each request to the appropriate area of responsibility, ensuring efficient allocation of resources and timely resolution of problems.
One of the key components of the proposed solution is a self-learning Artificial Intelligence (AI) model (3). This AI model, based on DenseNet121 (4), leverages mature image classification algorithms to accurately analyze and categorize reported issues, whether it's related to cleanliness, roads, signage, or illicit posters, facilitating efficient problem-solving.

The model was trained on approximately 300,000 photos and their associated typology leading to a 90% top-3 accuracy rate for assigning citizen requests to the correct category, surpassing initial expectations. The system guarantees a response time per request of less than one second.

We implemented an industrialized approach, leveraging Machine Learning Operations (MLOps) to automate, deploy, and monitor AI models in a production environment. This approach enables to continuously retrain the model, ensuring smooth evolution of the service as well as its scalability, rapid deployment, and seamless integration with existing systems. All the components used are Open Source.
Furthermore, we prioritize digital sobriety by adopting a lightweight approach that minimizes resource consumption. This solution is currently in production in a main French metropolis. The implementation relies on two virtual machines, with no GPUs, providing an optimal balance between performance and environmental sustainability.

Keywords: Computer vision, MLOps, Deep Learning, DenseNet

*Speaker

Automatic denoising of high-dimensional tissue images to improve the cell segmentation

Matteo Pentassuglia * [1], Marion L. Tiberti† [2], Francesco Galati [1], Benedicte Buttard [2], Clemence Ginet [2], Maria A. Zuluaga [1], Aida Meghraoui [2]

[1] Data Science Department, EURECOM, Sophia Antipolis – Eurecom [Sophia Antipolis] – France
[2] AMKbiotech, Sophia Antipolis – AMKbiotech – France

Imaging mass cytometry (IMC) is a powerful biological tissue analysis technique that combines the detection of over 40 features with spatial resolution at a single cell level on a unique sample1. By evaluating tissue architecture, phenotypic cell heterogeneity, and potential cell interactions. IMC offers crucial insights for the identification of disease biomarkers and new therapeutic targets 2,3.
However, the analysis of high-dimensional images is complex and remains a challenging task. The variations in tissue staining performance and signal to noise ratios across tissues represent an additional complexity level to data analysis, making image denoising and normalization a necessary and crucial analysis step. Existing denoising workflows require manual or semi-automated image by image and target by target processes and represents a significant source of bias4. This laborious approach hampers the analysis and limits the potential automatic and non-supervised image analysis.

To overcome the current limitations of established denoising workflows, we propose an innovative computational denoising tool that leverages the power of a U-net neural network. Unlike conventional methods that rely solely on pixel intensity, our approach employs a two-part symmetric architecture to capture the contextual information of the image. It thus robustly identifies the signal's location, enhancing denoising capabilities. To train our model, we used a diverse set of tissue (ie: human Skin, Tonsil, and Colon) including both histologically healthy and pathologic tissue (ie: Tumors, inflammation). Additionally, we incorporated a comprehensive panel of over 30 measured features. The output of our model is a binary mask that effectively distinguishes signal from noise for each feature. By leveraging our tool, we can efficiently clean up all measured features, enhance the quality of IMC data and eliminate the need for time-consuming manual tasks.

To evaluate the performance of the developed tool, we used, on every feature, the Intersection over Union (IoU) between the predicted signal areas and those found following a standard semi-automatic procedure. Our model achieved an average IoU score of 0.502 with an independent test set. This promising result, coupled with visually appealing outcomes, is an indication of the effectiveness of our tool, its robustness and generalizability. Moreover, the obtained improvements facilitate the analysis and comparison of multiple samples and allow the normalization of the dataset for the subsequent steps in the image analysis pipeline.

*Speaker
†Corresponding author: tiberti@amkbiotech.fr

1 Giesen C, et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. Nat Methods (2014) 11:417–22.

2 Jackson HW, et al. The single-cell pathology landscape of breast cancer. Nature (2020) 578:615–20.

3 Elaldi, R. et al. "High dimensional Imaging Mass Cytometry™ panel to visualize the tumor immune microenvironment contexture." Frontiers in Immunology 12 (2021): 666233.
4 Ijsselsteijn, ME, Somarakis, A, Lelieveldt, BPF, Höllt, T, de Miranda, NFCC. Semi-automated background removal limits data loss and normalizes imaging mass cytometry data. Cytometry. 2021; 99: 1187– 1197. https://doi.org/10.1002/cyto.a.24480

Keywords: imaging mass cytometry, denoising, image analysis, U, NET

Automatic denoising of high-dimensional tissue images to improve the cell segmentation

Matteo Pentassuglia * [1], Marion L. Tiberti† [2], Francesco Galati [3], Benedicte Buttard [2], Clemence Ginet [2], Maria A. Zuluaga [1], Aida Meghraoui [2]

[1] Data Science Department, EURECOM, Sophia Antipolis – Eurecom [Sophia Antipolis] – France
[2] AMKbiotech, Sophia Antipolis – AMKbiotech – France
[3] Data Science Department, EURECOM, Sophia Antipolis – AMKbiotech – France

Imaging mass cytometry (IMC) is a powerful biological tissue analysis technique that combines the detection of over 40 features with spatial resolution at a single cell level on a unique sample1. By evaluating tissue architecture, phenotypic cell heterogeneity, and potential cell interactions. IMC offers crucial insights for the identification of disease biomarkers and new therapeutic targets 2,3.
However, the analysis of high-dimensional images is complex and remains a challenging task. The variations in tissue staining performance and signal to noise ratios across tissues represent an additional complexity level to data analysis, making image denoising and normalization a necessary and crucial analysis step. Existing denoising workflows require manual or semi-automated image by image and target by target processes and represents a significant source of bias4. This laborious approach hampers the analysis and limits the potential automatic and non-supervised image analysis.

To overcome the current limitations of established denoising workflows, we propose an innovative computational denoising tool that leverages the power of a U-net neural network. Unlike conventional methods that rely solely on pixel intensity, our approach employs a two-part symmetric architecture to capture the contextual information of the image. It thus robustly identifies the signal's location, enhancing denoising capabilities. To train our model, we used a diverse set of tissue (ie: human Skin, Tonsil, and Colon) including both histologically healthy and pathologic tissue (ie: Tumors, inflammation). Additionally, we incorporated a comprehensive panel of over 30 measured features. The output of our model is a binary mask that effectively distinguishes signal from noise for each feature. By leveraging our tool, we can efficiently clean up all measured features, enhance the quality of IMC data and eliminate the need for time-consuming manual tasks.

To evaluate the performance of the developed tool, we used, on every feature, the Intersection over Union (IoU) between the predicted signal areas and those found following a standard semi-automatic procedure. Our model achieved an average IoU score of 0.502 with an independent test set. This promising result, coupled with visually appealing outcomes, is an indication of the effectiveness of our tool, its robustness and generalizability. Moreover, the obtained improvements facilitate the analysis and comparison of multiple samples and allow the normalization of the dataset for the subsequent steps in the image analysis pipeline.

*Speaker
†Corresponding author: tiberti@amkbiotech.fr

1 Giesen C, et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. Nat Methods (2014) 11:417–22.

2 Jackson HW, et al. The single-cell pathology landscape of breast cancer. Nature (2020) 578:615–20.

3 Elaldi, R. et al. "High dimensional Imaging Mass Cytometry™ panel to visualize the tumor immune microenvironment contexture." Frontiers in Immunology 12 (2021): 666233.
4 Ijsselsteijn, ME, Somarakis, A, Lelieveldt, BPF, Höllt, T, de Miranda, NFCC. Semi-automated background removal limits data loss and normalizes imaging mass cytometry data. Cytometry. 2021; 99: 1187– 1197. https://doi.org/10.1002/cyto.a.24480

Keywords: imaging mass cytometry, denoising, image analysis, U, NET

SalienceNet: an unsupervised Image-to-Image translation method for nuclei saliency enhancement in microscopy images

Emmanuel Bouilhol * [1], Edgar Lefevre [2], Barry Thierno [3], Florian Levet [3], Anne Beghin [4], Virgile Viasnoff [4], Xareni Galindo [3], Rémi Galland [3], Jean-Baptiste Sibarita [3], Macha Nikolski [2,5]

[1] ibv – Université Côte d'Azur (UCA), 3IA, iBV, Institut National de la Santé et de la Recherche Médicale - INSERM – France
[2] Centre de Bioinformatique de Bordeaux – CGFB – 146, rue Léo Saignat, 33076 Bordeaux, France
[3] Interdisciplinary Institute for Neuroscience [Bordeaux] – Université de Bordeaux, Centre National de la Recherche Scientifique – Institut des neurosciences/Bordeaux Neurocampus - Université de Bordeaux (Campus Carreire)- 38 Rue Albert Marquet, 33000 Bordeaux, France
[4] Mechanobiology Institute [Singapore] – 5A Engineering Drive 1Singapore 117411, Singapore
[5] Institut de biochimie et génétique cellulaires – Université Bordeaux Segalen - Bordeaux 2, Centre National de la Recherche Scientifique – 1 Rue Camille Saint-Saens 33077 BORDEAUX CEDEX, France

Cell nuclei segmentation is crucial for various applications, including cell detection, counting, tracking, morphology analysis, and quantification of molecular expression. Achieving accurate automatic segmentation is particularly important in high-throughput microscopy imaging, where it serves as the initial step for downstream quantitative data analysis. The precision of nuclei segmentation significantly influences the quality of downstream quantitative analyses, making it essential for drawing meaningful biological conclusions.
Automating this process presents challenges due to diverse image characteristics influenced by biological and acquisition conditions. Such aspects as morphological differences between nuclei from different tissues, heterogeneity of intensity and texture, variation in spatial organization such as the presence of both sparse or dense images with touching nuclei, or imaging artifacts (e.g., low signal-to-noise ratio or out-of-focus signal)(1). This results in the necessity to fine-tune numerous parameters between different image acquisitions, or even between individual images.

Recent advancements in deep learning tools, such as Cellpose(2) and StardDist(3), have alleviated the need to select specific parameters. However, despite these methodological breakthroughs, there is no universal combination of methods and parameters capable of automatically performing nuclei segmentation across all images due to the aforementioned heterogeneity in biological samples and technical artifacts(4). Particularly, live-cell imaging poses a challenge for these techniques, as the low light levels result in low signal-to-noise ratio and artifacts.

In this work, instead of focusing on the segmentation process itself, we propose to tackle this problem by enhancing the nuclei prior to segmentation. This approach simplifies the task for classical nuclei segmentation tools. Leveraging recent advancements in unsupervised generative adversarial networks, we utilize image translation techniques to enhance nuclei saliency in microscopy images. Our aim is to achieve strong signal differences between nuclei and the back-

*Speaker

ground, facilitating straightforward segmentation.

We introduce SalienceNet, a novel unsupervised Deep Learning-based approach for nuclei saliency enhancement in microscopy images. SalienceNet does not require image annotation when training on new data with different characteristics. We demonstrate its efficacy in enhancing organoid images acquired under low light conditions, without the need for prior annotation. SalienceNet adapts the domain style transfer framework to this specific task, enabling automatic nuclei segmentation without extensive annotation. Our approach employs a ResNet-based CycleGAN trained with a customized loss function dedicated to nuclei enhancement. The intensity of nuclei, particularly their borders, is made more salient, independent of contrast, intensity, textures, or shapes present in the input data.

Furthermore, we evaluate the impact of nuclei enhancement on downstream segmentation using conventional methods. We demonstrate that incorporating SalienceNet into a standard segmentation pipeline eliminates the need for manual parameter fine-tuning. Our experiments reveal that SalienceNet successfully increases the saliency of low-light microscopy images to the desired level. We assess its impact on segmentation using Otsu thresholding and StarDist, showing that nuclei enhancement with SalienceNet improves segmentation results by 30% using Otsu thresholding and 26% using StarDist in terms of Intersection over Union (IOU), compared to segmentation on non-enhanced images. These findings highlight the potential of SalienceNet as a preprocessing step to automate nuclei segmentation pipelines for low-light microscopy images.

(1) Zhou, Yanning, et al. "Cia-net: Robust nuclei instance segmentation with contour-aware information aggregation." 26th International Conference, IPMI 2019, Hong Kong, China, June 2–7, 2019, Proceedings 26. Springer International Publishing, 2019.

(2) Stringer, Carsen, et al. "Cellpose: a generalist algorithm for cellular segmentation." Nature methods 18.1 (2021): 100-106.

(3) Schmidt, Uwe, et al. "Cell detection with star-convex polygons." Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11. Springer International Publishing, 2018.

(4) Hollandi, Reka, et al. "Nucleus segmentation: towards automated solutions." Trends in Cell Biology (2022).

Keywords: Deep Learning, Bioimaging, Image Processing, Microscopy, Automatic segmentation, GAN

Online Federated Learning with Mixture Models

Othmane Marfoq * [1,2]

[1] Inria Sophia Antipolis – Inria Sophia Antipolis - Méditerranée, Inria Sophia Antipolis - Méditerranée – France
[2] Accenture Technology Labs – Accenture – Sophia Antipolis, France

Federated Learning (FL) (1) aims at collaboratively training machine learning models over remote devices/clients while keeping their data local due to privacy concerns or limited communication resources. Most previous works on federated learning assume that every client collects and stores all the samples before training starts. Learning on static datasets can be sub-optimal in many cases, because new samples collected during training are ignored, and clients may have limited memory capacities, and cannot store a large number of data samples.
Recent works (2-7) relaxe the assumptions of offline federated learning and explore online federated learning from data streams. However, these works have various limitations and fail to capture the full potential of collaboration. Specifically, (2) assumes independent samples from a fixed unknown distribution. While this assumption is suitable for certain scenarios, it does not hold in many real-world applications. On the other hand, (3) eliminates any assumptions about the data stream and considers a worst-case scenario where an adversary provides clients with the worst instances at each time-step. Although this approach accounts for adversarial scenarios, it overlooks the significant benefits that collaboration can offer. In a recent work (4), authors addressed continual federated learning, introducing a regularization-based method to minimize interference between tasks and enable knowledge transfer. However, they lacked theoretical learning guarantees. Other works (5, 6) focused on temporally shifting distributions; (5) used a block-cyclic pattern, and (6) used a mixture model with daytime-nighttime shifts, but with strong temporal priors and limited theoretical guarantees. While (7) used clustering for drift adaptation, limiting knowledge transfer to clusters, our soft-clustering approach allows each client to benefit from all other clients' datasets. Our framework extends (7) and provides theoretical learning guarantees unlike (7).

