Munich Center for Machine Learning

Machine Learning Conference

at Imperial College London

Advances in Foundations, Language, Vision and Medicine

Program

13-14th April 2026 |
at Imperial College London

CONFERENCE PROGRAM – OVERVIEW

DAY 1, MONDAY, APRIL 13TH 2026		
1 - 1.45pm:	**Arrival and check-in at the conference** *Location: Huxley Building, 180 Queen's Gate, rooms 341 and 342*	
2 pm:	**Welcome by the organisers (Imperial College & MCML)**	
2.10 pm:	**Keynote by** *Daniel Rückert (MCML): AI and the Future of Medicine*	
2.35 pm:	**Keynote by** *Francesca Toni (Imperial College): Debating Explainable AI*	
WORKSHOPS IN PARALLEL SETTINGS		
	I) AI in Medicine: Applications and Intersections	***II) Natural Language Processing, Computer Vision, Foundations***
03:00 pm:	**Julia Schnabel (MCML)** *Precision MR(A)I - from imaging sensor to making sense of images*	**Barbara Plank (MCML)** *Human-centered LLMs for Inclusive Language Technology*
03:20 pm:	**Chen Qin (Imperial)** *Learning Across Modalities Within and Beyond Medical Imaging*	**Tolga Birdal (Imperial)** *Higher-order (Topological) Representation Learning for Computer Vision and AI4Science*
03:40 pm:	**Nassir Navab (MCML)** *AI for Sensing, Perception, Decision, and Action in High-Intensity, Dynamic Surgical Environments*	**Daniel Cremers (MCML)** *3D Computer Vision for Autonomous Systems*
04:00 pm:	**Ben Glocker (Imperial)** *Imaginable Imaging: The Role of Causality in Medical Imaging AI*	**Alessio Lomuscio (Imperial)** *Robustness Verification of Machine Vision Systems*
04:20 pm:	**Andreas Triantafyllopoulos (MCML)** *Speech & language analysis for reliable & interpretable mental health assessments*	**Michael Hedderich (MCML)** *Technical and Human-Centric Perspectives on Understanding and Controlling LLM Behavior*
04:40 pm:	**Fabrizio Russo (Imperial)** *Collaborative Causal Discovery for AI-Aided Decision-Making*	**Florian Eichin (MCML)** *ExPLAIND: Unifying Model, Data, and Training Attribution to Study Model Behavior*
05:00 to 06:00 pm:	**NETWORKING WITH SNACKS**	

CONFERENCE PROGRAM

DAY 2, TUESDAY, APRIL 14TH 2026		
9:00 am:	**Registration, check-in** *Location: Huxley Building, 180 Queen's Gate, rooms 341 and 342*	
9.30 am:	**Keynote by Imperial College:** *Aldo Faisal: Nightingale AI*	
10.00 am:	**Keynote by MCML:** *Björn Ommer: Efficient Models for Visual Intelligence*	
WORKSHOPS IN PARALLEL SETTINGS		
	I) AI in Medicine: Applications and Intersections	***II) Natural Language Processing, Computer Vision, Foundations***
10:30 am:	**Björn Schuller (MCML/Imperial)** *Multimodal Foundation Models for Medicine: From Signals to Care*	-
10:50 am:	**Qingjie Meng (University of Birmingham)** *Population-Scale Cardiac Digital Twins from Clinical Imaging*	**Pier Luigi Dragotti (Imperial)** *AI for scientific imaging*
11:10 am:	**Björn Eskofier (MCML)** *AI-supported medical therapy decisions @ LMU Munich*	**Nefta Kanilmaz (MCML)** *Structure-Preserving Clustering of Event Sequences*
11:30 am:	**Xinzhe Luo (Imperial)** *Breaking the Ground-Truth Barrier: Unsupervised Parallel MRI Reconstruction via Projected Conditional Flow Matching*	**Alessandra Russo (Imperial)** *From Symbols to Systems: How Neuro-Symbolic AI Can Deliver Interpretable and robust AI at Scale*
11:50 am:	**David Bani-Harouni (MCML)** *Language Models for Clinical Decision Making*	**Vincent Fortuin (UTN)** *Decision-oriented Uncertainty Quantification: Pathologies and Promises*
12:10 pm:	**Ayush Bandari (Imperial)** *Noise for Humans, Signal for Machines*	**Yingzhen Li (Imperial)** *Variational Uncertainty Decomposition in In-Context Learning*
12:30 pm:	**LUNCH BREAK**	

