International Workshop

Shaping the Future of AI and Health

Inaugural Oxford-Québec-France workshop in the series "Shaping the Future of AI".

Date: 8-10 September 2024
Venue : Maison Française d'Oxford (Oxford, UK)
Partners for the event: Maison française d'Oxford, French Embassy in the UK, Délégation du Québec à Londres, IVADO, University of Oxford and University of Montreal
Number of invitees: 20-25 in total (from Canada, France and UK), with invited PhD students and post-docs
Conveners:
Prof. Catherine Régis (Université de Montréal, IVADO and MILA)
Prof. Lionel Tarassenko (University of Oxford, Reuben College)
Prof. Pascal Marty (Maison Française d'Oxford – CNRS-MEAE)

Context

This event was the inaugural Oxford-Québec-France workshop in the series "Shaping the Future of AI", building on comparative and interdisciplinary perspectives to explore key societal opportunities and challenges related to the development and deployment of AI.

The UK and France are the top two countries in Europe for AI research; across the Atlantic, Canada is second only to the US in this domain. In all three countries, there is a strong drive to apply the latest AI algorithms to the field of healthcare. The Maison française d'Oxford (MFO) has strong links with the University of Oxford and French research centres of excellence. It is also developing a partnership with Québec institutions.

It was therefore timely for it to organise an invitation-only workshop bringing together leading AI researchers from Oxford, Québec, and France to share the latest developments in the application of AI to healthcare.

The event was held at the MFO on 9th and 10th September 2024. As the MFO is a multi-disciplinary research centre, the workshop focused not only on algorithms and clinical applications, but also on key ethical and legal questions.

It addressed three main issues:

- **Fundamental Research at the Intersection of AI and Health**
- **Embedding AI in Clinical Settings**
- **Challenges and Opportunities for Responsible AI**

Workshop booklet

9 September: *First session* – Fundamental Research at the Intersection of AI and Health

Session chaired by **Prof. Lord (Lionel) Tarassenko** (Oxford University, President of Reuben College)

Summaries prepared by **Dr Cristian Roman, Senior Researcher** in the Biomedical Signal Processing & Machine Learning Group (Tarassenko group).

Dr Pierrick COUPÉ (CNRS and Univ of Bordeaux, CNRS unit LABRI, Bordeaux)
Title: Artificial Intelligence for Big Data Analysis of Brain MRI

Abstract: *In this talk, I will introduce advanced methods for brain segmentation and computer-aided diagnosis of brain pathologies. I will begin by unveiling groundbreaking advancements in the quantitative analysis of cerebral MRIs, showcasing a novel brain segmentation method powered by collective artificial intelligence. Next, I will demonstrate our user-friendly, open-access platform, which integrates sophisticated MRI analysis tools. This platform, used by 10,000 users worldwide, has automatically processed 500,000 MRIs since its launch. Finally, I will present our recent Big Data studies on lifespan modelling, highlighting new insights into various neurodegenerative diseases and demonstrating how lifespan modelling can enhance computer-aided diagnosis and prognosis.*

Summary:
The talk explored cutting-edge techniques that leverage large datasets and application of artificial intelligence for medical imaging. The speaker emphasised the 5Vs of machine learning, focusing the talk on velocity, volume, and veracity, representing crucial aspects in processing big data for brain MRI analysis. A key highlight was the use of **collective artificial intelligence (AI)**, where multiple intelligences collaborate to address complex and unknown problems, improving the generalisation and accuracy of diagnostic models.

The talk delved into whole-brain segmentation, a method that typically takes two weeks to complete by a human expert. The intra-rater accuracy is 72%, while inter-rater accuracy stands at 60%, with an out-of-domain accuracy of 37% for current single AI-based methods, indicating challenges in generalising the model to new datasets. To enhance the method's efficiency, collective AI employs 250 neural networks (nets)—125 for coarse decisions at 2 mm resolution and 125 for refinement at 1 mm resolution. This approach, inspired by the split observed in certain **bicameral political systems** (House of Commons – House of Lords), achieved 74% accuracy with a processing time of just over 10 minutes, showcasing its potential in lesion segmentation, deep grading, and aiding differential diagnoses in complex cases. Brain structure age estimation was also presented as an innovative method to assist with differential diagnoses, particularly in challenging cases.