In this work, we provide a novel formulation for the problem of online federated learning based on the assumption that clients' data distributions are mixtures of a finite number of unknown underlying distributions with varying mixing weights. In comparison to previous work, our assumption allows the clients to provably benefit from collaboration, while allowing clients' data distributions to vary in a (constrained) adversarial manner.

We propose Federated Expectation-Maximization Online Mirror Descent (FEM-OMD), a federated variant of the OMD algorithm, where the gradient of the cost function is estimated through an EM-like algorithm at each time-step. FEM-OMD leverages all of the data stored across clients to learn the parameters of the underlying distributions using Expectation-Maximization updates, while enabling each client to adapt to the temporal variation of its data distribution. We analyze the regret guarantees of FEM-OMD in the case of well-separated spherical Gaussian mixture models. Specifically, we establish a $O(\sqrt{T}\log(m) + T/\sqrt{n})$ regret bound, where T is the time horizon, m is the number of the underlying distributions, and n is the number of

*Speaker

samples received by each client.

Through experimental results on synthetic datasets and FL benchmarks, we demonstrate the effectiveness of our approach in online federated settings and show that our scheme allows the clients to benefit from collaboration. We believe that our work opens up new directions for research in online federated learning, where clients' data distributions are allowed to vary in constrained adversarial manners, and we hope that our proposed algorithm will pave the way for further improvements in this field.

References

(1) McMahan et al. "Communication-efficient learning of deep networks from decentralized data." AISTATS 2017.
(2) Chen et al. "Asynchronous online federated learning for edge devices with non-iid data." IEEE International Conference on Big Data 2020.
(3) Mitra et al. "Online federated learning." IEEE CDC 2021.
(4) Yoon et al. "Federated continual learning with weighted inter-client transfer." ICML 2021.
(5) Eichner et al. "Semi-cyclic stochastic gradient descent." CML 2019.
(6) Zhu et al. "Diurnal or nocturnal? federated learning of multi-branch networks from periodically shifting distributions." ICLR 2022.
(7) Jothimurugesan et al. "Federated learning under distributed concept drift." AISTATS 2023.

Keywords: Federated learning, Online learning, Mixture models

Trustworthy ModSecurity: Countering Adversarial SQL Injections with Robust Machine Learning

Biagio Montaruli * [1,2], Battista Biggio [3], Luca Demetrio [4], Luca Compagna [5], Andrea Valenza [6], Davide Balzarotti [2]

[1] SAP Security Research – SAP Labs France – 805 avenue Maurice Donat, 06250 MOUGINS, France
[2] Eurecom [Sophia Antipolis] – EURECOM – Campus SophiaTech, 450 Route des Chappes, CS 50193 - 06904 Biot Sophia Antipolis cedex, France
[3] University of Cagliari – Italy
[4] University of Genoa – Italy
[5] SAP Security Research – SAP Labs France – 805 Avenue du Docteur Maurice Donat, BP 1216, 06254 Mougins Cedex, France
[6] Prima Assicurazioni – Italy

ModSecurity is widely recognized as the standard open-source Web Application Firewall (WAF), maintained by the OWASP Foundation. It detects malicious requests by matching them against the Core Rule Set (CRS), identifying well-known attack patterns.
Each rule in the CRS is manually assigned a weight, based on the severity of the corresponding attack, and a request is detected as malicious if the sum of the weights of the firing rules exceeds a given threshold. In this work, we show that this simple strategy is largely ineffective for detecting SQL injection (SQLi) attacks, as it tends to block many legitimate requests, while also being vulnerable to adversarial SQLi attacks, i.e., attacks intentionally manipulated to evade detection.
To overcome these issues, we design a robust machine learning model, named TrustModSec, which uses the CRS rules as input features, and it is trained to detect adversarial SQLi attacks. Our experiments show that TrustModSec, being trained on the traffic directed towards the protected web services, achieves a better trade-off between detection and false positive rates, improving the detection rate of the vanilla version of ModSecurity with CRS by 21%.
Moreover, our approach is able to improve its adversarial robustness against adversarial SQLi attacks by 42%, thereby taking a step forward towards building more robust and trustworthy WAFs.

Keywords: trustworthy AI, adversarial machine learning, web application firewalls

*Speaker

Mapping of Vulnerability Advisories onto their Fix Commits Using Heuristics and NLP approaches

Antonino Sabetta , Rocio Cabrera Lozoya * [1], Daan Hommersom , Dario Di Nucci , Bonaventura Coppola , Damian A. Tamburri

[1] SAP Labs France – SAP Labs France – 805 Avenue du Docteur Maurice Donat, BP 1216, 06254 Mougins Cedex, France

The study and understanding of software vulnerabilities and their corrections face a significant challenge due to the lack of comprehensive and accurate sources. We present an approach that combines heuristics stemming from practical experience and natural language processing (NLP) to address this problem. Our method consists of three phases: First, an advisory record containing key information about a vulnerability is extracted from an advisory (expressed in natural language). Second, using heuristics, a subset of candidate fix commits is obtained from the source code repository of the affected project by filtering out commits that are known to be irrelevant for the task at hand. Finally, for each such candidate commits, our method builds a numerical feature vector reflecting the characteristics of the commit that are relevant to predicting its match with the advisory at hand. The feature vectors are then exploited for building a final ranked list of candidate fixing commits. The score attributed by the ML model to each feature is kept visible to the users, allowing them to interpret the predictions. To evaluate the effectiveness of our approach, we conducted experiments on a carefully curated dataset consisting of 2,391 known fix commits corresponding to 1,248 public vulnerability advisories. Our results demonstrate the efficacy of our method in identifying fix commits for vulnerabilities. When considering the top-10 commits in the ranked results, our implementation successfully identified at least one fix commit for up to 84.03% of the vulnerabilities. Moreover, in 65.06% of the cases, the correct fix commit occupied the first position in the ranked list. In conclusion, our approach significantly reduces the effort required to search open-source software repositories for the specific commits that address known vulnerabilities. By combining practical heuristics with NLP techniques, we enhance the accuracy and efficiency of vulnerability analysis. Our method not only aids researchers in studying software vulnerabilities but also provides practical value to developers and users seeking to identify and mitigate vulnerabilities effectively.

Keywords: Open Source Software, Software Security, Common Vulnerabilities and Exposures (CVE), National Vulnerability Database (NVD), Mining Software Repositories, Code, level Vulnerability Data

*Speaker

Geometric and statistical analysis of the extracellular matrix

Faisal Jayousi * [1], Laure Blanc-Féraud [2], Xavier Descombes [3], Ellen Van Obberghen-Schilling [4]

[1] Université Côte d'Azur – Université Côte d'Azur, Inria, CNRS, I3S – France
[2] Université Côte d'Azur – Université Côte d'Azur, Inria, CNRS, I3S – France
[3] Université Côte d'Azur – Université Côte d'Azur, Inria, CNRS, I3S, France – France
[4] Université Côte d'Azur – iBV, Institut National de la Santé et de la Recherche Médicale - INSERM – France

The extracellular matrix (ECM) functions as the architectural scaffold of organs and tissues, providing a dynamic milieu of physical and biochemical signals to cells. Throughout tumour progression, the ECM undergoes pronounced topological and biophysical transformations. Our research aims to investigate the geometric and topological properties of the tumour microenvironment, with the goal of gaining insights into disease progression and identifying and characterising biomarkers that can effectively predict the efficacy of immunotherapy. By employing a combination of classical image processing techniques and deep learning approaches, we can extract quantitative information regarding the phenotypes and spatial distribution of cells within the tumour microenvironment. While the study of cellular components has been extensively explored, the characterisation of non-cellular elements, particularly the ECM, remains an underexplored domain. Our current focus lies in the geometric and topological analysis of the ECM utilising multispectral fluorescence imaging. Ultimately, our objective is to integrate ECM attributes into prediction models for immunotherapy outcomes.
It is hypothesised that fibronectin (FN), a key component of the tumour ECM landscape, can manifest in three discernible forms: aligned fibers, reticular-fiber-like structures, and aggregates. Aligned fibers pertain to FN structures characterised by a highly-organised and linear arrangement, exhibiting a distinct orientation within the tissue. In contrast, reticular-fiber-like structures resemble a network-like configuration of FN. Lastly, FN aggregates denote accumulations of FN molecules that exhibit a dense and compact morphology within the tissue.

Gaining a comprehensive understanding of the various forms of FN and their interactions with immune cells is crucial, as it can yield valuable insights into spatial ECM features and how they impact immunotherapy. In the scope of our research, our objective is to segment fibronectin images into the three aforementioned classes. This segmentation will facilitate quantitative analysis and characterisation of each form, allowing for a comprehensive assessment of the functional implications of FN. By achieving this, we aim to contribute to the broader understanding of FN's role and its potential implications in the context of tumour progression.

In the context of FN segmentation, we provide an overview of *some* of the methods utilised:

1. Deep Learning (AlexNet)

AlexNet is a convolutional neural network architecture that achieved groundbreaking results in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012. In transfer learn-

*Speaker

ing for segmentation, the pre-trained AlexNet model is used as an encoder in order to extract meaningful features from images. The SVM (Support Vector Machine) algorithm is a popular choice for classification tasks, and in this context, it can be employed to classify the extracted features into different segmentation classes. By utilising SVM on the extracted features, the model can make precise and detailed segmentation predictions based on the learned representations.

2. Gray Level Co-occurrence Matrix

The gray level co-occurrence matrix (GLCM) is a matrix with dimensions of L x L, where L represents the number of gray levels within an image. The GLCM captures the statistical relationship between pairs of pixels concerning their intensities and spatial positions. It quantifies the frequency of occurrence of various combinations of gray-level values at specified pixel distances and directions. By analyzing the GLCM, texture features such as contrast, homogeneity, and correlation can be derived. These features provide valuable information about the texture characteristics of the image.

In the context of FN segmentation, the GLCM texture features can be utilised to train a supervised model, such as an SVM. The model learns from labelled FN images, extracting and quantifying texture features from the GLCM. Once the model is trained, it can be applied to classify pixels within new images into one of the FN classes discussed earlier.
While we have presented two approaches here, it is important to note that there are additional approaches that will also be analysed in our study. The preliminary findings demonstrate excellent performance for AlexNet with SVM, surpassing AlexNet with its original classifier in the output layer and outperforming the co-occurrence matrices. However, it is worth noting that the co-occurrence matrices may provide valuable geometric information that deserves further exploration.

Quantifying the uncertainty of algorithms handling missing values with a conformal procedure

Andjela Todorovic * [1], Pierre-Alexandre Mattei [1], Aude Sportisse† [1]

[1] L'Institut National de Recherche en Informatique et en Automatique (INRIA) – L'Institut National de Recherche en Informatique et e n Automatique (INRIA) – France

Uncertainty estimation plays a vital role in assessing the trustworthiness and reliability of predictive models. It enables us to quantify the level of confidence associated with the model's predictions. In certain application domains, particularly those pertaining to medicine (1), the outcomes of erroneous predictions can be substantial. These domains often encounter significant challenges due to the prevalence of missing values, further emphasizing the need for accurate uncertainty estimation techniques.
Dealing with data that contains missing values brings challenges to the uncertainty estimation due to information loss, biased inferences, and reliance on imputation assumptions. Incomplete data is oftentimes limiting the model's ability to accurately estimate uncertainty and affecting the reliability of predictions.

Various algorithms have been developed to quantify missing data, including Conformal Prediction, Multiple Imputation and Maximum Likelihood Estimation (MLE). In our work, we focus on conformal prediction, as a straightforward way to create predictions for any model (2). Beginning with a fitted predicted model f it creates prediction sets for this model using a small amount of additional calibration data, by constructing a prediction set of possible labels C(X). Prediction set has the property of marginal coverage, ie. the probability that the prediction set contains the correct label is almost exactly 1 - alpha.

Conformal prediction has shown to be effective in determining confidence intervals, but still, when missing data is present, traditional conformal prediction methods may not perform optimally due to the inherent challenges of missing values. To overcome this, we propose an approach that combines the robustness of missForest algorithm() for the imputation of the data, and weighted quantile regression for determining the confidence interval and the conditional quantiles. (3) By integrating these two methods, we aim to develop an algorithm that effectively estimates uncertainty in the presence of missing data, improving the reliability and usefulness of conformal predictions.

We explore the benefits and limitations of different strategies and provide practical recommendations for implementing conformal prediction in the presence of missing values. Our findings contribute to enhancing the applicability and effectiveness of conformal prediction in real-world scenarios with missing data.

References:

*Speaker
†Corresponding author: aude.sportisse@inria.fr

(1) Ibrahim JG, Chu H, Chen MH. Missing data in clinical studies: issues and methods. J Clin Oncol. 2012 Sep 10;30(26):3297-303. doi: 10.1200/JCO.2011.38.7589.

(2) Angelopoulos, A. N., & Bates, S. (2021). A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification.
(3) Stekhoven, D. J., & Bühlmann, P. (2012). MissForest-Non-parametric missing value imputation for mixed-type data. Bioinformatics, 28(1), 112-118.