CONFERENCE PROGRAM

DAY 2, TUESDAY, APRIL 14TH 2026

1:30 pm- 2.15 pm:	**PANEL TOPIC DISCUSSION** *"AI for Science: The Do's and Don'ts"*	
Moderator: **Maja Sophie Klimaschewski** *(Bavaria in the UK)*	**Panelists:** **Barbara Plank** *(MCML)*, **Daniel Cremers** *(MCML)*, **Jason McEwen** *(Alan Turing Institute/ UCL)*, **Chen Qin** *(Imperial)*	
WORKSHOPS IN PARALLEL SETTINGS		
	I) AI in Medicine: Applications and Intersections	***II) Natural Language Processing, Computer Vision, Foundations***
02:30 pm:	**Jun Li (MCML)** *Vision-Language Models for Medical Imaging: Teaching AI to Perceive and Explain*	**Marek Rei (Imperial)** *The Reality Gap in LLM Reasoning and Evaluation*
02:50 pm:	**Emily Sullivan (University of Edinburgh):** *Alignment for AI for Science*	**Felicia Körner (MCML)** *Investigating the Emergence of Shared Concept Spaces in Multilingual LLMs*
03:10 pm:	**Julian Suk (MCML)** *Physics-informed graph neural networks for flow field estimation in carotid arteries*	**Deniz Gündüz (Imperial)** *Learning to Compress Without Learning to Reconstruct*
03:30 pm:	**COFFEE BREAK**	
03:50 pm:	**Adam Dejl (Imperial)** *Argumentation for Explainable and Globally Contestable Treatment Decision Support with LLMs*	**Andrew Davison (Imperial)** *A Robot Web for Many-Robot Localisation and Coordination*
04:10 pm:	**Martin Menten (MCML)** *Missing Capabilities of LLMs for Medical Image Analysis*	**Dominik Muhle (MCML)** *Self-Supervised (Semantic) Scene Completion*
04:30 pm:	**Robert Richer (MCML)** *AI-enabled Assessment of Human Physiology and Psychology using Contactless Digital Biomarkers*	**Martin Binder (MCML)** *CASHing In on the Rashomon Effect: Extending Rashomon Sets Across Model Classes*
04:50 pm:	***End of workshop (Day 2)***	

Book of Abstracts

Conference Day 1

Keynote by the MCML: Daniel Rückert: AI and the Future of Medicine
Over the past decade, AI has catalyzed transformative advances across the sciences. In medicine, AI has transitioned from inflated expectations to robust clinical utility, particularly in radiology, which currently accounts for 76% of FDA-authorized AI medical devices. This presentation examines the trajectory of AI toward enabling personalized diagnosis, disease stratification, and precise prognostic evaluation. Key developments include AI-driven image reconstruction techniques that accelerate MR image acquisition and promise to democratize global access to medical imaging. Furthermore, the integration of multimodal data architectures is unlocking latent diagnostic value in highly accessible modalities. Novel contrastive learning frameworks successfully align cost-effective electrocardiograms with high-fidelity cardiac MR imaging, significantly enhancing the prediction of cardiovascular disease. At the population level, large-scale analysis demonstrates that combining whole-body MRI, radiomics, and comprehensive clinical records substantially improves risk assessment for a diverse set of diseases. The emergence of Large Language Models (LLMs) presents a paradigm shift. However, while state-of-the-art LLMs achieve passing scores on standardized medical examinations, they exhibit critical limitations in real-world clinical decision-making, including poor adherence to clinical guidelines, inadequate laboratory interpretation, and sensitivity to information ordering. To bridge this gap, the development of specialized Vision-Language Models (VLMs), such as RetinaVLM, demonstrates superior clinical utility in biomarker recognition and disease staging. Realizing the full potential of clinical AI necessitates addressing profound technical and societal challenges. Standard federated learning paradigms remain vulnerable to adversarial interference, requiring advanced privacy-preserving frameworks that integrate differential privacy and secure aggregation. Beyond technical data security, the integration of AI into healthcare must navigate the fundamental limits of clinical predictability, characterized by inherent inter-observer variability. Finally, strategic focus must be directed toward mitigating the risk of clinician deskilling, establishing robust regulatory compliance, and ensuring that AI deployment fosters equitable care.

Keynote by Francesca Toni: Debating Explainable AI
The need for explainability in AI is widely agreed upon as crucial towards safe and trustworthy deployment of AI systems, especially given the plethora of opportunities for undesired behaviour, including misinformation, hallucination and bias. In this talk I aim at debating explainable AI, advocating in particular explainable AI approaches based on computational argumentation that can (1) interact to progressively explain outputs and/or reasoning as well as assess grounds for contestation provided by humans and/or other machines, and (2) revise decision-making processes to redress any issues successfully raised during

contestation. I will ground the talk in LLM-based claim verification, image classification, bias detection and causal discovery.