Outlier detection methods were highlighted, illustrating how they help determine how far a case deviates from normal patterns. For instance, Alzheimer's disease detection, using

multiple lifespan models combined with the **HAVA score**, achieved 88% accuracy and a 73% prognostic value. This demonstrates the efficacy of AI in assessing the risk and progression of neurological conditions with relatively simple models—using only six parameters compared to the millions typically employed by deep learning models.

Scalability and accessibility were addressed through the system's recent upgrade to microservices infrastructure and Docker, which allowed it to process over 581,000 submissions from 10,391 users in around 3,500 institutions. This was achieved through collaboration with a team of dedicated computer scientists, which focus on the scalability aspect. The platform aims to be **user-friendly**, with a goal of enabling others to retrieve results in fewer than **five clicks**. This effort to streamline usability was part of a broader discussion on ensuring that AI technology is accessible to the wider scientific community.

In the discussion session, the difference between male and female brain volumes was explored, raising questions about normalisation methods. Additionally, the integration of blood biomarkers, PET scans, and MRI data was proposed as a means of improving diagnostic accuracy. The challenges of **data-sharing agreements** were also discussed, with federated learning identified as a potential solution for continuous learning from shared data across institutions, while maintaining privacy. This technology, however, remains strictly for research purposes and has not yet received CE or FDA approval for clinical use.

The talk concluded by emphasising the importance of interdisciplinary collaboration, particularly between neuroscientists and neurologists, to enhance the effectiveness of AI-driven diagnostics. The platform volbrain.net was provided as an example of how such technologies are being applied in real-world research environments.

Prof. Guillaume Lajoie (Université de Montréal /Mila)

Title: Foundation Models for Neuroscience: Uncovering Universal Neural Patterns for Many Uses

Abstract: *In this talk, I will present recent progress toward building AI foundation models for neuroscience. Much like large language models have done for text, neuro-foundation models aim to capture general and universal patterns of neural activity from different acquisition modalities, across brain regions, tasks, and even species. The learned representations can then be used to fine tune for varied tasks from diagnostic assistance, to machine learning tools for neuro-technology. I will discuss how such foundation models are poised to assist various areas of healthcare.*

Summary:

The talk outlined a roadmap for centralising the vast amount of neural data that has already been collected. The key message was the importance of **scale**, a lesson underscored by the success of large language models (LLMs). The talk emphasised that the brain's activity is dynamic and complex, requiring models that can process heterogenous, multimodal data. One example given was the **Flamingo 80B** model, which outperforms tasks completed individually when applied to different signals and patterns.

The speaker explained why these models work: pretraining on diverse neural recordings creates latent representations that capture the shared information across different brains. A model trained on **POYO**, animal data, serves as an example of this concept. Interestingly, the model does not require a predetermined time resolution; instead, it autonomously decides how to bin the brain's activity, removing the need for resolution as a hyperparameter. The talk highlighted the use of **Transformer architectures** for processing neural activity, where each channel is treated as a token, enabling the model to generalise to new tasks across approximately 100 regions of interest in the brain.

One key application is **Brain-Computer Interfaces (BCIs)**, where pretrained models are fine-tuned to achieve significant improvements. For example, patient images of writing letters show a task error rate of 16%, but after pretraining on monkey data, this is reduced to 4%. The transferability of the models was demonstrated through training on animal data and applying it to human subjects, showing the potential to decode and output human brain activity.

The talk also touched on the challenges associated with the **scale of data** needed for these foundation models, particularly in academic settings. Although the models are powerful, the sheer amount of compute required is substantial, and academia may not always have the resources to handle such demands. However, collaborations with commercial entities provide potential solutions to this problem.

Finally, the model's versatility was emphasised, as it is trained across different sessions and tasks, making it generalise well. Recordings typically last five seconds, and combining methods from multiple models could enhance future applications. The speaker concluded by stressing that while the groundwork has been laid, scaling up these models and partnerships will be essential to further progress.