Keywords: conformal prediction, theoretical machine learning, uncertainty, missing data

Deep Decision-theoretic Anomaly Detection for Time Series

Nour Jamoussi [*†] [1,2], Alix Lhéritier [* ‡] [1], Nicolas Bondoux [1], Rodrigo Acuna Agost [1], Maurizio Filippone [2]

[1] Amadeus – Amadeus – Nice, France
[2] Eurecom [Sophia Antipolis] – EURECOM – Campus SophiaTech, 450 Route des Chappes, CS 50193 - 06904 Biot Sophia Antipolis cedex, France

Time series anomaly detection is a crucial task in various domains including network security, financial fraud detection, system monitoring, Internet-of-Things device monitoring, predictive maintenance, and healthcare monitoring. Its wide range of applications highlights its importance in identifying abnormal patterns and deviations in time-dependent data to ensure system reliability and security (see, e.g., (1)).

Model-based unsupervised anomaly detection methods, an important category of techniques, include two important subcategories: predictive and reconstructive. Reconstructive methods aim to learn and reconstruct the input data, utilizing the reconstruction error as an anomaly score. An example of reconstructive model is based on reconstructing the input data from the Principal Component Analysis (PCA) projection (2). On the other hand, predictive methods are based on predicting the subsequent sequence or symbol in a time series, typically employing the prediction error as an anomaly score.

In this work, we propose a decision-theoretic approach to anomaly detection. A minimax strategy takes the decision that minimizes the regret —the increase in loss with respect to the optimal decision in hindsight— for the worst-case scenario. In a supervised setup, the decisions can be probability assignments on the target value given some features. The predictive Normalized Maximum Likelihood distribution (3) is a minimax strategy for the logarithmic loss in a supervised setup and yields a constant regret whatever the observed target is. When the class of possible decisions is restricted to Gaussian linear regression models, analytical expressions for the minimax strategy and its corresponding regret can be obtained (4).

The minimax regret has been shown to be a powerful score for out-of-distribution detection in an image classification setup (5). In the case of linear regression models, the minimax regret is low when the features of the test sample lie mostly in the subspace spanned by the principal directions of the training data and high otherwise (4). Although the regret is defined in predictive terms, in the Gaussian linear regression case, it depends only on the features and is connected to PCA, thus bridging predictive and reconstructive worlds.

As a key contribution, we propose minimax regret scores for anomaly detection on time series, based on feature transformations given by different deep neural network architectures. Our experimental results demonstrate that our approach yields competitive performance when compared to state-of-the-art methods on standard anomaly detection benchmarks.

*Speaker
†Corresponding author: nour.jamoussi@eurecom.fr
‡Corresponding author: alix.lheritier@amadeus.com

REFERENCES

(1) Choi, K., Yi, J., Park, C., & Yoon, S. (2021). Deep learning for anomaly detection in time-series data: review, analysis, and guidelines. IEEE Access, 9, 120043-120065.
(2) Shyu, M. L., Chen, S. C., Sarinnapakorn, K., & Chang, L. (2003). A novel anomaly detection scheme based on principal component classifier. Miami Univ Coral Gables Fl Dept of Electrical and Computer Engineering.
(3) Fogel, Y., & Feder, M. (2018). Universal supervised learning for individual data. arXiv preprint arXiv:1812.09520.
(4) Bibas, K., Fogel, Y., & Feder, M. (2019). A new look at an old problem: A universal learning approach to linear regression. In 2019 IEEE International Symposium on Information Theory (ISIT) (pp. 2304-2308). IEEE.
(5) Bibas, K., Feder, M., & Hassner, T. (2021). Single layer predictive normalized maximum likelihood for out-of-distribution detection. Advances in Neural Information Processing Systems, 34, 1179-1191.

Keywords: Time Series, Unsupervised Anomaly Detection, Minimax Theory, Normalized Maximum Likelihood

Enhancing Landfill Detection through Multi-Spectral Satellite Imagery and Deep learning Techniques

Noelia Abascal Zorrilla *† [1], Yoann Rey-Ricord [1], Julie Remenant [1], Anne Vallete [1], François-Régis Martin-Lauzer [1]

[1] Argans – ARGANS Limited – France

The monitoring and management of dumpsites have emerged as crucial components of environmental governance worldwide. Swift identification and appropriate disposal of these sites are essential to prevent further contamination and protect ecological integrity (Sun et al., 2023). However, the timely acquisition of dumpsite location information remains challenging for local government agencies and environmental groups.
In recent years, the combination of remote sensing techniques and deep learning algorithms has shown considerable promise in addressing diverse environmental challenges. This study focuses specifically on investigating the efficacy of Convolutional Neural Networks (CNNs) models in comparison to Conditional Generative Adversarial Networks (cGANs) for the detection of dumpsites in satellite imagery, utilizing Sentinel 2 data (BigEarthNet, (Charfuelan et al., 2019). Sentinel 2 is a satellite within the Copernicus program, offering high-resolution imagery and data that is freely available, making it a cost-effective solution for monitoring dumpsites worldwide. The results of this study demonstrate the promising potential of using deep learning methods for detecting dumping sites from Sentinel 2 satellite imagery through sampled areas around the world.

By offering a more efficient and cost-effective approach to monitoring and identifying potential dumping sites, this technology can contribute to sustainable environmental management and public health promotion. Additionally, the utilization of freely accessible Sentinel 2 data from the Copernicus program enables worldwide monitoring at a low cost, facilitating global efforts in addressing the challenges associated with dumpsite management.

Sun, X., Yin, D., Qin, F. et al. Revealing influencing factors on global waste distribution via deep-learning based dumpsite detection from satellite imagery. Nat Commun 14, 1444 (2023). https://doi.org/10.1038/s41467-023-37136-1
Charfuelan, M., Demir, B., Markl, V., and Sumbul, G.. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. pp. 5901–5904, 2019.

Keywords: dumping sites, detection, Sentinel, 2 data, CNN, cGAN

*Speaker
†Corresponding author: nzorrilla@argans.eu

SparseGEMINI: joint discriminative clustering and feature selection

Louis Ohl * [1,2,3]

[1] Université côte d'azur – Université Côte d'Azur (UCA) – France
[2] Equipe Maasai, Inria Sophia Antipolis – Institut National de Recherche en Informatique et en Automatique – France
[3] Université Laval [Québec] – 2325, rue de lÚniversité Québec G1V 0A6, Canada

Clustering is a fundamental learning task which consists in separating data samples into several categories, each named cluster. This task hinges on two main questions concerning the assessment of correct clustering and the actual number of clusters that may be contained within the data distribution.

The most common approaches for clustering rely on a generative modelling of the clusters (1): the cluster assignment is a latent variable which explains the observed data. Here, we focus on inferring clusters directly from the data, that is the discriminative setting (2). By modelling the cluster distribution thanks to (deep) neural networks, this setting brings the question of training neural network without labels, i.e. deep clustering.

In the last decade, recent successes in deep clustering majorly involved the mutual information (MI) as an unsupervised objective for training neural networks with increasing regularisations. While the quality of the regularisations have been largely discussed for improvements (3), little attention has been dedicated to the relevance of mutual information as a clustering objective. We first highlight how the maximisation of mutual information does not lead to satisfying clusters by identifying the Kullback-Leibler divergence as the main reason of this behaviour. Hence, we generalise the mutual information by changing its core distance, introducing the generalised mutual information (GEMINI) (4): a set of metrics for unsupervised neural network training. Unlike mutual information, some GEMINIs do not require regularisations when training. Some of these metrics are geometry-aware thanks to distances or kernels in the data space. Finally, we highlight that GEMINIs can automatically select a relevant number of clusters, a property that has been little studied in deep clustering context where the number of clusters is a priori unknown.

Moreoever, the discriminative formulation of clustering brings a different focus on difficult joint tasks. In this sense, we present SparseGEMINI, an extension of GEMINI dedicated to feature selection. Indeed, feature selection in clustering is a hard task which involves simultaneously the discovery of relevant clusters as well as relevant variables with respect to these clusters. Contrary to model-based algorithms that would achieve selection through optimised model selection or strong assumptions on p(x), sparseGEMINI just requires a simple L1 penalty (5). This algorithm avoids the burden of combinatorial feature subset exploration and is easily scalable to high-dimensional data and large amounts of samples while only designing a clustering model p(y—x). We demonstrate the performances of Sparse GEMINI on synthetic datasets as well as large-scale datasets. Our results show that Sparse GEMINI is a competitive algorithm and has the ability to select relevant subsets of variables with respect to the clustering without using

*Speaker

relevance criteria or prior hypotheses.

(1) Bouveyron, C., Celeux, G., Murphy, T. B., & Raftery, A. E. (2019). Model-based Clustering and Classification for Data Science: With applications in R. Cambridge University Press.
(2) Minka, T. (2005). Discriminative Models, not Discriminative Training. Technical Report MSR-TR-2005-144, Microsoft Research.
(3) Tschannen, M., Djolonga, J., Rubenstein, P. K., Gelly, S., & Lucic, M. (2020). On Mutual Information Maximization for Representation Learning. International Conference on Learning Representations.
(4) Ohl, L., Mattei, P.-A., Bouveyron, C., Harchaoui, W., Leclercq, M., Droit, A., & Precioso, F. (2022). Generalised Mutual Information for Discriminative Clustering. Advances in Neural Information Processing Systems.
(5) Hastie, T., Tibshirani, R., & Wainwright, M. (2015). Statistical Learning with Sparsity. Monographs on Statistics and Applied Probability, 143, 143.

Keywords: clustering, deep learning, neural networks, sparsity, discriminative models

Data fusion of remote sensing observations for sustainable development

Pierre-Laurent Cristille *† [1], Jeronimo Bernard-Salas *

[1], Nick Cox [1], Emmanuel Bernhard [1], Antoine Mangin [1]

[1] ACRI-ST – ACRI-ST – 260 route du Pin Montard 06904 Sophia-Antipolis, France

The European Copernicus programme is the most ambitious Earth observation program in the world, providing satellite data to improve environmental management, understand and mitigate the effects of climate change, and ensure civil security. To achieve this, the European Space Agency (ESA) is developing and operating the Sentinels satellites, which are equipped with a wide range of instruments for monitoring the Earth, oceans and atmosphere. These observations and the vast amount of data they generate play a major role in supporting and monitoring the 17 Sustainable Development Goals (SDGs), including many of its targets. This information is freely available and accessible to users, including service providers, public authorities, and international organizations.

A main challenge for the exploitation of Copernicus data is to combine data from multiple sources. Data fusion can enhance the accuracy and reliability of the resulting information by producing high-quality images and spectra. While many traditional methods have been developed in the context of low-resolution and hyper-spectral images (LR-HSI) and high-resolution and multi-spectral images (HR-MSI) fusion, machine learning and deep learning approaches offer new methods to address this challenge by optimizing performance criteria based on statistical analysis using examples or past experiences.

In recent years, the attention mechanism (via the transformer model, Vaswani et al. 2017) has obtained outstanding achievements in the field of natural language processing and computer vision, as such, transformer-based network architecture is a good candidate for the fusion of LR-HSI and HR-MSI. In this talk we will present our our methodology for the fusion of the Sentinel 2 data (with 10-60m spatial resolution in 12 broad band filters) with Sentinel 3 data (with 300m spatial resolution and 22 narrow filters).

We have investigated two architectures that utilize (cross-) attention mechanisms. The first is the Coupled Unmixing Network with Cross-Attention (CUCaNet, Yao et al. 2020), and the second is a Transformer-based Fusion Approach for Hyperspectral Image Super-resolution (FUSFORMER, Hu et al. 2022). The figure illustrates the results obtained with CUCaNet, where the reconstructed HR-HSI image (right panel) retains the spatial information of the HR-MSI (left panel), and spectral information of the LR-HSI (middle panel). These results show that data fusion can drastically improve the quality of remote sensing data. This is promising for a more comprehensive and accurate understanding of the Earth and its processes to support the achievement of the SDGs.

*Speaker
†Corresponding author: pierre-laurent.cristille@acri-st.fr

References:

- Jin-Fan Hu, Ting-Zhu Huang, Liang-Jian Deng, Hong-Xia Dou, Danfeng Hong, and Gemine Vivone. Fusformer: A transformer-based fusion network for hyperspectral image super-resolution. IEEE Geoscience and Remote Sensing Letters, 19:1–5, 2022.
- Jing Yao, Danfeng Hong, Jocelyn Chanussot, Deyu Meng, Xiaoxiang Zhu, and Zongben Xu. Cross-attention in coupled unmixing nets for unsuper(1)vised hyperspectral super-resolution. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceed(1)ings, Part XXIX 16, pages 208–224. Springer, 2020
- VASWANI, Ashish, SHAZEER, Noam, PARMAR, Niki, et al. Attention is all you need. Advances in neural information processing systems, 2017, vol. 30.

Keywords: Data fusion, Deep learning, Earth observations, Copernicus, Sustainability

Amadeus Ancillaries Dynamic Pricing: an AI approach

Riccardo Petitti *† [1], Hugo Alexander Gonzalez Reyes * ‡ [1], Bettina Alexieva * § [1], Norbert Leon¶ [1], Davide Sanapo‖ [1], Gerardo Ayala Solano [1]

[1] Offer Data Engineering, Science and Innovation – Amadeus SAS – 485 Rte du Pin Montard, 06410 Biot, France

The airline industry is characterized by a high susceptibility to economic downturns, slim margins, and downward pressure on fares. To address these challenges and hedge against an unpredictable business climate, airline boardrooms are increasingly focusing on diversifying revenue streams and exploring new commercial opportunities. Ancillary revenues represent one such important shift in the way the industry operates – from all-inclusive bundled fares to base fares to which desired products are added according to the context (e.g. seat choice, wi-fi or meal on board...) Through this "a la carte" business model airlines aim to increase profit margins by influencing the right differentiated purchases at the right time. Improving digital and analytical capabilities is therefore now more important than ever, with Artificial Intelligence becoming the paramount driver of this paradigm shift and an enabler of the evolution towards contextualized and personalized service offering.
Ancillary Dynamic Pricing is Amadeus Offer Management's response to this new era of airline operations, generating tailored offers and dynamically optimized prices through an "Earn-while-Learn" model, which enables the exploration of new price domains, while learning in parallel to generate optimal prices from customer purchasing insights.