Workshop: AI in Medicine: Applications and Intersections

Julia Schnabel: MR(A)I - from imaging sensor to making sense of images
Medical imaging is undergoing a paradigm shift, moving from a fragmented pipeline of image acquisition, reconstruction and analysis toward integrated, intelligent imaging solutions. This talk explores the journey of embedding AI into Magnetic Resonance Imaging, from the imaging sensor to image interpretation. Starting from self-supervised and physics-informed deep learning approaches, high-quality image reconstructions from undersampled or motion-corrupted sensor data are obtained. Further down the imaging pipeline, vision-tabular models enable the transition from unimodal image analysis to clinically-informed multimodal learning. Leveraging patient-level tabular data such as demographics and clinical history to guide the learning of visual representations, semantic similarities between different patients that visual data alone might miss can be found. Using Precision MR(A)I, we can thus transform raw MR signals into precise, actionable insights, ultimately paving the way for more robust and personalised diagnostic reasoning

Chen Qin: Learning Across Modalities Within and Beyond Medical Imaging
Abstract: Clinical decision-making relies on synthesizing diverse streams of patient information, yet developing AI to handle real-world, heterogeneous data remains a major challenge. Following a brief overview of our broader medical imaging pipeline, this talk explores how we learn across different modalities for robust image analysis and downstream diagnosis. We frame this as a progressive multimodal learning problem. First, we address information fusion within the imaging domain. By utilizing multi-contrast MRI, we demonstrate how leveraging complementary information can robustly align images and enhance overall image quality. Second, we expand this fusion beyond the pixel, introducing architectures that combine high-dimensional medical images with incomplete tabular electronic health records. By deploying advanced representation learning strategies that natively handle missing clinical variables, we provide grounded steps toward learning the robust representations required for the complex data found in real-world clinical environments.

Nassir Navab: **AI for Sensing, Perception, Decision, and Action in High-Intensity, Dynamic Surgical Environments**
Over the past decade, rapid advances in machine learning have profoundly transformed many domains, with growing impact in medicine. In this talk, I will explore the opportunities and challenges of integrating these technologies into computer-assisted interventions, where safety, robustness, and clinical acceptance are paramount.

The focus will be on Surgical Data Science, holistic operating room (OR) modeling, and the development of surgical world models that enable context-aware perception and decision-making. A central theme is the need to foster trust and acceptance of machine learning and robotic systems in clinical practice—an objective that often requires a progression through Intelligence Amplification (IA) before achieving higher levels of autonomy. In this context, Augmented Reality (AR) plays a key role by enhancing human perception and cognition, providing intuitive interfaces that bridge human expertise and machine intelligence.
Building on research conducted at the Chair for Computer-Aided Medical Procedures and the Munich Center for Machine Learning (MCML), I will present a set of novel approaches addressing the unique challenges of high-intensity surgical environments. These include robotic ultrasound imaging, multimodal data fusion, and semantic scene graphs for holistic modeling of the surgical domain, ultimately contributing to the construction of surgical world models.
Finally, I will demonstrate how AR-driven solutions can act as catalysts for the adoption of AI in the operating room, enabling a gradual transition from human-centered assistance and telemanipulation toward AI-supported decision-making and autonomous action. This talk highlights a unifying perspective at the intersection of machine learning, computer vision, intelligent user interfaces, and medical robotics, outlining a pathway from intelligence amplification to artificial intelligence in future computer-assisted interventions.

Ben Glocker: Imaginable Imaging: The Role of Causality in Medical Imaging AI
Abstract: In this talk, I will discuss how causal reasoning and counterfactual image synthesis can strengthen the robustness and reliability of medical imaging AI. I will briefly outline our methods for generating high fidelity counterfactual images and demonstrate how they enable counterfactual contrastive learning, targeted stress testing, and more resilient image segmentation in the presence of pathology. I will discuss the role of causality in better understanding key challenges such as performance drift under data distribution shift. The talk will conclude with perspective on how causal AI may lead to more trustworthy algorithms that are safe for clinical deployment.

Andreas Triantafyllopoulos: **Speech & language analysis for reliable & interpretable mental health assessments**
In recent years, there has been increasing interest in using speech and language analysis for mental health assessments. While there is substantial evidence to show that mental disorders manifest in how people talk and what they say, ensuring the reliability of trained model remains a significant open problem. This talk will highlight recent developments and open problems in this direction.