Prof. David Clifton (Royal Academy of Engineering Chair of Clinical Machine Learning, University of Oxford)

Title: Generative Medical AI and the Role of the University

Abstract: *In a world in which the large AI multi-nationals dominate the creation of next-generation capabilities, what is the role of the University? This short presentation will make the case for the unique contributions to the field for which University research is particularly well-suited, acting as a complement to the massive-scale investment of industry in generic AIs.*

Summary:
The talk explored the evolving landscape of research outside medical imaging, focusing primarily on time-series data. The speaker humorously pointed out that data collection outside the realm of medical imaging is significantly easier. A key takeaway was the shift from simple approaches applied in time series data to more complex ones, such as those used in large language models (LLMs), highlighting that amidst the complexity, **a principled approach** to research remains essential. The current state of data handling for time series

was described as a "wild west," particularly in managing highly heterogeneous multimodal datasets.

The timescales of academic research were discussed, noting that collaborations with clinicians play a crucial role in shaping the understanding of datasets. As an example, through Prof. Clifton's collaborations, a recently acquired Monkeypox dataset from a low and middle-income country (Vietnam) was presented, which has already been patented, underscoring the **power of datasets** and their growing importance in research. However, challenges persist, such as the difficulty of anonymising free-text data, which is crucial for privacy and ethical considerations.

In clinical settings, the ability of AI to recommend drug dosages was discussed, tying into the development of more advanced interfaces, such as **Llama 3**. The talk addressed the issue of AI hallucinations—instances where the system provides false or misleading information. One proposed solution was to require the system to explain its reasoning by pointing to specific references or evidence. If it cannot provide a clear justification, the response may be a hallucination, thus creating **guardrails in medical AI**.

The speaker questioned whether academic research lab members are functioning as acquirers rather than innovators, particularly when it comes to building tools such as retrieval-augmented generation (RAG). The discussion raised an important point: even if researchers gain access to certain data or models, do they fully understand how to use them effectively? Another theme was whether merely **training AI models** can be as beneficial as having full access to proprietary datasets, especially when working with medium-sized language models.

The responsible use of AI tools was also emphasised. While these systems could be incredibly valuable as **discussion tools** for clinicians, they are unlikely to be classified as medical devices and thus won't face regulation from bodies like the FDA. However, the speaker suggested that other regulatory frameworks, like the **MHRA**, might step in to address what aspects could be regulated and how to design products that fit these gaps.

The global nature of AI was highlighted, particularly the challenge of different countries having varying medical guidelines. The speaker noted that in many cases, AI models do not need to be retrained for each country's guidelines. Training medical professionals in AI was also discussed, with programs such as the **MIT-Harvard PhD in Clinical AI** cited as an exemplary model for combining medical training with AI expertise. In addition, short-term fellowships for medical professionals were mentioned that aim to help clinicians better understand the growing field of clinical AI.

Summaries prepared by **Jonathan Carter, PhD student** in the Biomedical Signal Processing & Machine Learning Group (Tarassenko group).

Prof. Guillaume Dumas (Université de Montréal/Mila)
Title: Social Neuro-AI for Inter-Personalized Psychiatry

Abstract: *Precision medicine faces a significant challenge in mental health: addressing stakeholders needs by integrating cutting-edge technologies with existing scientific knowledge across biological, behavioral, and social scales. Scientific Machine Learning (SciML) tackles this challenge by incorporating domain-specific scientific knowledge into algorithms, combining hypothesis-driven and data-driven computational methods. Neuroinspired Artificial Intelligence (NeuroAI) further enhances this approach by embedding brain mechanisms into algorithms, providing a computational sandbox to better understand cognition in health and disease. We argue that human cognition emerges from the dynamic interplay between biological foundations and social contexts. Therefore, it is crucial to model developmental, interpersonal, and cultural processes alongside neurobiological mechanisms. We propose Social Neuro-AI as a novel paradigm that integrates these multidimensional factors, enabling the development of more inter-personalized, predictive, and ethically aligned AI systems for psychiatry. This approach has the potential to improve patient outcomes but also support the design of preventive and adaptive mental health systems.*

Summary:

In this talk, Professor Guillaume Dumas gave an overview of the field of computational psychiatry, discussed current challenges and introduced the paradigm of "Social Neuro-AI": using neural AI methods to integrate the social dimension into computational psychiatry, which is often overlooked.