In this talk, we will briefly review our AI models, then we will show the challenges to deploy in production an "AI-based" ancillaries dynamic pricing model. We will discuss hybrid architecture (On Premise – Cloud), Data concerns, Model robustness, and lifecycle.

References

Naman Shukla, Arinbjörn Kolbeinsson, Ken Otwell, Lavanya Marla, Kartik Yellepeddi:Dynamic Pricing for Airline Ancillaries with Customer Context. KDD 2019: 2174-2182
Fiig and All., Forecasting and Optimization of fare families, Journal of Revenue and Pricing Management, 2011, Vol. 11, 3, 322–342

Fiig and All., Dynamic pricing of airline offers, Journal of Revenue Management 2018

*Speaker
†Corresponding author: riccardo.petitti@amadeus.com
‡Corresponding author: hugoalexander.gonzalezreyes@amadeus.com
§Corresponding author: bettina.alexieva@amadeus.com
¶Corresponding author: norbert.leon@amadeus.com
‖Corresponding author: davide.sanapo@amadeus.com

Keywords: Revenue management, Travel Industry, Artificial Intelligence, Earn, while, Learn, Dynamic Pricing, Hybrid environment

Application of quantum computing in Revenue Management

Antoine Boulanger *† [1], Gerardo Ayala Solano [1]

[1] Offer Data Engineering, Science and Innovation – Amadeus SAS – 485 Rte du Pin Montard, 06410 Biot, France

With the massive development of quantum technologies, which promise to dramatically reduce computation time and solve highly combinatorial problems that are out of reach for classical computers, many industries have taken an interest into it and thought of use cases. It is especially true for some fields such as drug discovery, finance and artificial intelligence. We, then, decided to investigate in some use cases of Airline Revenue Management that could potentially benefit from quantum advantage.

In this study, application of quantum computing in Revenue Management is evaluated, both considering the quantum devices that exist today as well as assessing the potential in the future with an increasing number of qubits and reduced error rates. We develop and implement a quantum algorithm for solving the single flight leg optimization problem by mapping the Bellman equation to an Ising model (1) (spin model used to describe magnetism). This allows execution on existing hardware from D-Wave Systems. The results are limited to toy scenarios, but still demonstrate the state-of-the-art of quantum computing. Further, we develop a quantum algorithm for solving a Choice-Based Deterministic Linear Program (CDLP) (2) for network optimization with customer choice. CDLP, although expected to be superior to existing approaches, is far too complex for existing classical computers. Utilizing a quantum computer, we obtain a quadratic speed up, which allow for more realistic problem sizes. Finally, because we have also witnessed an upheaval in the use of AI in solution development, in particularly in the Revenue Management field (3), we discuss the ways and improvements we could expect from applying quantum computing within the AI field (4).

References:

(1) F. Glover, G. Kochenberger, and Y. Du, "A tutorial on formulating and using qubo models," arXiv preprint arXiv:1811.11538, 2018.

(2) Q. Liu and G. Van Ryzin, "On the choice-based linear programming model for network revenue management," Manufacturing & Service Operations Management, vol. 10, no. 2, pp. 288–310, 2008.

(3) N. Bondoux, A. Q. Nguyen, T. Fiig, and R. Acuna-Agost, "Reinforcement learning applied to airline revenue management," Journal of Revenue and Pricing Management, vol. 19,

*Speaker

†Corresponding author: antoine.boulanger@amadeus.com

no. 5, pp. 332–348, 2020.

(4) M. Cerezo, G. Verdon, H.-Y. Huang, L. Cincio, and P. J. Coles, "Challenges and opportunities in quantum machine learning," Nature Computational Science, vol. 2, no. 9, pp. 567–576, 2022.

Keywords: Quantum Computing, Revenue Management, NP, Hard, Choice Model, Travel Industry, Artificial Intelligence

Generative AI in Amadeus: The Engineering Perspective

Norbert Leon [*†] [1], Francois Hantry [* ‡] [1], Gerardo Ayala Solano[§] [1]

[1] Offer Data Engineering, Science and Innovation – Amadeus SAS – 485 Rte du Pin Montard, 06410 Biot, France

From ChatGPT (1) (2) (3) release in winter 2022, many companies have been investing on ML teams to follow Generative AI hype and to deliver new Generative AI features in their products or internal tools. Despite it is easy for anyone to discuss with a Chat Bot and imagine the potential of the underpinning generative AI models, there is a journey to make generative AI models (1) relevant for a non-native AI company and (2) "production ready".
In this talk, we will share Amadeus status about industrialization of Generative AI models. The talk will sketch different use cases and corresponding end-to-end technical solutions for production. Indeed, despite Large Language Models (LLMs) are already unavoidable "AI component", we will describe how this AI momentum must come back to reality especially regarding the two following points: engineering and consistency.

First, we will discuss engineering concerns like security, legal, cost and business opportunities, scalability and service SLAs, LLMs openness, Foundational models and related model factorization. Second, we will share our vision about the necessary consistency that must stick to any generative AI model. We will discuss possible technical solutions to pave the way to the corresponding neuro-symbolic grail (4) (5) (2).

References

(1) OpenAI, *GPT-4 Technical Report,* 2023.

(2) A. Liu, X. Hu, L. Wen and P. S. Yu, *A comprehensive evaluation of ChatGPT's zero-shot Text-to-SQL capability,* 2023.

(3) J. Yang, H. Jin, R. Tang, X. Han, Q. Feng, H. Jiang, B. Yin and X. Hu, *Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond,* 2023.

(4) L. Wong, G. Grand, A. K. Lew, N. D. Goodman, V. K. Mansinghka, J. Andreas and J. B. Tenenbaum, *From Word Models to World Models: Translating from Natural Language to the Probabilistic Language of Thought,* 2023.

(5) E. Karpas, O. Abend, Y. Belinkov, B. Lenz, O. Lieber, N. Ratner, Y. Shoham, H. Bata, Y. Levine, K. Leyton-Brown, D. Muhlgay, N. Rozen, E. Schwartz, G. Shachaf, S. Shalev-Shwartz, A. Shashua and M. Tenenholtz, *MRKL Systems: A modular, neuro-symbolic architecture that combines large language models, external knowledge sources and discrete reasoning,* 2022.

[*]Speaker
[†]Corresponding author: norbert.leon@amadeus.com
[‡]Corresponding author: francois.hantry@amadeus.com
[§]Corresponding author: gerardo.ayalasolano@amadeus.com

(6) langchain, *langchain.*

Keywords: LLMs, neuro, symbolic, foundational models, Generative AI industrial architecture, security, FinOps

SemiPy: a simple Semi-Supervised Learning Python library

Lucas Boiteau * [1], Pierre-Alexandre Mattei [2], Hugo Schmutz [3], Aude Sportisse [2], Quentin Oliveau [4], Alexandre Gensse [4]

[1] Inria – L'Institut National de Recherche en Informatique et e n Automatique (INRIA) – France
[2] Université Côte d'Azur, INRIA, Maasai, LJAD, CNRS – Université Cote d'Azur, Institut National de Recherche en Informatique et en Automatique, Centre Sophia-Antipolis Méditerranée – France
[3] Université Côte d'Azur, TIRO-MATOS, UMR CEA E4320 – Université Côte d'Azur, UMR E4320 – France
[4] Naval Group – Naval Group – France

Semi-Supervised Learning (SSL) is a machine learning approach that aims to train models using a combination of labelled and unlabelled data, with the goal of improving the model's performance. SSL has emerged as a powerful machine learning strategy, bridging the gap between the efficiency of unsupervised learning and the performance of supervised learning. The main idea is that the unlabelled data contains additional information that can help improve the model's generalization and performance. By using the unlabelled data, the model can learn from the underlying structure of the data distribution, even when explicit labels are unavailable. By leveraging large amounts of unlabelled data together with limited labelled data, SSL algorithms enable the training of robust and accurate models in scenarios where obtaining labelled data is expensive or time-consuming.
In this talk, we introduce SemiPy, a versatile PyTorch-based Python library designed to facilitate and accelerate the development of SSL models. SemiPy combines the flexibility of Python with a comprehensive set of SSL algorithms and techniques, empowering researchers and users to tackle realworld problems with confidence. With SemiPy, users can integrate semi-supervised learning into their existing workflows, unlocking new opportunities for improved model's performance and efficiency.

The talk will begin with an overview of SSL and its significance in the machine learning domain, highlighting the challenges and opportunities it presents. We will discuss the motivation behind the development of SemiPy and its key features, such as the implementation of a custom sampler that allows to load both labelled and unlabelled data using only one dataloader with respect to customizable parameters (for example ratio of labelled/unlabelled items in each batch), or the inclusion of some useful datasets for benchmarking like MedMNIST (1) or CIFAR (2). MedMNIST is a collection of MNIST-like datasets (12 datasets for 2D and 6 for 3D) specifically designed for medical image analysis and classification tasks. It aims to provide a similar standardized evaluation platform for medical imaging tasks. It is therefore more realistic and more difficult to handle than CIFAR.

An overview of the different SSL methods included in the library will be presented, such as Pseudolabel or FixMatch (3). PseudoLabel is a method where unlabelled data is assigned pseudo-labels based on model's predictions. A loss on unlabelled data can then be computed and incorporated in the total loss, allowing training with unlabelled data. FixMatch is an SSL algorithm that uses pseudo-labelling method and data augmentation to both labelled and

*Speaker

unlabelled samples to enforce consistency between predictions made on augmented versions of unlabelled samples.

To provide a hands-on experience, we will showcase several practical use cases where SemiPy has been successfully applied, such as images classification not only with famous benchmarks datasets like CIFAR but also with less used datasets like MedMNIST. Through these examples, attendees will gain insights into the practicality and effectiveness of SemiPy in real-world scenarios.

Finally, we will discuss future directions and ongoing developments in SemiPy, including integration with the latest advancements in SSL research such as simple debiasing for safe SSL (4). We encourage researchers and developers to actively participate in the SemiPy community to drive further innovation and empower the wider adoption of SSL techniques.

By the end of this talk, attendees will be equipped with a comprehensive understanding of SemiPy's capabilities and will be inspired to leverage its potential in their own SSL projects.

(1) J. Yang et al. MedMNIST v2 - A large-scale lightweight benchmark for 2D and 3D biomedical image classification. In Sci Data 10, article number 41 (2023)

(2) A. Krizhevsky, G. Hinton. Learning multiple layers of features from tiny images. (2009)

(3) K. Sohn et al. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. In Advances in Neural Information Processing Systems, pages 596-608 (2020)
(4) H. Schmutz, O. Humbert, and P.-A. Mattei. Don't fear the unlabelled: safe deep semi-supervised learning via simple debiasing. In 11th International Conference of Learning Representations. arXiv preprint arXiv:2203.07512v3 (stat.ML), 2023

Keywords: semi, supervised, library, python, pytorch

SMEs and AI: Triumphs, Tools, and Tomorrow's Trends

Dr. Anna Lahtinen *† [1]

[1] Haaga-Helia University of Applied Sciences [Helsinki, Finland] – Ratapihantie 13, 00700 Helsinki, Finland, Finland

The proposed talk titled "SMEs and AI: Triumphs, Tools, and Tomorrow's Trends" seeks to shed light on how small and medium-sized enterprises (SMEs) can benefit from Artificial Intelligence (AI). Large corporations currently reap the lion's share of AI's benefits; nonetheless, SMEs – with their sheer numbers and significant economic contribution – are ripe for AI integration (Lahtinen & Humala, 2023).
Research indicates that the proper integration of AI solutions accelerates the product development cycle, curtails expenses, and enhances product quality (Latvakoski, 2021). Regarding the diverse spheres of AI application, Finnish SMEs perceive product and service development and manufacturing as areas with immense potential (Ruohonen, 2021). Key challenges for SMEs to adopt AI include time, money, and skills.

SMEs primed for AI utilization include those that have digitalized their operations and those planning to use AI to refine internal processes or innovate new products and services (Lahtinen & Humala, 2023). However, many SMEs, particularly those focused on export, lack a deep understanding of AI and its potential. The readiness of SMEs to transition to AI varies, making Finland an ideal country to pilot AI solutions due to its agile business environment fostering cooperation among businesses, research institutes, municipal and city organizations, and society (Nikina-Ruohonen, SanMiguel, & Kauttonen, 2021).

This talk will provide insights from over two years of AI piloting by over a hundred of Finnish SMEs within the AI-TIE project framework. AI success stories in SMEs, tools for AI deployment and future trends will be based on the results and outcomes of AI-TIE – AI Technology Innovation Ecosystems for Competitiveness of SMEs (2021-2023), with the mission to support SMEs in business development and growth by promoting the use artificial intelligence solutions through AI accelerators, training and an AI ecosystem. AI-TIE was financed by EAKR (REACT-EU) and Regional Council of Uusimaa. Additionally, sister project AI-TIE Southern Finland is supported by Regional Council of Kymenlaakso.

Examples of practical AI use cases and applications in Finnish SMEs will be introduced. The talk will also present various tools for AI deployment in SMEs, many of which have emerged from the AI-TIE project. These tools that are available free of charge include the following:

> Online portal AI stories of Finnish SMEs (www.aistories.fi), AI in Finland interview series (www.aistories.fi/suomi), where Finnish public figures and social influencers describe their experiences of artificial intelligence and the potential it brings to businesses, "Empowering SMEs with Artificial Intelligence" guide (www.haaga-helia.fi/en/ai-tie) that provides examples of AI solutions and company experiences, ,knowledge and understanding of AI,

*Speaker
†Corresponding author: anna.lahtinen@haaga-helia.fi

tools to support companies' adoption of AI, Online course "AI in Business" (www.ai-in-business.fi) that helps to brainstorm, analyse, plan and prioritise AI use cases.

As for future trends, the business interest and willingness to engage in AI experimentation have grown significantly. By 2030, AI-based solutions will support continuous innovation by SMEs on a large scale. Moreover, many SMEs will have succeeded in creating disruptive, niche products and services, driving new business and possibly revolutionizing their sectors in ways similar to the advent of the internet (Lahtinen & Humala, 2023).