Fabrizio Russo*:* Collaborative Causal Discovery for AI-Aided Decision-Making
This talk introduces CausAlly, a collaborative causal discovery framework for AI-aided decision-making that brings together expert knowledge and data to build, challenge, and refine causal models. The talk will show how causal graphs, argumentative reasoning, and explanatory representations can support uncertainty-aware, transparent, and contestable AI systems, enabling clinicians,

researchers, and domain experts to interrogate assumptions, compare alternative causal hypotheses, and generate more trustworthy insights for decision-making.

Workshop: Natural Language Processing, Computer Vision, Foundations

Barbara Plank: **Human-centered LLMs for Inclusive Language Technology**
Large Language Models (LLMs) have advanced rapidly but often still cater only to a narrow set of users. This talk advocates for a human-centered approach to NLP technology—one that embraces linguistic variation, improves reasoning and safety, and better serves diverse language communities. In this talk I will outline paths toward more inclusive and trustworthy language technology.

Tolga Birdal: **Higher-order (Topological) Representation Learning for Computer Vision and AI4Science**
This is a rapidly growing field that pertains to the development of deep learning models for data supported on topological domains such as simplicial complexes, cell complexes, and hypergraphs, which generalize many domains encountered in scientific computations. In this talk, Tolga will present a unifying deep learning framework built upon an even richer data structure that includes widely adopted topological domains.
Specifically, he will begin by introducing combinatorial complexes, a novel type of topological domain. Combinatorial complexes can be seen as generalizations of graphs that maintain certain desirable properties. Similar to hypergraphs, combinatorial complexes impose no constraints on the set of relations. In addition, combinatorial complexes permit the construction of hierarchical higher-order relations, analogous to those found in simplicial and cell complexes. Thus, combinatorial complexes generalize and combine useful traits of both hypergraphs and cell complexes, which have emerged as two promising abstractions that facilitate the generalization of graph neural networks to topological spaces.
Second, building upon combinatorial complexes and their rich combinatorial and algebraic structure, Tolga will develop a general class of message-passing combinatorial complex neural networks (CCNNs), focusing primarily on attention-based CCNNs. He will additionally characterize permutation and orientation equivariances of CCNNs, and discuss pooling and unpooling operations within CCNNs. The performance of CCNNs on tasks related to mesh shape analysis and graph learning will be provided. The experiments demonstrate that CCNNs have competitive performance as compared to state-of-the-art deep learning models specifically tailored to the same tasks. These findings demonstrate the advantages of incorporating higher-order relations into deep learning models and shows great promise for AI4Science.

Daniel Cremers: Convolutional Networks beyond the Euclidean Space
Over the past decade, the work of Daniel Cremers has helped redefine how machines perceive the three-dimensional world, pioneering a shift from sparse, feature-based pipelines to dense, direct formulations of visual SLAM. This talk presents a unifying vision in which direct and deep SLAM fuse geometric

optimization with learned representations, enabling autonomous systems to achieve unprecedented levels of accuracy, robustness, and adaptability in complex, real-world environments.

Michael Hedderich: **Technical and Human-Centric Perspectives on Understanding and Controlling LLM Behavior**

Large Language Models (LLMs) and Generative AI have transformed the field of artificial intelligence and are drastically lowering the entrance barrier for building AI applications. However, the black-box nature of AI systems presents critical challenges in understanding the behavior and effectively controlling their outputs, both for AI-experts and novice users. For example, even minor prompt variations can lead to unexpected and problematic behaviors, like the emergence of strong biases. In this talk, I'll present interdisciplinary perspectives from machine learning, natural language processing and human-computer interaction on how to support users to comprehensively evaluate and understand, as well as effectively control the behavior of AI technology; thus enabling diverse, reliable and safe applications of AI.

Alessio Lomuscio: Robustness Verification of Machine Vision Systems
Ensuring the reliability of ML-based computer vision remains a significant hurdle, as traditional extensive testing often fails to provide the rigorous coverage required for safety-critical deployment. This talk explores verification methods that offer formal guarantees of model performance within dense neighborhoods of selected inputs, addressing diverse perturbations ranging from geometric transformations to semantic shifts in latent space. By integrating these verification techniques with robust learning, we can develop models that are inherently more resilient than those produced by standard methods. The presentation will provide a high-level overview of the key algorithmic advancements that have enabled these methods to scale to increasingly complex systems, including modern transformer architectures and vision-language models.

Florian Eichin: ExPLAIND: Unifying Model, Data, and Training Attribution to Study Model Behavior
Post-hoc interpretability methods typically attribute a model's behavior to its components, data, or training trajectory in isolation. This leads to explanations that lack a unified view and may miss key interactions. While combining existing methods or applying them at different training stages offers broader insights, such approaches usually lack theoretical support.
In this talk, I will present ExPLAIND, a unified framework that integrates all these perspectives. Jointly interpreting model components and data over the training process, I will present some of our most recent findings on generalization in smaller settings like Grokking, as well as in LLMs generalizing over multilingual data.