Giving examples of how AI techniques have been used in this context, Prof. Dumas explained how his team have used mixed reality (MR) applications to frame cognitive tasks as games, enabling the tracking of cognitive development. Using techniques such as auto-encoders, this has enabled the analysis of learning trajectories and clinical phenotyping [1]. At the macro-level, his group have also used natural language processing (NLP) techniques to investigate popular and scientific discourse around autism through sources such as Twitter and PubMed [2].

Prof. Dumas additionally described recent advances in brain-based predictive models. For example, hybrid models which combine the best of data-driven and hypothesis-driven approaches [3]. He noted that, as recently seen in other problem domains, general purpose 'foundation models' could have a range of applications across computational psychiatry.

In addition to integrating the social dimension into computation psychiatry, another key challenge discussed was the importance of integrating AI skills into medical training [4]. To address this, Prof. Dumas' home institution, Université de Montréal has launched the first French-speaking course in computational medicine, aimed at both clinicians and computer scientists, on the fundamentals of AI in medicine.

Finally, an important point raised in the discussion was the importance of co-production and patient engagement in clinical research, especially in data collection methodologies. Prof. Dumas provided an example of this from a study involving the collection of continuous electroencephalogram (EEG) data from patients. To help alleviate any patient concerns about the equipment, a designated member of staff was always 'on-call' to answer any questions.

[1] M. Hafsia, R. Trachel, and G. Dumas, 'Towards Clinical Phenotyping at Scale with Serious Games in Mixed Reality', presented at the Empowering Communities: A Participatory Approach to AI for Mental Health, Oct. 2022.

[2] C. Gauld, J. Maquet, J.-A. Micoulaud-Franchi, and G. Dumas, 'Popular and Scientific Discourse on Autism: Representational Cross-Cultural Analysis of Epistemic Communities to Inform Policy and Practice', Journal of Medical Internet Research, vol. 24, no. 6, p. e32912, Jun. 2022, doi: 10.2196/32912.

[3] G. Abrevaya et al., 'Learning Brain Dynamics With Coupled Low-Dimensional Nonlinear Oscillators and Deep Recurrent Networks', Neural Computation, vol. 33, no. 8, pp. 2087–2127, Jul. 2021, doi: 10.1162/neco_a_01401.

[4] C. Gauld, J-A. Micoulaud-Franchi, G. Dumas, 'Comment on Starke et al.:'Computing schizophrenia: ethical challenges for machine learning in psychiatry': from machine learning to student learning: pedagogical challenges for psychiatry', Psychological medicine, vol. 51, no. 14, p. 2509-11, Oct. 2021, doi: 10.1017/S0033291720003906.

Prof. Alison Noble (Technikos Professor of Biomedical Engineering, University of Oxford)
Title: Leveraging AI to Simplify Ultrasound

Abstract: *In this talk I will discuss how the emergence of machine learning-based image analysis is changing medical ultrasound, using examples from my group's research on fetal ultrasound video and multi-modal imaging to highlight recent technical advances and clinical translational opportunities. I will also highlight how research is no longer just about the algorithm but depends on effective collaboration of inter-disciplinary research teams to progress the understanding of AI technology acceptance and adoption in existing and novel clinical settings.*

Summary:
In this talk, Professor Alison Noble gave a tour of the field of automated sonography, a field she first started working in after a chance conversation sitting next to a colleague working in obstetrics. Since that initial conversation over 25 years ago, Prof. Noble's research has addressed a range of important problems in the field of medical imaging and led her to co-found Intelligent Ultrasound, one of the first companies successfully to create first-in-market clinical ultrasound AI products.