References:

Lahtinen, A., & Humala, I. (Eds.). (2023). Empowering SMEs with Artificial Intelligence. Haaga-Helia University of Applied Sciences. https://www.haaga-helia.fi/en/ai-tie

Latvakoski, M. (2021). Tekoälyn hyödyntäminen tuotekehityksessä. (Deployment of AI in product development.) LUT University.

Nikina-Ruohonen, A., SanMiguel, E., Kauttonen, J. (2021). Cleantech AI: suomalaisten PK-yritysten matka kohti tekoälykypsyyttä. Uusiouutiset-lehden blogi. https://www.uusiouutiset.fi/cleantech-ai-suomalaisten-pk-yritysten-matka-kohti-tekoalykypsyytta/
Ruohonen, A. (2021). Puhtaan teollisuuden kiihdyttämön yritysten alkukartoitus. Esitys AI-TIE tekoälykiihdyttämön kick-off -tilaisuudessa 4.11.2021. (An initial survey of the companies in a cleantech AI accelerator. A presentation at the kick-off event of the AI-TIE artificial intelligence accelerator 4 November 2021.)

Keywords: Small and Medium, sized Enterprises (SMEs), Artificial Intelligence (AI), AI adoption, AI stories, AI tools, AI Trends

Sustainable AI Integration: A New Era of Empowered Knowledge Workers and Collaborative Innovation with HEIs

Dr. Anna Lahtinen *† [1], Dr. Iris Humala *

[1] Haaga-Helia University of Applied Sciences [Helsinki, Finland] – Ratapihantie 13, 00700 Helsinki, Finland, Finland

Artificial Intelligence (AI), as a transformative, general-purpose technology, is not just a technological marvel, but a catalyst for sustainable organizational development. It automates tasks that were conventionally performed by human intellect, enabling a wide range of applications - from customer segmentation and automated accounting to warehouse optimization, all guided by sustainability principles.
Transitioning to sustainable AI is not just a technological shift; it's a comprehensive transformation requiring the knowledge workers' multifaceted expertise. These workers act as pivotal change navigators, whose potential needs to be effectively harnessed to integrate sustainable AI into work processes. Knowledge workers form a notable strategic asset for organizations, not in the least in the context of technological transformation. Their naturally performed organizational roles such as experts, team members, coordinators, assistant, IT, communications, HR and other functions place them at the central conjunctions of knowledge flows. (Kärnä et al. 2021) Furthermore, involvement of personnel and the positive attitude of in all digitalization processes are crucial (Barann et al. 2019).

The integrative role of Higher Education Institutions (HEIs) is central in bringing AI-related research and the industry's AI based development work closer together to enable the next growth stage (e.g., Fortune & Shelton, 2012; Nikina-Ruohonen 2021). One key aspect of sustainable AI is the emerging collaboration between the industry and HEIs. Sustainable AI fosters new pathways for applied research, innovation, and development, broadening the roles undertaken by HEIs - from research leadership to facilitation and mentoring. This partnership creates a cooperative ecosystem, where sustainable AI is not merely adopted but co-developed.

Our presentation draws from a longitudinal research process (2021-2022) within a large Finnish municipal organization. Through a quantitative survey, qualitative interviews, and action research workshops, we offer a rich perspective on preparing knowledge workers for the co-development of sustainable technology.

Our proposal presents an implementation framework for sustainable AI (Appendix 1), supported by the unique alliance of HEIs and industry in research, development, and innovation. Beyond theoretical perspectives, we provide actionable insights into practical aspects of integrating sustainable AI within an organizational context. As the landscape of AI continues to evolve towards sustainability, our talk emphasizes the critical role of knowledge workers and HEIs in ensuring a smooth, effective, and sustainable transition to AI-powered organizations.

*Speaker
†Corresponding author: anna.lahtinen@haaga-helia.fi

References:

Barann, B., Hermann, A., Cordes, A. K., Chasin, F., & Becker, J. (2019, January). Supporting digital transformation in small and medium-sized enterprises: a procedure model involving publicly funded support units. In Proceedings of the 52nd Hawaii International Conference on System Sciences.

Fortune, A. & Shelton, L. (2012). R&D effort, effectiveness, and firm performance in the pharmaceutical sector. *Journal of Business and Management*, 18(1), 97–115.

Kärnä, E., Nikina-Ruohonen, A., & Humala, I. (2021). Entrepreneurial spirit of knowledge workers as a key asset in strategic change. *eSignals Research*, HHBIC 2020, 17– 18.11.2020, Online.
Nikina-Ruohonen, A. (2021). Leading and Managing Areas of Innovation: The MultiStakeholder Perspective. *eSignals Research*.

Keywords: Sustainable Artificial Intelligence, Knowledge Workers, Organizational Sustainability, Higher Education Institutions (HEIs), Technological Transformation, AI Adoption

Leveraging AI in Finland's Health and MedTech Sector: A Focus on SMEs

Dr. Anna Lahtinen *† [1]

[1] Haaga-Helia University of Applied Sciences [Helsinki, Finland] – Ratapihantie 13, 00700 Helsinki, Finland, Finland

In this proposed talk, we will delve into the integration of Artificial Intelligence (AI) into Finland's health and medtech sector, with a particular emphasis on small- and medium-sized enterprises (SMEs). As SMEs constitute 99% of all businesses in Europe, the potential for AI-driven growth is enormous (Ruohonen & Kallasvaara, 2022). This becomes even more compelling considering that health technology has contributed nearly €14 billion to Finland's foreign trade over the past 20 years (Healthtech Finland 2022), making it a sector ripe for AI advancements. The health and medtech sector in Finland has distinguished itself in AI research and development, as evidenced by its significant contribution to Europe's patent filings – responsible for 1% of all submissions. This has placed Finland in an impressive 8th position among the EU27 for AI-related healthcare patents and 11th for academic publications (European Commission, 2021).

The talk will underscore Finland's core strengths in AI, such as its expertise, high-quality data, and effective collaboration (ETLA, 2019). The country's promising growth in AI solution developers, coupled with the increasing demand for such professionals among SMEs, will be examined.

The Finnish strategy emphasizes unique AI applications rather than economies of scale. The COVID-19 pandemic has further catalyzed the growth of AI-focused health and medtech enterprises, highlighting the potential for AI to revolutionize diagnosis and treatment.

In addition to providing key insights of AI in Finland's health and medtech, a significant focus of the talk will be on the experiences within AI-TIE accelerator, Finland's pioneering Health and Med Tech AI accelerator launched in 2022. It embarked on an innovative journey with 16 Finnish SMEs to identify and leverage AI-related business opportunities in the health and medtech sector. The accelerator program showcases Finland's commitment to deepening industry-specific AI expertise and supporting the companies operating within this vibrant industry (Ruohonen & Kallasvaara, 2022).

During the talk, firsthand stories from the participating Finnish health and medtech SMEs will be shared, providing an inside look at their experiences with AI integration and development. Some of the AI use cases will be introduced to the audience as examples of the practical solutions developed by Finnish SMEs in the sector. Another key aspect of the talk will be the presentation of various tools for AI deployment in SMEs that have been developed through collaboration between higher education institutions and industry for Research, Development, and Innovation (RDI). The need for a multidisciplinary approach to AI in higher education, research, development, and innovation initiatives will be emphasized, underlining the importance of a holistic approach in shaping the future of AI in Finland's health and medtech sector.

*Speaker
†Corresponding author: anna.lahtinen@haaga-helia.fi

References

ETLA. (2019). Digital Barometer. (link)

European Commission. (2021). Study on eHealth, Interoperability of Health Data and Artificial Intelligence for Health and Care in the European Union. Artificial Intelligence for health and care in the EU. PwC. (link)

Healthtech Finland. (2022). The export of healthtech continues to grow - Corona has driven the export of laboratory diagnostics. (link)
Ruohonen, A., & Kallasvaara, H. (2022). AI and medtech ecosystem in Finland – opportunities for SMEs. MedTech News. (link)

Keywords: Artificial Intelligence (AI), Health and MedTech, Small and Medium, sized Enterprises (SMEs), Finland, AI Ecosystem, Industry, Academia Collaboration

Prescribing with Caution: Assessing the Impact of Potential Drug-Drug Interactions in the Pediatric Intensive Care Unit

Pradeep Singh * [1], Akshaya Devadiga *

[1], Ridam Pal [1], Pranam R Betrabet [2], Jhuma Sankar [3], Tavpritesh Sethi [1], Rakesh Lodha† [3]

[1] Indraprastha Institute of Information Technology [New Delhi] – Okhla Industrial Estate,Phase III(Near Govind Puri Metro Station)New Delhi, 110020, India
[2] Manipal Academy of Higher Education – MANIPAL, KARNATAKA, India
[3] All India Institute of Medical Sciences [New Delhi] – Ansari Nagar, New Delhi - 110029, India

Background: The intricate nature of pediatric patients' physiology and the frequent administration of multiple medications in pediatric intensive care units (PICU) expose hospitalized children to potential drug-drug interactions (pDDIs)1,2. Therefore, the objective of this study was to examine and characterize the clinically relevant potential DDIs (pDDIs) in pediatric patients admitted in pediatric intensive care units.
Method: In this retrospective study, we analyzed the prescriptions for children admitted to the PICU over five years (2016–2021) for pDDI. The identification of pDDIs in this study was accomplished by utilizing the DDInter database and determining the clinical relevance using the database's severity ratings (Major, Moderate, and Minor). Using univariate and multivariate linear regression, the association between the severity of pDDI and the variables age and number of prescribed medications was investigated. Additionally, the predictors for PICU length of stay were investigated3.

Findings: The examination of 8010 prescriptions for the 899 patients hospitalized in the PICU revealed 3884 significant pDDIs. The most frequent interacting drug pairs were Midazolam + Fluconazole, Fluconazole + Fentanyl, Fentanyl + Dexamethasone, and Vancomycin + Piperacillin. Among the 49·6% (446) of pediatric patients admitted to the PICU, who had at least one significant (major, moderate, or minor) interaction, 38·8% (173) had a major pDDI. In this investigation, a higher number of prescribed drugs was significantly linked with a higher incidence of pDDIs (odds ratio (OR): 12·79; 95% confidence interval (CI): 8·26–20·6; $p< 0·001$). The presence of major and moderate pDDIs and patient age $\geq$ one year ($p< 0·001$), as well as the prescription of five or more medications ($p< 0·001$), were also strongly associated with the length of ICU stay.

Interpretation: The study suggests that pDDIs are frequent in PICU. The majority of these interactions were moderate, followed by major severity. There is a statistically significant association between the number of drugs provided and the prevalence of pDDIs. The findings of this study could aid in the planning and execution of future studies as well as the monitoring and prevention of pDDIs in children receiving intensive care. This study introduces "Contraindicator,"

*Speaker
†Corresponding author: rlodha1661@gmail.com

a graphical user interface that presents empirical findings and promotes awareness regarding pDDIs.

Bibliography:

1 Hassanzad M, Arenas–Lopez S, Baniasadi S. Potential Drug-Drug Interactions Among Critically Ill Pediatric Patients in a Tertiary Pulmonary Center. J Clin Pharmacol, 2018; 58(2):221-227. DOI: 10.1002/jcph.996

2 Mousavi S, Ghanbari G. Potential drug–drug interactions among hospitalized patients in a developing country. Casp J Intern. Med. 8, 2017; 282–288. doi:10.22088/cjim.8.4.282

3 Lima EdC, Camarinha BD, Ferreira Bezerra NC, Panisset AG, Belmino de Souza R, Silva MT and Lopes LC. Severe Potential Drug–Drug Interactions and the Increased Length of Stay of Children in Intensive Care Unit. Front. Pharmacol. 2020; 11:555407. DOI: 10.3389/fphar.2020.555407

Keywords: pediatric patients, potential drug–drug Interaction, intensive care units, named entity recognition, medical prescriptions, length of stay

Characterizing the Emotion Carriers of COVID-19 Misinformation and Their Impact on Vaccination Outcomes in India and the United States Characterizing the Emotion Carriers of COVID-19 Misinformation and Their Impact on Vaccination Outcomes in India and the United States

Ridam Pal * [1], Sanjana S [2], Deepak Mahto [2], Kriti Agrawal [2], Gopal Mengi [2], Sargun Nagpal [3], Akshaya Devadiga [2], Tavpritesh Sethi†

[1] Indraprastha Institute of Information Technology [New Delhi] – Okhla Industrial Estate,Phase III(Near Govind Puri Metro Station)New Delhi, 110020, India
[2] IIITD – India
[3] NYU – United States

Background: The COVID-19 Infodemic had an unprecedented impact on health behaviors and outcomes at a global scale. While many studies have focused on a qualitative and quantitative understanding of misinformation, including sentiment analysis, there is a gap in understanding the emotion-carriers of misinformation and their differences across geographies. In this study, we characterized emotion carriers and their impact on vaccination rates in India and the United States. Our large language model was trained on 2.3 million tweets carrying COVID-19 misinformation and validated prospectively on a manually annotated set of 2000 misinformation tweets across the globe.
Methods: A manually labelled dataset was created and collated with three publicly available datasets (CoAID, AntiVax, CMU) to train deep learning models for misinformation classification. Misinformation labelled tweets were further analyzed for behavioral aspects by leveraging Plutchik Transformers to determine the emotion for each tweet. Time series analysis was conducted to study the impact of misinformation on spatial and temporal characteristics. Further, categorical classification was performed using transformer models to assign categories (Authority, Symptoms, Cure, Vaccination and Spread) for the misinformation tweets.

Findings: Word2Vec+BiLSTM was the best model for misinformation classification, with an F1-score of 0.92. The US had the highest proportion of misinformation tweets (58.02%), followed by the UK (10.38%) and India (7.33%). Disgust, anticipation, and anger were associated with an increased prevalence of misinformation tweets. Disgust was the predominant emotion associated with misinformation tweets in the US, while anticipation was the predominant emotion in India. For India, the misinformation rate exhibited a lead relationship with vaccination, while

*Speaker
†Corresponding author: tavpriteshsethi@iiitd.ac.in

in the US it lagged behind vaccination.