Conference Day 2

Workshop: AI in Medicine: Applications and Intersections

Björn Schuller: Multimodal Foundation Models for Medicine: From Signals to Care
Multimodal Foundation Models are beginning to transform medicine and healthcare not only through better diagnostics, but through a broader capacity to understand the rich signals of human health and behaviour. From speech and language to physiological, behavioural, and clinical data, multimodal foundation models offer new ways to detect risk earlier, monitor change more continuously, and support more personalised care and intervention across diverse medical domains. This talk explores how recent advances in multimodal and generative AI can help bridge the gap between raw data and meaningful clinical insight in everyday data. It will highlight emerging opportunities in areas such as mental health, neurology, and rare diseases, where subtle patterns across modalities may reveal changes in health and wellbeing that are difficult to capture through conventional assessment alone. At the same time, the talk will address the critical translational questions that will determine real-world impact: robustness, efficiency, explainability, fairness, privacy, and clinical validation. By connecting methodological progress with pressing healthcare needs, the presentation outlines a vision of AI in medicine that moves beyond narrow task automation toward systems that are preventive, interventive, adaptive, and human-centred. The ultimate goal is not simply smarter models, but more responsive and effective care.

Qingjie Meng: Population-Scale Cardiac Digital Twins from Clinical Imaging
Current clinical workflows primarily rely on 2D imaging for cardiac examinations, such as cardiac cine magnetic resonance (CMR) . These modalities provide only 2D views, which severely limits our understanding of the true 3D structure and dynamic motion of the heart. At the same time, the high privacy of medical data lead to data scarcity, further hindering progress in cardiac research. This talk will present recent advances by our team in AI-powered cardiac modelling, covering several works accepted by IEEE Transactions on Medical Imaging and MICCAI. The content includes 3D digital heart reconstruction from multi-view 2D CMR images, 3D cardiac motion tracking methods, and the use of diffusion models to generate high-quality dynamic medical images (e.g., echocardiography videos). These methods demonstrate significant advantages in modelling accuracy, data generation capability, and clinical applicability, offering new directions for the application of AI in cardiovascular image analysis.

Björn Eskofier: AI-supported medical therapy decisions @ LMU Munich
Artificial Intelligence (AI) and Machine Learning (ML) methods are currently a “hot topic” in medicine. The driver of AI or ML method employment, in medicine as well as in other domains, is the availability of digital data. Here, the potential of delivering more objective, precise, and personalized medical diagnosis and care decisions is by far not reached. The reason is that the current healthcare data infrastructure, both nationally and internationally, lacks interoperability and

interfaces on several different levels (individual, institutional, device, and provider level, just to name a few). The health data infrastructure for future healthcare needs to address this. One favored solution is “personal health dataspaces”, which put individuals at the center of health data (figure) and create new opportunities for AI and ML applications.
The talk will present the core idea of personal health dataspaces and hint at some opportunities for the future AI-driven “digital” healthcare system that emerge from it. It will open up new possibilities in healthcare, which will hopefully contribute to delivering more objective, precise, and personalized medical diagnosis and care decisions.

Xinzhe Luo: **Breaking the Ground-Truth Barrier: Unsupervised Parallel MRI Reconstruction via Projected Conditional Flow Matching**
Reconstructing high-quality images from undersampled MRI data is a challenging inverse problem. Supervised deep learning requires impractical fully sampled ground-truth data, while current unsupervised methods struggle at high acceleration rates. We introduce UPMRI, an unsupervised framework utilizing Projected Conditional Flow Matching (PCFM). UPMRI learns the prior distribution of fully sampled multi-coil MRI using only undersampled k-space measurements. By theoretically linking measurement-space vector fields to the PCFM objective, we develop a robust dual-space cyclic sampling algorithm. Evaluations on brain and cardiac datasets demonstrate UPMRI significantly outperforms unsupervised baselines and matches leading supervised methods without needing any fully sampled training data.

David Bani-Harouni: Language Models for Clinical Decision Making
Clinical decision making is a dynamic, iterative process in which clinicians draw on both medical knowledge and clinical experience to formulate hypotheses, select diagnostic tests, and refine their differential diagnosis under uncertainty and resource constraints. This talk explores how large language models can be trained to replicate this process by employing Reinforcement Learning for cost-effective test selection and for calibrated confidence expressions. Thus, enabling LLMs to perform efficient differential diagnosis and accurately assess hypothesis probabilities.