One of the key enablers of her field of research has been the significantly increased availability of data. Indeed, as seen in many other fields adopting deep learning-based methods, the focus has become "less the AI, more the data". Prof. Noble highlighted that some of her early work using classical computer vision techniques used only a small number of image sequences collected from two patients [1]. In contrast, more recent work from her group has used over 200 hours of multi-modal data including ultrasound, audio and video, capturing the entire sonography process [2]. Beyond diagnostic tasks which only use the (uni-modal) ultrasound data, the collection of multi-modal data could lead to new applications in

automated sonography, such as video summarization [3], and AI-assisted probe navigation [4].

Prof. Noble's talk sparked broader, important discussions on automation and the human-AI interface. The use of assistive AI tools for the clinician poses risks such as the de-skilling of future generations. Boeing's now infamous MCAS system was highlighted as a particularly severe example of how automation-induced changes in human behaviour can have unintended consequences. Additionally, AI may help to automate away easy tasks, but this may not necessarily ease the burden on the clinician. There is a risk that they may leave the clinician with only the difficult tasks to solve in a more compressed timeframe, increasing cognitive burden.

In the light of these challenges, Prof. Noble highlighted the need to build up case studies and design best practices. After 25 years, her field continues to be an exciting, interdisciplinary example of research at the human-AI interface, with many open problems left to solve.

[1] G. Jacob, J. A. Noble, and A. Blake, 'Robust contour tracking in echocardiographic sequences', in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, Jan. 1998, pp. 408–413. doi: 10.1109/ICCV.1998.710751.

[2] H. Zhao *et al.*, 'Towards Unsupervised Ultrasound Video Clinical Quality Assessment with Multi-modality Data', in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds., Cham: Springer Nature Switzerland, 2022, pp. 228–237. doi: 10.1007/978-3-031-16440-8_22.

[3] X. Guo, Q. Men, and J. A. Noble, 'MMSummary: Multimodal Summary Generation for Fetal Ultrasound Video', Aug. 07, 2024, *arXiv*: arXiv:2408.03761. doi: 10.48550/arXiv.2408.03761.

[4] J. Karim, J. Lander, Q. Men, R. Ahuja, A. Noble, and A. Papageorghiou, 'EP01.34: Towards AI-enabled navigation of the second trimester anomaly scan', *Ultrasound in Obstetrics & Gynecology*, vol. 64, no. S1, pp. 116–116, 2024, doi: 10.1002/uog.28048.

Dr. Laurence Watier (Inserm - CESP, U1018); Institut Pasteur, Paris, and University of Versailles-Saint Quentin

Title: *Healthcare Data in France and The BactHub Project*
Abstract: *Large, high-quality databases are key to the development of AI. After a description of the French national healthcare database (SNDS), its pitfalls and the administrative procedures for accessing it, an illustration focusing on the study of care pathways using state sequence analysis will be presented. The final part of the talk will illustrate how clinical data warehouses can be used to enrich SNDS database, through the BactHub project "Impact of individual exposure to antibiotics on bacteremia caused by resistant bacteria".*

Summary:
In this talk, Dr Laurence Watier discussed the development of the French National Healthcare Database (SNDS) and its value for healthcare research. Since the initial

development of the *Programme National de Médicalisation des Systèmes d'Information* (PMSI) in 1982, the scope and scale of healthcare data available from France for academic research has steadily increased.

Following the creation of the SNDS in 2016, it is now possible to get a complete overview of care pathways for the entire French population, with linked data on hospital stays, health insurance reimbursement, and causes of death. After an extension of its scope in 2019, this care pathway data now spans a period of up to 20 years. Covering over 67 million people, this makes the SNDS the largest and most comprehensive healthcare data resource available in Europe.

Dr Watier highlighted the value of linking and combining data from multiple sources and modalities. For example, combining data from clinical data warehouses and SNDS databases will enable her team to study the role of antibiotic consumption on microbial resistance through the ongoing BactHub research project [1].