Interpretation: Our study deciphered that emotions acted as differential carriers of misinformation across geography and time. These carriers can be monitored to develop strategic interventions for countering misinformation, leading to improved public health.
References

1. Cinelli, M. et al. The COVID-19 social media infodemic. Sci. Rep. 10, 16598 (2020).
2. Cheng, C. & Lau, Y.-C. Social Media Addiction during COVID-19-Mandated Physical Distancing: Relatedness Needs as Motives. Int. J. Environ. Res. Public. Health 19, 4621 (2022).
3. Karhu, M., Suoheimo, M. & Häkkilä, J. People's Perspectives on Social Media Use during COVID-19 Pandemic. in 20th International Conference on Mobile and Ubiquitous Multimedia 123–130 (ACM, 2021). doi:10.1145/3490632.3490666.

A Prospectively Validated Generalizable Model for Outcome Prognostication Using Shock Index in Intensive Care Units

Aditya Nagori [1], Pradeep Singh * [1], Sameena Firdos [1], Vanshika Vats [1], Arushi Gupta [1], Harsh Bandhey [1], Anushtha Kalia [1], Arjun Sharma [1], Prakriti Ailavadi [2], Raghav Awasthi [1], Wrik Bhadra [1], Ayushmaan Kaul [1], Akshaya Devadiga [1], Rakesh Lodha [3], Tavpritesh Sethi†

[1] Indraprastha Institute of Information Technology [New Delhi] – Okhla Industrial Estate,Phase III(Near Govind Puri Metro Station)New Delhi, 110020, India
[2] Netaji Subhas University of Technology – Delhi, India
[3] All India Institute of Medical Sciences [New Delhi] – Ansari Nagar, New Delhi - 110029, India

Background- Shock Index (SI) is a widely accepted indicator for prognosticating outcomes in critical care and emergency settings(1). This study aimed at building a generalizable and prospectively validated model for SI(2), followed by its clinical profiling in disease subsets(3). Model development was carried out using a multi-centric eICU database, with 208 ICUs across the United States, for early prediction of abnormal SI with up to 8 hours of lead times.
Methods- A total of 16,246 ICU stays of adult patients eligible for the development cohort from the eICU database were taken for modeling. Deep learning models were compared against a parsimonious model with features selected from 3970 statistical and complexity-based characteristics of time series, followed by interpretability analysis. External validation of the selected model was performed on (i) publicly available adult MIMIC-III data and (ii) a pediatric age group in an ICU in New Delhi, India (SAFE-ICU). Prospectively validated over a period of six months in the Indian pediatric ICU and profiled for performance over 15 disease subsets.

Results- We found mortality rate was associated with the proportion of the length of stay with abnormal shock index (Pearson's correlation r = 0.89, p-value = 4.8×10-4), underscoring the importance of early detection of abnormal shock index. In the e-ICU dataset, our model SIgnose identified 92% of all the events of shock index abnormality (median > 0.7 over 30 minutes) with a lead-time of 8 hours and achieved an AUROC of 86% (SD= 1.2), AUPRC of 93% (SD =1.1). External validation on the MIMIC-III cohort achieved an AUROC of 87% (SD =1.6) and an AUPRC of 92 % (SD=1.5). Finally, prospective validation of SIgnose in the SAFE-ICU(n=88; Age =19.2 (26.7) months) resulted in an AUROC of 88.5% and AUPRC of 91%, demonstrating generalizability across geographies and age groups. In the pediatric cohort, our model had the highest sensitivity and positive predictive value (PPV) for patients with Sepsis (93% and 100%, respectively), while in the MIMIC cohort, it had comparable sensitivity (87%) and PPV (92%) in pneumonia, atrial fibrillation and acute kidney injury, signifying broad applicability of our model across critical care illnesses.

Conclusion & Added value of this study:- The study provides a generalizable, prospectively verified, clinically supported and openly available prediction model for prognosticating shock in-

*Speaker
†Corresponding author: tavpriteshsethi@iiitd.ac.in

dex up to 8 hours in advance from vitals time series alone. We believe this to be the first study assessing the transfer of a predictive trained on adult ICU data from developed settings in the United States to a pediatric ICU in Indian settings, thus providing a framework for building critical care models. We also show the superiority of time-series features over black-box deep learning, thus improving clinical interpretability and trust. Finally, the open-source release of models as a pre-configured Docker container can be used by other clinical settings to fine-tune the models for advancing research and nuanced clinical utility of shock index-based outcome prognostication.- https://github.com/tavlab-iiitd/ShoQPred/tree/master.

References

1. Koch E, Lovett S, Nghiem T, Riggs RA, Rech MA. Shock index in the emergency department: utility and limitations. Open Access Emerg Med OAEM. 2019;11:179–99.

2. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, et al. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016 Feb 23;315(8):762–74.

3. Mao Q, Jay M, Hoffman JL, Calvert J, Barton C, Shimabukuro D, et al. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. BMJ Open. 2018;8:17833.

Keywords: Shock Index, Intensive Care, Predictive Modeling, Generalization, Clinical, Profiling, Prospective Validation

Bias Amplification in Intersectional Subpopulations for Clinical Phenotyping by Large Language Models

Ridam Pal * [1], Hardik Garg *

[1], Shaswat Patel [1], Tavpritesh Sethi†

[1] Indraprastha Institute of Information Technology [New Delhi] – Okhla Industrial Estate,Phase III(Near Govind Puri Metro Station)New Delhi, 110020, India

Abstract

Large Language Models (LLMs) have demonstrated remarkable performance across diverse clinical tasks. However, there is growing concern that LLMs may amplify human bias and reduce performance quality for vulnerable subpopulations. Therefore, it is critical to investigate algorithmic underdiagnosis in clinical notes, which represent a key source of information for disease diagnosis and treatment. This study examines prevalence of bias in two datasets - smoking and obesity - for clinical phenotyping. Our results demonstrate that state-of-the-art language models selectively and consistently underdiagnosed vulnerable intersectional subpopulations such as young-aged-males for smoking and middle-aged-females for obesity. Deployment of LLMs with such biases risks skewing clinicians' decision-making which may lead to inequitable access to healthcare. These findings emphasize the need for careful evaluation of LLMs in clinical practice and highlight the potential ethical implications of deploying such systems in disease diagnosis and prognosis.

References

1. Seyyed-Kalantari L, Zhang H, McDermott M, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations — Nature Medicine (Internet). (cited 2023 Feb 25). Available from: https://www.nature.com/articles/s41591-021-01595-0

2. Zhang H, Lu AX, Abdalla M, McDermott M, Ghassemi M. Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings (Internet). arXiv; 2020 (cited 2023 Mar 12). Available from: http://arxiv.org/abs/2003.11515

3. Mulyar A, Schumacher E, Rouhizadeh M, Dredze M. Phenotyping of Clinical Notes with Improved Document Classification Models Using Contextualized Neural Language Models (Internet). arXiv; 2020 (cited 2023 Mar 12). Available from: http://arxiv.org/abs/1910.13664

Keywords: bias, large language model, natural language processing, fairness, phenotyping, electronic health records

*Speaker

†Corresponding author: tavpriteshsethi@iiitd.ac.in

A cross-silo Federated Learning based framework to make the industrial AI product lifecycle more sustainable

Hongliu Cao * [1]

[1] Amadeus – Amadeus – Nice, France

With generative models catching the attention of general users across the world in the past years, many industries have accelerated their adoption of Artificial Intelligence (AI) solutions to their products. AI technology can be used to improve the efficiency and productivity of a lot of people, but it also comes at a cost to the planet. AI product lifecycle may consume significant energy and resources, from data storage (data collection, processing, transmission, management, etc.), to model development (data engineering, model training, testing, evaluation, re-training, tuning, etc.), to model deployment and maintenance. Today, it is estimated that training a single deep learning model can generate as much CO2 as the total lifetime of five cars (1). Hence, there is a growing need to ensure that the AI product lifecycle is sustainable and environmentally responsible.
Federated Learning is a distributed learning framework known for its privacy-preserving ability, where a bunch of clients collaboratively train a global model under the coordination of a central server. Instead of data transmission, clients train models using their private local data and upload model updates. When participated clients are small, distributed entities with relatively small amount of local data, it is called cross-device Federated Learning. In cross-silo Federated Learning, the clients are companies or organizations: small number of participants each with large amount of local data (2). In the state-of-the-art studies, most of them focus on Federated Learning combating Non-IID data and trustworthy Federated Learning. A few existing solutions studied the energy efficient training strategies for federated learning (1,3,4). However, they mainly focus on reducing the energy cost of model training only for cross-device Federated Learning and lack of sustainability comparison with the centralized solution.

Cross-silo Federated Learning is now starting to be deployed at a global scale by more companies as businesses need to comply with new legal requirements and policies committed to privacy protection. However, the worrying situation is that we currently have little to no understanding of its impact on carbon emissions. In this work, we focus on how companies can reduce their carbon footprint by leveraging cross-silo Federated Learning, where no existing solution is found in the literature. The main contributions of this work include:

- This is the first work in the literature: studying the sustainability aspect of cross-silo Federated Learning, across the AI product lifecycle instead of focusing only on the model training in the literature, with the comparison to the centralized method, with the objective to help companies to be more cost effective and more environmentally friendly.
- Analytical Carbon Footprint analysis for cross-silo Federated Learning: We provide a more holistic quantitative cost and CO2 emission estimation method for cross-silo Federated

*Speaker

Learning, with both operational and embodied emissions, including data storage, data transmission, model training, communication, etc.

- Real-world experiments: cross-silo Federated Learning cost and carbon footprint analysis with multiple real-world industrial scenarios on real hardware, network and security settings of Federated Learning on Azure Cloud are reported and analyzed (the first work in the literature).
- Direction towards Carbon-friendly cross-silo Federated Learning: Based on the experiment analysis, we propose a cross-silo Federated Learning based framework for companies to improve their sustainability level. We also provide future research directions in developing carbon-friendly federated learning.

(1) Güler, Başak, and Aylin Yener. "A framework for sustainable federated learning." 2021 19th International Symposium on Modeling and Optimization in Mobile, Ad hoc, and Wireless Networks (WiOpt). IEEE, 2021.

(2) Li, Qinbin, et al. "Federated learning on non-iid data silos: An experimental study." 2022 IEEE 38th International Conference on Data Engineering (ICDE). IEEE, 2022.

(3) Zeng, Qunsong, et al. "Energy-efficient radio resource allocation for federated edge learning." *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2020.

(4) Qiu, Xinchi, et al. "A first look into the carbon footprint of federated learning." *Journal of Machine Learning Research,* 24 (2023): 129-1.

Keywords: Sustainable AI, Cross silo, Federated Learning, AI product lifecycle

Standardization in support of AI and Management

Antoine Mouquet *† [1]

[1] ETSI – European Telecommunications Standards Institute [Sophia Antipolis] – 650 route des Lucioles - 06921 Sophia Antipolis cedex, France

Artificial Intelligence (AI) is accelerating the digital transformation and ETSI, one of European Standardization Organizations (ESOs) based in Sophia Antipolis, is at the heart of digital.
AI enhances all things digital
Artificial Intelligence is a complex and powerful tool to identify patterns buried in masses of information, sometimes the meaning and usage of those patterns needs to be confirmed by human beings. Patterns can point to many kinds of improvements to optimize capacity of communication networks, to diagnose cancers from protein fragments in blood samples, to extract semantic meaning from spoken sounds, to detect and deflect virus attacks on computer networks, to identify a safe route for driving along the street, and a thousand other applications. ETSI focuses on using AI to improve applications using information and communications technologies (ICT), and in identifying technical means to constrain various lapses or misuses of AI.

An array of AI activities in the ETSI's standardization work
AI is a horizontal activity which encompasses almost every aspect of future ICT systems; AI is not the end goal but a means to achieve the goal. It is thus impossible to define "the single AI solution" – rather, ETSI Technical Committees (TC) and Industry Specification Groups (ISG) consider the technology for their specific purposes in a less harmonized way. ETSI has chosen to implement a distributed approach to standardization activities related to AI – specialized communities meet in technically focused groups. Examples include:

- ISG SAI (Securing Artificial Intelligence) considers errors in, or misuses of, AI algorithms and how to minimize those.
- TC Cyber works with ISG SAI to respond to Standardization Requests (SRs) from the European Commission on topics related to AI Security.
- TC ITS (Intelligent Transport Systems) is of course updating its guidelines and specifications for interworking of car-to-X communications to take account of the needs of AI-enabled autonomous vehicles and also city traffic optimization.
- TC eHealth (Electronic Heath) is focused on ensuring interworking of health and personal health solutions which use AI, and – most importantly – ensuring that the solutions can be designed to comply with European Union requirements for public and personal health, safety and mental health (including privacy).

*Speaker
†Corresponding author: Antoine.Mouquet@etsi.org

These are four of the 14 groups currently working on AI related technologies within ETSI. The first initiative dates back to 2016 with the publication of a White Paper describing GANA (the Generic Autonomic Networking Architecture).
ETSI specifications as well as workshops, webinars, guides and whitepapers can offer practical support to the introduction of AI into many areas.
During this talk, the ETSI representative will share more about the new European regulatory framework on Artificial Intelligence and how standardization can help shape the future of AI in Europe and globally. It is understood that AI is a technology with immense opportunities giving the potential to offer a wide array of economic and societal benefits to European citizens and industries and beyond. On the other hand, care must be taken to avoid potential new risks or negative consequences for individuals or the society. Those may indeed arise if AI is used without any rules or boundaries. Use cases of ongoing standardization activities detailed in the white paper published early 2023: Artificial Intelligence and future directions for ETSI , will be shared and support the talk.