Ayush Bandari: Noise for Humans, Signal for Machines
As AI moves ever closer to the physical world, the question of how machines acquire data becomes as important as how they process it. No matter how sophisticated the algorithms, the quality of inference is ultimately bounded by the quality of acquisition. Yet conventional digital sensing remains grounded in the classical Shannon–Nyquist paradigm, where quantization is viewed as inherently lossy and quantization noise as an undesirable artifact of digitization. This talk challenges that foundation. We will show that signals can be recovered purely from quantization noise, reversing the long-standing intuition that this seemingly discarded part of the measurement is useless. In the Unlimited Sensing Framework, what was once treated as loss becomes a powerful digital representation for machines. Through examples in audio, medical imaging, and

radar, we will argue that future AI systems may require not only better algorithms, but new principles for machine-native sensing, perception, and inference.

Jun Li: **Vision-Language Models for Medical Imaging: Teaching AI to Perceive and Explain**
Large vision-language models have shown strong potential for medical imaging, but their performance often remains unreliable under real-world clinical conditions such as limited supervision, long-tailed disease distributions, and domain shift. In this talk, I will present recent work on vision-language models for medical imaging with a focus on perception, explanation, and robustness. In particular, I will highlight recent work on abnormality grounding and test-time adaptation, and discuss how these approaches can help make multimodal models more accurate, reliable, and clinically useful.

Emily Sullivan: Alignment for AI for Science
This talk is about the need for aligning AI for science with scientific values: traceability, coherence, verifiability, calibration, and intelligibility. The talk will also discuss the social epistemic value of science and presents a skeptical narrative toward automating science.

Julian Suk: "Physics-informed graph neural networks for flow field estimation in carotid arteries"
Abstract: Hemodynamic quantities are valuable biomedical risk factors for cardiovascular pathology such as atherosclerosis. Non-invasive, in-vivo measurement of these quantities can only be performed using a select number of modalities that are not widely available, such as 4D flow magnetic resonance imaging (MRI). In this work, we create a surrogate model for hemodynamic flow field estimation, powered by machine learning. We train graph neural networks that include priors about the underlying symmetries and physics, limiting the amount of data required for training. This allows us to train the model using moderately-sized, in-vivo 4D flow MRI datasets, instead of large in-silico datasets obtained by computational fluid dynamics (CFD), as is the current standard. We create an efficient, equivariant neural network by combining the popular PointNet++ architecture with group-steerable layers. To incorporate the physics-informed priors, we derive an efficient discretisation scheme for the involved differential operators. We perform extensive experiments in carotid arteries and show that our model can accurately estimate low-noise hemodynamic flow fields in the carotid artery. Moreover, we show how the learned relation between geometry and hemodynamic quantities transfers to 3D vascular models obtained using a different imaging modality than the training data. This shows that physics-informed graph neural networks can be trained using 4D flow MRI data to estimate blood flow in unseen carotid artery geometries.

Adam Dejl: **Argumentation for Explainable and Globally Contestable Treatment Decision Support with LLMs**

Large language models (LLMs) exhibit strong general capabilities, but their deployment in high-stakes domains is hindered by their opacity and unpredictability. Recent work has taken meaningful steps towards addressing these issues by augmenting LLMs with symbolic post-hoc reasoning based on computational argumentation, providing faithful explanations and enabling users to contest incorrect decisions. However, this paradigm is limited to pre-defined binary choices and only supports local contestation for specific instances, leaving the underlying decision logic unchanged and prone to repeated mistakes. In this talk, I will introduce ArgEval, a framework that shifts from instance-specific reasoning to structured evaluation of general decision options. Rather than mining arguments solely for individual cases, ArgEval systematically maps task-specific decision spaces, builds corresponding option ontologies, and constructs general argumentation frameworks (AFs) for each option. These frameworks can then be instantiated to provide explainable recommendations for specific cases while still supporting global contestability through modification of the shared AFs. The effectiveness of ArgEval has been demonstrated in the context of providing treatment recommendations for glioblastoma, an aggressive brain tumour, where it was found to produce explainable guidance aligned with clinical practice.

Martin Menten: Missing Capabilities of LLMs for Medical Image Analysis
Driven by exceptional interest in both research and real-world applications across many fields, the capabilities of large language models (LLMs) continue to advance rapidly. However, the requirements for LLMs in medical image analysis differ from those of general-purpose language models.This talk will establish that the current generation of LLMs still lacks several critical capabilities required for the reliable analysis of medical images. It will then present a series of recent works that aim to enhance LLMs in three key areas: numeracy, pixel-wise segmentation, and domain-specific medical knowledge.

Robert Richer: AI-enabled Assessment of Human Physiology and Psychology using Contactless Digital Biomarkers
Recent advances in sensing and machine learning enable the contactless assessment of human psychophysiology through video, audio, and radar data. This talk presents a series of works demonstrating how digital biomarkers derived from body movements, pulse wave propagation, and respiration can be used to capture stress, pain, and related states. It will show how movement patterns (e.g., stress-induced movement inhibition), facial dynamics, and remotely extracted cardiovascular and respiratory signals can be jointly modeled to characterize autonomic and neuroendocrine responses. Together, these approaches outline a scalable framework for AI-enabled psychophysiological assessment, linking observable behavior and physiology to underlying regulatory processes in naturalistic environments.

Workshop: Natural Language Processing, Computer Vision, Foundations

Pier Luigi Dragotti: AI for scientific imaging
Today, machine learning systems provide state-of-the-art solutions to a wide range of imaging problems in science and medicine. However, these approaches typically rely on access to large quantities of ground-truth data. This requirement poses a significant limitation for exploratory scientific imaging, where the objective is to capture phenomena that have not been previously observed.
In this talk, we present several case studies demonstrating that incorporating acquisition physics and prior knowledge about the imaged phenomena into deep neural network models enables the development of scalable algorithms. These approaches achieve state-of-the-art performance using semi-supervised and self-supervised learning techniques. In particular, we demonstrate, for the first time, the monitoring of activity of large populations of neurons using light-field microscopy, as well as the resolution of molecular structures using electron microscopy, specifically cryogenic electron microscopy.

Nefta Kanilmaz: Structure-Preserving Clustering of Event Sequences
Event logs capturing real world processes arise across many domains and are characterized by high variability. Unsupervised clustering of such sequences is a key technique for detecting underlying structures. However, existing approaches often rely on vector-based representations that disregard the sequential nature of the data and lose important temporal dependencies. This talk presents a structure-preserving framework for unsupervised clustering of event sequences. Rather than embedding into a vector space, our approach operates directly on sequences and represents clusters as model-based prototypes using alignment-based similarity. This enables the discovery of clusters based on behavioral similarity rather than feature overlap. We will introduce the method, discuss empirical results, and conclude with an outlook on complementary research directions, in particular active learning for event sequences with a focus on outcome prediction.

Alessandra Russo: From Symbols to Systems: How Neuro-Symbolic AI Can Deliver Interpretable and robust AI at Scale
AI is making rapid, transformative advances, yet its adoption in domains such as healthcare, security, robotics, and other safety-critical settings remains constrained by the need for models that are transparent, robust and capable of generalising from data in a reliable manner. Deep learning and foundation models still struggle to address these challenges. Symbolic AI methods offer a complementary approach: they can learn human-understandable models from data, grounded in domain knowledge, that generalise across tasks, and support robust inference even under noise and uncertainty. In this talk, I will first outline progress in symbolic machine learning. I will then describe how it can serve as a core component of a neuro-symbolic architecture to improve LLM robustness through semantic control of the generative process.

Vincent Fortuin: Decision-oriented Uncertainty Quantification: Pathologies and Promises

Uncertainty quantification is an important task in machine learning and is often motivated by virtue of making better, more reliable decisions. However, when benchmarking uncertainty quantification algorithms, researchers often use a range of different metrics. How these metrics relate to downstream decisions has long been unclear. In this talk, I shed light on how decision-aligned common uncertainty metrics are (spoiler: they are not), and I propose a path towards designing more decision-aligned metrics.

Yingzhen Li: Variational Uncertainty Decomposition in In-Context Learning

As large language models (LLMs) gain popularity in conducting prediction tasks in-context, understanding the sources of uncertainty in in-context learning becomes essential to ensuring reliability. The recent hypothesis of in-context learning performing predictive Bayesian inference opens the avenue for Bayesian uncertainty estimation, particularly for decomposing uncertainty into epistemic uncertainty due to lack of in-context data and aleatoric uncertainty inherent in the in-context prediction task. However, the decomposition idea remains under-explored due to the intractability of the latent parameter posterior from the underlying Bayesian model. In this work, we introduce a variational uncertainty decomposition framework for in-context learning without explicitly sampling from the latent parameter posterior, by optimising auxiliary queries as probes to obtain an upper bound to the aleatoric uncertainty of an LLM's in-context learning procedure, which also induces a lower bound to the epistemic uncertainty. Through experiments on synthetic and real-world tasks, we show quantitatively and qualitatively that the decomposed uncertainties obtained from our method exhibit desirable properties of epistemic and aleatoric uncertainty.