Keywords: Standardization, EU IA Act, ICT (Information and Communications technologies), Cybersecurity

Enabling Lung Cancer Screening with iBiopsy® AI-based Software as Medical Device

Benoit Huet * [1], Pierre Baudot , Van Khoa Le , Danny Francis , Ezequiel Geremia , Gwendoline De Bie

[1] MEDIAN Technologies – MEDIAN Technologies – Les Deux Arcs B1800 route des Crêtes06560 Valbonne, France

Median's iBiopsy Lung Cancer Screening (LCS) Software as Medical Device (SaMD) offers a unique end-to-end Computer Aided Detection/Diagnosis approach (CADe/CADx (5)) based on AI/ML technologies. With this upcoming SaMD, Median Technologies aims at drastically improving the ability to detect and characterize lung cancers at their earliest stages within low-dose computed tomography (CT), enabling a better patient care while avoiding unnecessary medical tests and procedures, and reducing healthcare costs.
Lung cancer is the leading cancer killer worldwide and the second most commonly diagnosed cancer, accounting for 18% of all cancer deaths (1). The I-ELCAP study (2) showed a 80% survival rate at 20 years when Lung cancer is diagnosed at early stage, yet when diagnosed at the earliest stage (stage 1) the survival rate jumps to 92% while dropping to only 5% for stage 4. Unfortunately, lung cancer symptoms appear generally at later stages. This further strengthen the importance of a screening program as exposed and demonstrated by NLST (3) and NELSON (4) among others...

With this breakthrough innovation, Median's iBiopsy LCS SaMD brings a unique AI-powered solution for clinicians to fight lung cancer, the deadliest cancer worldwide, by detecting and characterizing malignant lung nodules in low dose CT scans (LDCT). A key enabling solution for screening the vast eligible population, estimated to over 130M worldwide. Latest results, obtained on a very large cohort of about 10K patients show excellent performance of the iBiopsy® LCS CADe/CADx Software as Medical Device (SaMD) with a sensitivity (6) of 96.5% at a specificity (7) of 97.2%. The Area Under the Receiver Operating Curve of the of lung nodule characterization stand at 0.974.

Strong of such performance, and further to the completion of the Q-Submission phase with the FDA (the Food and Drug Administration, USA), Median is on trajectory for execution of the upcoming pivotal studies of this device with an aim at obtaining 510(k) clearance in the course of 2024. *Our next milestone is the execution of the pivotal studies, based on imaging and clinical data we started to collect at the end of 2022 from very prestigious clinical sites and cancer centers in the US and Europe.*

(1) https://gco.iarc.fr/

(2) https://www.mountsinai.org/about/newsroom/2022/lung-cancer-screening-dramatically-increases-long-term-survival-rate

*Speaker

(3) NLST (2011) Reduced lung-cancer mortality with low- dose computed tomographic screening.

(4) NELSON (2020) Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial

(5) A radiological CADe device is *"intended to identify, mark, highlight or otherwise direct attention to portions of an image that may reveal abnormalities during interpretation of images by the clinician.*" A CADx device is "*intended to provide information beyond identifying abnormalities, such as an assessment of disease." Source: FDA*

(6) Sensitivity is the ability to correctly generate positive results for cancer patients.
(7) Specificity is the ability to correctly generate negative results for non-cancer patients

Keywords: AI in Medical Imaging, Early Cancer Detection, Lung Cancer Screening

TRACTIVE: TowaRds A Computational mulTImodal analysis of film discursiVe aEsthetics

Lucile Sassatelli * [1], Hui-Yin Wu [2], Virginie Julliard [3], Sarah Lécossais [4], Damon Mayaffre [5], Florence Sèdes [6]

[1] Université Côte d'Azur, CNRS, I3S – Université Côte d'Azur (UCA), Université Côte d'Azur (UCA) – France

[2] Biologically plausible Integrative mOdels of the Visual system : towards synergIstic Solutions for visually-Impaired people and artificial visiON – Inria Sophia Antipolis - Méditerranée – 2004 route des LuciolesBP 9306902 Sophia Antipolis, France

[3] Groupe de recherches interdisciplinaires sur les processus d'information et de communication – Sorbonne Universite – CELSA - 77, rue de Villiers - 92522 Neuilly- sur-Seine Cedex, France

[4] Université Sorbonne Paris Nord – Sorbonne Paris Nord – 99, avenue Jean-Baptiste Clément, 93430 Villetaneuse, France

[5] Bases, Corpus, Langage (UMR 7320 - UCA / CNRS) – Centre National de la Recherche Scientifique, Université Côte d'Azur – Laboratoire BCL - UMR 7320MSHS Sud-Est Pôle Universitaire Saint Jean d'Angely Bâtiment SJA 3 24 avenue des Diables bleus, 06357 Nice Cedex 4, France

[6] Université Toulouse III - Paul Sabatier – Université de Toulouse – 118 route de Narbonne - 31062 Toulouse, France

Surveys on the quantitative representation of women in visual media are insufficient to grasp the issue of gender inequality. In film studies, two relevant visual discursive regimes have been identified and recently revisited: the male gaze and the female gaze. Male gaze is defined as a set of mise-en-scène techniques to represent a character as an object of desire rather than a subject of action. On the contrary, female gaze techniques are employed to convey to the audience the experience felt by a character. These cinematographic gestures are difficult to pinpoint but correspond to complex, subtle and wide-spread visual discourse patterns, that may convey biased gender representation. How can we characterize them and quantify the extent of their respective usage?

This poster presents the multi-disciplinary work ongoing in the TRACTIVE project, stemming from this strong societal motivation. TRACTIVE is a four-year research project funded by the French National Research Agency (ANR) and involving six academic laboratories in computer science and social science and humanities. The scientific objective of TRACTIVE is to characterize and quantify gender representation and objectification in films and visual media by designing AI-driven multimodal (visual, textual, audio) discourse analyses. To address this objective, TRACTIVE brings together researchers from computer science, media studies, linguistics, and gender studies.

The poster will be organized to reflect the articulation of the multiple complementary approaches taken in TRACTIVE. First, a joint work between computer science and media studies enable the definition of a new concept, multimodal gender objectification, to c**reate two new annotated film datasets enabling data-centric AI approaches**. Second, learning approaches are taken to address the task of **multimodal objectification detection, based on graph-based transformers**. Third, **knowledge-driven approaches** are adopted to describe the data in a structured way, and approach the task of objectification detection with **graph infer-**

*Speaker

ence. Fourth, approaches in computational linguistics, assisted with **explainable** deep neural networks, enable **to pinpoint linguistic patterns** characteristic of objectification.

Keywords: video analysis, multimodality, transformers, dataset creation

Human trajectory forecasting in 3D contexts with simulated visual impairments

Franz Fabini Franco Gallo *† [1], Hui-Yin Wu [1], Lucile Sassatelli *

[2,3]

[1] Centre Inria d'Université Côte d'Azur – L'Institut National de Recherche en Informatique et e n Automatique (INRIA) – France
[2] Université Côte d'Azur, CNRS, I3S – Université Côte d'Azur (UCA), Université Côte d'Azur (UCA) – France
[3] Institut Universitaire de France – Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de l´ducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de l'Education nationale, de l'Enseignement supérieur et de la Recherche – Maison des Universités 103 Boulevard Saint-Michel 75005 Paris, France

Predicting human trajectory is crucial to have efficient systems in various application areas, such as autonomous driving. Self-driving cars must indeed consider pedestrians' behavior in their decisions. Human trajectory forecasting is a difficult problem depending on several factors: human intents, physical elements occupying the space, social distances, interactions, and visual attention. Existing approaches for pedestrian trajectory forecasting in the street generally rely on datasets made of GPS, video and possible LIDAR data captured from the car itself (1). Creating rich human-centered data is therefore key to improve human trajectory forecasting. This should allow to better understand the motion factors at play, specifically catering for people with mobility impairments, who risk being averaged out in current datasets. Virtual environments, in the form of Virtual Reality or Augmented reality, enable the collection of human data in controlled and varied interactive scenarios where users can move naturally.
The objective of this poster is to present ongoing work within the CREATTIVE3D ANR project.

First, rich multimodal human data (translation and rotation motion with gaze direction, heart rate and electro-dermal conductance) is collected to design models to predict motion and attention in complex environments. We present the dataset stemming from the GusT-3D framework (2), a new dataset of 40 users moving in a virtual environment of 10 meters by 5 meters with a road-crossing task carried out with various types of intents and levels of interactions (e.g., carrying a box across the street) and, importantly, different levels of vision impairments, namely normal vision and simulated Age Macular Degeneration.

Second, we aim to investigate how much low vision can impact the generalization error of models trained on motion of normal-vision persons when predicting motion of low-vision persons at inference time. We benchmark two types of models on this task, specifically an adapted version of a structural RNN for attention prediction (3) and the GIMO model (4), made of 3 cross-attentional modules inspired by the Perceiver IO architecture (5). The modalities attend-

*Speaker
†Corresponding author: franz.franco-gallo@inria.fr

ing to each other are the position, the neighborhood of the scene point cloud around the body and around the gaze target. We also propose a new model able to benefit both from unstructured data (time series of trajectories and scene point cloud) and from the structured description available for the 3D dynamic scene, the series of interaction and movements, and the task model.

References

(1) A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets Robotics: The KITTI Dataset," International Journal of Robotics Research (IJRR), 2013.

(2) F. A. S. Robert, H.-Y. Wu, L. Sassatelli, S. Ramanoel, A. Gros, and M. Winckler, "An Integrated Framework for Understanding Multimodal Embodied Experiences in Interactive Virtual Reality," in 2023 ACM International Conference on Interactive Media Experiences (IMX), 2023.

(3) M. F. Romero Rondón, L. Sassatelli, R. Aparicio-Pardo and F. Precioso, "TRACK: A New Method From a Re-Examination of Deep Architectures for Head Motion Prediction in 360◦ Videos," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 9, pp. 5681-5699, 2022.

(4) Y. Zheng et al., "GIMO: Gaze-Informed Human Motion Prediction In Context," in Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, pp. 676–694.
(5) A. Jaegle et al., "Perceiver IO: A General Architecture for Structured Inputs & Outputs," in International Conference on Learning Representations, 2022.

Keywords: time series forecasting, autonomous cars, transformers, virtual environments

Fully Decentralized Multi-Agent Deep Reinforcement Learning Framework For Distributed Packet Routing in the context of Cloud Overlay Networks

Redha Abderrahmane Alliche *† [1,2], Ramon Aparicio Pardo‡ [1,2], Lucile Sassatelli [1,2,3], Tiago Da Silva Barros [1,2,4]

[1] Université Côte dÁzur – Université Côte d'Azur (UCA) – Parc Valrose, 28, avenue Valrose 06108 Nice Cedex 2, France

[2] Signal, Images et Systèmes – Laboratoire d'Informatique, Signaux, et Systèmes de Sophia Antipolis – Laboratoire I3SCS 4012106903 Sophia Antipolis Cedex, France

[3] Institut Universitaire de France – Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de l´ducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de l'Education nationale, de l'Enseignement supérieur et de la Recherche – Maison des Universités 103 Boulevard Saint-Michel 75005 Paris, France

[4] Inria Sophia Antipolis - Méditerranée – Institut National de Recherche en Informatique et en Automatique – 2004 route des Lucioles BP 93 06902 Sophia Antipolis, France

In the last years, Reinforcement Learning (RL) (1) using Deep Neural Networks (DNNs), also called Deep Reinforcement Learning (DRL), has obtained ground-breaking results in solving highly complex tasks (2). In communication networks, DRL has also been widely used in many networking technologies and problems. One of them is the Distributed Packet Routing (DPR). The DPR is the problem of finding optimal paths for packets, where each node decides locally which neighbor to forward a packet to. Multi-Agent Reinforcement Learning (MA-RL) is a suitable approach for addressing DPR. Each node can work as an agent that learns a local policy based on its observations and rewards, resulting in scalable and robust solutions. Moreover, MA-RL can leverage the advances of Deep Reinforcement Learning (DRL) techniques (MA-DRL), which enables the agents to handle complex and high-dimensional state and action space. Several studies applied MA-DRL to DPR in many situations (3). In this work, we extend MA-RL to overlay network. Overlay networks are virtual or logical networks built on top of a physical network (called underlay networks). The problem of routing the traffic between the overlay links becomes challenging, since the underlay routing policies are unknown and can involve different routing protocols. To handle these challenges, we propose a MA-DRL framework, allowing each agent to adapt its policy to meet the dynamically changing demand and overcome the lack of knowledge about the underlay network. The proposed framework handles the two main issues of training fully decentralized MA-DRL agents in partially observable environments, namely the high communication overhead between the neighbors, and the stability in term of convergence. For the first issue, we propose a new information sharing protocol between the agents, named digital-twin, that minimizes the overhead. For the second issue, we propose a

*Speaker

†Corresponding author: alliche@i3s.unice.fr

‡Corresponding author: raparicio@i3s.unice.fr

guided learning based on RCPO (4) that improves the convergence time and the stability of the algorithm. Finally, we test our solution in prisma-v2(5), which is an open-source realistic network simulator for overlay networks. We demonstrate, by extensive experimentations, that our framework can reach solutions close to the optimum of the centralized version of the problem, which has actually access to a complete view of the network with minimal overhead.

References

(1) R. S. Sutton et al., Reinforcement learning: An introduction. MIT press, 2018.

(2) Nolan B et al (2020) The hanabi challenge: a new frontier for ai research. Artif Intell 280:103216

(3) X. You et al., "Toward packet routing with fully distributed multiagent deep reinforcement learning," IEEE Transactions on Systems, Man, and Cybernetics: Systems, pp. 1–14, 2020.

(4) Sobhan Miryoosefi et al. "Reinforcement learning with convex constraints". In: Advances in Neural Information Processing Systems 32 (2019)

(5) Redha A. Alliche et al. "prisma-v2: Extension to Cloud Overlay Networks". In: 23rd International Conference on Transparent Optical Networks (ICTON 2023). Bucharest, Romania, July 2023.