Marek Rei: The Reality Gap in LLM Reasoning and Evaluation
In this talk, we investigate the gap between perceived LLM capabilities and real-world reasoning performance, highlighting the vulnerabilities of automated evaluation paradigms. We will discuss AgentCoMa, a novel compositional benchmark evaluating mixed-type reasoning. Experiments on this benchmark show that LLMs can successfully solve tasks that focus on either commonsense or mathematical reasoning, but suffer a substantial performance drop when a single task requires combining different types of reasoning. We also investigate the robustness of the LLM-as-a-judge evaluation paradigm, which is widely used in major benchmarks. A new method of Reinforcement Learning for Reverse Engineering (RLRE) allows us to perform RL updates straight through a frozen LLM, based on the feedback from the benchmark judge. Using this technique, we show how an adversarial preamble generator can be trained to improve evaluation results, without actually changing the weights of the main LLM.

Felicia Körner: **Investigating the Emergence of Shared Concept Spaces in Multilingual LLMs**

Training Large Language Models (LLMs) with high multilingual coverage is becoming increasingly important. However, performance lags for many languages, in

particular those with limited data. In such low-resource settings, cross-lingual transfer is crucial to closing the performance gap. Current interpretability studies suggest that cross-lingual transfer is supported by language-agnostic concept spaces, shared semantic representations that models use to process multilingual input. In this talk, I will present our work investigating the development of such shared concept spaces during the pretraining of multilingual language models.

Deniz Gündüz: **Learning to Compress Without Learning to Reconstruct**
The dominant paradigm in neural image compression trains deep networks end-to-end to minimise reconstruction error. This talk challenges that paradigm from three directions. First, we show that compression-ready representations already reside in untrained neural networks: by searching for subnetworks within a randomly initialised architecture, one can match the rate-distortion performance of fully trained overfitted codecs at a fraction of the decoding cost. Second, we demonstrate that decomposing the compression problem into modular, region-based implicit representations, each capturing a semantically coherent part of the image, yields a flexible, layered codec that surpasses state-of-the-art classical standards while enabling progressive and selective decoding. Third, we abandon pixel-level reconstruction altogether and compress images into the shared feature space of multimodal foundation models, preserving the ability to caption, classify, and reason about image content at rates orders of magnitude below conventional methods. The unifying insight is that effective compression is less about faithfully reproducing pixels and more about discovering compact neural structures that preserve the information that matters, whether structural, perceptual, or semantic.

Andrew Davison: A Robot Web for Many-Robot Localisation and Coordination
Safe and useful robots for complex environments must use their on-board sensors and computation to map, understand and localise within their surroundings, and we can envision a future where many such devices, with different functions and made by different companies, should operate in the same space. Is there a more modular way for this to work than all devices needing to use the same unified cloud-based "maps" system?
I will present and demonstrate our Robot Web proposal for distributed solutions to many robot localisation and planning based on per-device local computation and storage, and peer to peer communication between heterogenous devices via standardised open protocols. Our method uses Gaussian Belief Propagation-based distributed inference on full non-linear factor graph, and is highly robust and scalable while remaining simple and modular.

Dominik Muhle: **Self-Supervised (Semantic) Scene Completion**
Modern vision algorithms are already very capable in reconstructing scene geometry to facilitate better scene understanding. This talk presents a line of work where we tackle the task of (Semantic) Scene Completion. A task that unlike depth prediction also includes the reconstruction of occluded scene regions from images. In this line of work we tackle the problem in a self-supervised manner to

avoid costly 3D ground truth data. This data often requires extensive post processing in the form of dynamic objects handling and semantic annotations. This results in coarse, voxel-based ground truth data with a rigid class-based semantic structure. Self- or image-based supervision does not need to suffer from these drawbacks but proposes a unique challenge: we are missing a direct supervision signal for our models. In our works, we looked at the reconstruction from different perspectives: improving fidelity, reducing input data, and including modalities such as semantics.

Martin Binder: CASHing In on the Rashomon Effect: Extending Rashomon Sets Across Model Classes

The Rashomon effect describes the phenomenon where multiple models achieve similar predictive performance yet differ in their structure, predictions, and reliance on features. Rashomon sets formalize this by collecting all near-optimal models, but existing approaches are typically restricted to a single model class, even though the best class is rarely known in advance. This talk presents "CASHomon" sets, which extend Rashomon sets to the combined algorithm selection and hyperparameter optimization (CASH) setting, encompassing multiple model classes within a single set of near-optimal models. Using permutation feature importance, we show that feature importance values can vary substantially not only within a single model class but also across different model classes that achieve comparable performance. Our findings suggest that interpretations derived from a single-class Rashomon set may provide an incomplete picture: features deemed irrelevant by one model class can be highly important to another equally competitive class.