Keywords: Multi Agent Learning, Deep Reinforcement Learning, Deep Q Network, packet routing, cloud overlay networks

Deep Learning to detect objectification in films

Julie Tores *† [1,2], Lucile Sassatelli [2,3], Hui-Yin Wu [4], Frédéric Precioso [5]

[1] Modèles et algorithmes pour l'intelligence artificielle – Inria Sophia Antipolis - Méditerranée, Université Nice Sophia Antipolis (1965 - 2019), Laboratoire Jean Alexandre Dieudonné, Scalable and Pervasive softwARe and Knowledge Systems – 2004 route des Lucioles BP 93 06902 Sophia Antipolis, France

[2] Laboratoire d'Informatique, Signaux, et Systèmes de Sophia Antipolis – Université Nice Sophia Antipolis (1965 - 2019), Centre National de la Recherche Scientifique, Université Côte d'Azur – 2000, route des Lucioles - Les Algorithmes - bât. Euclide B 06900 Sophia Antipolis, France

[3] Institut Universitaire de France – Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de l´ducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de l'Education nationale, de l'Enseignement supérieur et de la Recherche – Maison des Universités 103 Boulevard Saint-Michel 75005 Paris, France

[4] Biologically plausible Integrative mOdels of the Visual system : towards synergIstic Solutions for visually-Impaired people and artificial visiON – Inria Sophia Antipolis - Méditerranée, Université Côte d'Azur – 2004 route des LuciolesBP 9306902 Sophia Antipolis, France

[5] Modèles et algorithmes pour l'intelligence artificielle – Inria Sophia Antipolis - Méditerranée, Scalable and Pervasive softwARe and Knowledge Systems, Université Côte d'Azur – 2004 route des Lucioles BP 93 06902 Sophia Antipolis, France

Media plays a crucial role in our perception of sociological constructs, such as gender or race. In movies, gender is often portrayed in a stereotypical manner. Analyses in information and communication science and media studies show that the disparities in how women and men are represented in visual media stem from multimodal film signs (cinematographic, iconographic, textual, sound), which produce gendered characteristics. In particular, such multimodal techniques of mise-en-scène can produce the objectification of a character, who rather appears as an object of desire than a subject of action. This poster presents a part of the work carried out in the ANR TRACTIVE project, which seeks to characterize and quantify character objectification with an AI-driven multimodal approach. Social science partners of the project have established that a person may be objectified because of 12 reasons including the type of shot, visible body parts, speech, voice, activity or narration. Detecting objectification caused by clothing requires using the video while detecting objectification caused by speech requires using the script. That is why objectification detection is a difficult task.

To address the new task of objectification detection defined in this project, we first introduce our augmentation of the MovieGraphs dataset (1) with expert annotations. This dataset consists of 51 movies divided into 7637 clips annotated with graphs that represent who is in each clip, the interactions and relationships between characters, and the reasons behind certain interactions. Each movie clip is annotated with 4 levels of objectification. If objectification is present, at least one of the 12 reasons described previously will be associated to the clip.

*Speaker

†Corresponding author: julie.tores@univ-cotedazur.fr

We then introduce deep models to assess their performance on this new learning task. We first adapt a mono-modal Video-Scene Graph model, VidSGG-BIG (2), based on an encoder-decoder transformer with visual tracks, pre-trained on interaction classification, to detect objectification. We then extend this scene-graph model to benefit from textual information in the form of scripts or subtitles. We compare with the LIReC model (3) first introduced for this film dataset. Finally, we present various ways of combining the visual and textual modalities by building upon the recent foundation models inspired by the Perceiver IO architecture (4).

(1) Vicol, P., Tapaswi, M., Castrejon, L., and Fidler, S. Moviegraphs: Towards understand-ing human-centric situations from videos. In IEEE Conference on Computer Vision and Pattern-Recognition (CVPR 2018).

(2) Gao, K., Chen, L., Niu, Y., Shao, J., and Xiao, J. Classification-then-grounding: Reformulating video scene graphs as temporal bipartite graphs. In IEEE/CVF conferenceon computer vision and pattern recognition (CVPR 2022).

(3) Kukleva, A., Tapaswi, M., and Laptev, I. Learning interactions and relationships betweenmovie characters. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR' 2020)

(4) A. Jaegle et al., Perceiver IO: A General Architecture for Structured Inputs & Outputs. In International Conference on Learning Representations, 2022.

Keywords: Video Analysis, Transformers, Scene Graphs, Multimodality, (Cinematography)

IT jobs outsourcing, IA needs and Firms productivity

Sarah Guillou * [1], Catherine Laffineur * †

[1] Sciences Po - OFCE – Sciences Po – France

Firms will need more and more Information and Telecommunications (IT) skills and production is going to use more and more artificial intelligence (AI) in the process of production itself, in the marketing, in the management of ressources, suppliers, clients as well as of shipments. AI may not be associated with in-house IT jobs because it could be embbeded into software and cloud services, but the increasing digitalization of production is poised to impose the need for having some dedicated in-house IT jobs, by which we mean jobs associated with computer, communication or digital skills. Considering that increasing dimensions of production are impacted by the use of digital tools and by the exploitation of data and that a new source of value of firms can be found in the systematic learning from their daily productive and market information, it is hard to believe that AI is not going to change the way firms produce and grow. We think it should also change the composition of their workforce, and at first, with regards the share of IT jobs. We indeed observe an increase in the share of IT workers in the French whole workforce of business sectors, running from a 3% share in 2009 to 6% in 2019.
The increasing need for IT tasks may entail a change in firm's IT workforce. Some firms might increase their IT jobs proportion, or instead increase their purchase of IT services.

Could firms take the AI turn without in-house IT jobs ? While it appears firms have few reasons to outsource IT jobs, it turns that they do. In France, during the period 2009 to 2019, we observe that 7 to 15% firms each year diminished their share of IT jobs (over their total workforce) while augmenting their overheads costs share.

In this paper, we want to explore what is the impact of this decision on firm's productivity. While we expect in-house IT jobs, at least to deal with AI penetration, to be a driver of productivity gain (Ding et al. 2022, Brynjolfsson et al. 2023, Noy & Zhang 2023), it is also possible that competition for talents on the IT and AI job market may lead the firm to prefer to outsource its IT needs in order to augment the quality of the IT skills it requires. It is then important to observe the average level of skills which were outsourced by the firm. At the same time, the lack of high-skilled IT

workers could create unequality among firms and only the fringe of firms capable of attracting the best IT workers will demonstrate gain in productivity creating an increasing gap between firms.

Based on employer-employee data for France (DADS) as well as balance sheets data (FARE), we build a measure of outsourcing based on Goldschmidt & Schmieder (2017). We use a total factor productivity (TFP) measure borrowing from Forlani et al. (2023) and control the robustness of our results on Levinsohn and Petrin (2003)'s more standard indicator of TFP. We are capable of

*Speaker
†Corresponding author: catherine.laffineur@gredeg.cnrs.fr

building a measure of worker quality (worker fixed effect) using the DADS Panel, which allows to follow workers year-on-year and in which firm they work, based on the Abowd et al. (1999) decomposition.

Our first preliminary results show that the effects of outsourcing on firm productivity are negatively large. Outsourcing IT jobs reduces firms productivity between 5 to 20 percentage points up to four years after the event. It turns that the potential gain in productivity expected from digitalization and AI could be hampered by tensions on IT and AI job markets.

References

Abowd, J. M., Kramarz, F. & Margolis, D. N. (1999), 'High wage workers and high wage firms', Econometrica 67(2), 251–333.

Brynjolfsson, E., Jin, W. & Wang, X. (2023), Information technology, firm size, and industrial concentration, Working Paper 31065, National Bureau of Economic Research.

Ding, X., Fort, T. C., Redding, S. J. & Schott, P. K. (2022), Structural change within versus across firms : Evidence from the united states, Working Paper 30127, National Bureau of Economic Research.

Forlani, E., Martin, R., Mion, G. & Muuls, M. (2023), 'Unraveling Firms : Demand, Productivity and Markups Heterogeneity', The Economic Journal . uead031.

Goldschmidt, D. & Schmieder, J. F. (2017), 'The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure*', The Quarterly Journal of Economics 132(3), 1165–1217. URL: https ://doi.org/10.1093/qje/qjx008

Noy, S. & Zhang, W. (2023), 'Experimental evidence on the productivity effects of generative artificial intelligence', Science 381(6654), 187–192.

Keywords: IT jobs, AI job market, outsourcing, productivity, workforce composition

Cross-modal distillation to predict human attention in VR

Quentin Guimard *† [1], Lucile Sassatelli [2,3]

[1] Université Côte d'Azur – Université Côte d'Azur, CNRS, Laboratire I3S, Sohpia-Antipolis, France – France
[2] Université Côte d'Azur – Université Côte d'Azur, CNRS, Laboratire I3S, Sohpia-Antipolis, France – France
[3] Institut Universitaire de France – Ministère de l'Education nationale, de l'Enseignement supérieur et
de la Recherche, Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche,
Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de
lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de lÉducation
nationale, de l'Enseignement supérieur et de la Recherche, Ministère de lÉducation nationale, de
l'Enseignement supérieur et de la Recherche, Ministère de lÉducation nationale, de l'Enseignement
supérieur et de la Recherche, Ministère de lÉducation nationale, de l'Enseignement supérieur et de la
Recherche, Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère
de lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de lÉducation
nationale, de l'Enseignement supérieur et de la Recherche, Ministère de lÉducation nationale, de
l'Enseignement supérieur et de la Recherche, Ministère de lÉducation nationale, de l'Enseignement
supérieur et de la Recherche, Ministère de lÉducation nationale, de l'Enseignement supérieur et de la
Recherche, Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche,
Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de
lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de lÉducation
nationale, de l'Enseignement supérieur et de la Recherche, Ministère de lÉducation nationale, de
l'Enseignement supérieur et de la Recherche, Ministère de lÉducation nationale, de l'Enseignement
supérieur et de la Recherche, Ministère de lÉducation nationale, de l'Enseignement supérieur et de la
Recherche, Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche,
Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de
lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de lÉducation
nationale, de l'Enseignement supérieur et de la Recherche, Ministère de lÉducation nationale, de
l'Enseignement supérieur et de la Recherche, Ministère de lÉducation nationale, de l'Enseignement
supérieur et de la Recherche, Ministère de lÉducation nationale, de l'Enseignement supérieur et de la
Recherche, Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche,
Ministère de lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de
lÉducation nationale, de l'Enseignement supérieur et de la Recherche, Ministère de lÉducation
nationale, de l'Enseignement supérieur et de la Recherche, Ministère de lÉducation nationale, de
l'Enseignement supérieur et de la Recherche, Ministère de lÉducation nationale, de l'Enseignement
supérieur et de la Recherche, Ministère de lÉducation nationale, de l'Enseignement supérieur et de la
Recherche, Ministère de l'Education nationale, de l'Enseignement supérieur et de la Recherche – France

With the rise in popularity of virtual reality (VR) and the very high resolution of newer head-mounted displays, dynamically adapting the quality to the user is critical, as not everything can be downloaded or rendered in full quality. Predicting human attention in VR environments is key to maximizing the quality of experience.
To predict the area seen by users in VR (viewport), recent works use deep learning models that learn to predict future head positions from past positions (1) as well as saliency maps extracted from 360∘ videos (2). However, we may be able to use more information from diverse modalities to make better predictions. For instance, it has been shown that the emotional state of the user

*Speaker
†Corresponding author: quentin.guimard@univ-cotedazur.fr

influences movements when watching immersive content (3). This emotional state may depend on content and other factors and is not explicitly considered in recent state-of-the-art viewport prediction methods (1, 2).
It is possible to measure physiological signals such as electrodermal activity (EDA) or heart rate that vary as the user's emotional state changes. Recently, datasets have been published where physiological signals, as well as continuous emotional annotations, were collected while participants were watching 360◦ immersive content (4). Recent work shows that it is possible to do fine-grained emotion recognition from these physiological signals using deep learning.
How can we learn from additional modalities, such as emotions and physiological signals, to improve the prediction of human attention in VR? How can we ensure that the knowledge learned from these additional modalities can be transferred at test time, when the model may not have access to such information?
The objective of this poster is to present ongoing work on an architecture that enables cross-modal distillation to better predict human attention in VR.
Models that use cross-modal distillation have been shown to outperform their non-distilled counterparts for various tasks (5) because they are able to take advantage of the extra information present in other modalities to learn a better latent representation.
Building upon this previous work, we present a cross-modal Transformer-based architecture able to benefit from all positional, VR content and emotional data to predict human motion, which can be distilled to a student model. We perform a thorough evaluation of the student models when no emotional data is available and compare against state-of-the-art methods for viewport prediction.

(1) Quentin Guimard, Lucile Sassatelli, Francesco Marchetti, Federico Becattini, Lorenzo Seidenari, and Alberto Del Bimbo. "Deep variational learning for multiple trajectory prediction of 360◦ head movements." In Proceedings of the 13th ACM Multimedia Systems Conference, pp. 12-26. 2022.
(2) Miguel Fabián Romero Rondón, Lucile Sassatelli, Ramón Aparicio-Pardo, and Frédéric Precioso. "TRACK: A New Method From a Re-Examination of Deep Architectures for Head Motion Prediction in 360◦ Videos." IEEE Transactions on Pattern Analysis and Machine Intelligence 44, no. 9 (2021): 5681-5699.
(3) Quentin Guimard, and Lucile Sassatelli. "Effects of emotions on head motion predictability in 360◦ videos." In Proceedings of the 14th International Workshop on Immersive Mixed and Virtual Environment Systems, pp. 37-43. 2022.
(4) Quentin Guimard, Florent Robert, Camille Bauce, Aldric Ducreux, Lucile Sassatelli, Hui-Yin Wu, Marco Winckler, and Auriane Gros. "PEM360: A dataset of 360 videos with continuous Physiological measurements, subjective Emotional ratings and Motion traces." In Proceedings of the 13th ACM Multimedia Systems Conference, pp. 252-258. 2022.
(5) Zineng Tang, Jaemin Cho, Hao Tan, and Mohit Bansal. "Vidlankd: Improving language understanding via video-distilled knowledge transfer." Advances in Neural Information Processing Systems 34 (2021): 24468-24481.

Keywords: deep learning, virtual reality