Dr Watier's talk gave rise to a particularly lively discussion on the ongoing issues around data access in the three respective nations. The suggestion that France might lag behind in this regard was met with resistance, with attendees from the UK and Canada quick to highlight pitfalls in their own healthcare systems. What was clear was that many attendees felt that bureaucracy (e.g. acquiring signatures) and the poor integration of different technological systems (e.g. electronic health record platforms) were continuing barriers to data access and research across healthcare systems.

Finally, a subtle but important point raised by Dr Watier was the need for extra care when performing analysis on evolving databases, where recommended treatments, care pathways and even basic definitions such as the 'primary diagnosis' can change over time. This problem, known as distributional shift in the AI community, often harms the generalisability of models deployed into real-world healthcare systems. Careful interpretation of the data and the integration of domain knowledge, as Dr Watier's team has done, remain some of the most simple but effective ways of addressing this.

[1] S. Abbara *et al.*, 'Antimicrobial Resistance and Mortality in Hospitalized Patients with Bacteremia in the Greater Paris Area from 2016 to 2019', *Clinical Epidemiology*, vol. 14, pp. 1547–1560, Dec. 2022, doi: 10.2147/CLEP.S385555.

***Second session*: Embedding AI in Clinical Settings**
Session chaired by **Dr Katia Wehbe** (French Embassy, Science Attaché for life sciences, medical sciences and health)

Summaries prepared by **Dr Shaun Davidson, Research Assistant** in the Biomedical Signal Processing & Machine Learning Group (Tarassenko group).

Prof. Charis Antoniades (Professor of Cardiovascular Medicine, University of Oxford, and Consultant Cardiologist, Oxford University Hospitals)

Title: *Using AI to interrogate CT images: Predicting future cardiovascular events*

Abstract: *AI is being is used for interpretation of multi-dimensional clinical/patient data and has many applications in medical imaging: for improving the workflow in image processing (automated segmentations, etc), improving accuracy in image interpretation (see things we cannot see), helping with prioritization (fast exclusion of "normal" to allow focus on "abnormal"), fast quantification of structures (e.g. automated calcium score, coronary plaque analysis, etc) or even fast diagnosis and future risk prediction (guided therapies). Technologies that use coronary CT angiograms to quantify the degree of coronary inflammation (e.g. fat attenuation index score), provide important input in AI-driven prognostic models together with atherosclerotic plaque burden and clinical risk factor data, leading to extremely powerful tools for cardiovascular risk prediction and management of patient care. The large ORFAN international programme aims for 250k cardiac CT scans linked with lifelong outcomes data and is being used to train and validate novel imaging biomarkers for early diagnosis and future prediction of a wide range of cardiometabolic diseases.*

Summary:
AI and image recognition can be used to identify disease patterns in medical data – similar to facial recognition etc. For example, a model can be trained to detect the risk of coronary artery disease from images of faces enabling potential passive screening (from a Chinese study).

However, AI is not always the solution – sometimes regression models, etc are better suited. Another major challenge is interpretability of these models – especially with regards to clinical trust. There are also issues with bias, and regulatory approval (especially if the model weights are not fixed).

AI has numerous potential applications in radiology – interpretation of data, segmentation, prioritisation (focus on 'abnormal' images), quantification (calcium score, etc). These can be used to make future predictions or guide therapy.

Heart attacks stem from narrowed arteries (which cause chest pain) with plaques which break, causing a heart attack. The pain provides an early warning, and corrective treatment (e.g., stents) can be applied to good effect. However, detection requires labour intensive manual analysis of images. AI is being used for automated image screening, for example detecting different types of plaques that have different risk profiles for heart attacks.

Another example is an AI model can be used for automated segmentation of the pericardium (membrane around the outside of the heart). These models have achieved human expert-level agreement and operate far faster than humans can. This enables larger scale screening, and in turn better predictive models.

The number of heart attacks is increasing, suggesting that improvements can be made in treatment pathways. In particular, predicting the growth of plaques would enable preventative care. Plaque growth is tied to inflammation of the arteries, and heart attacks can

occur when plaques are still small and unlikely to cause pain, and thus unlikely to be detected. Notably, the fat surrounding the artery sees an increase in water content when an artery is inflamed. In the longer term, this leads to fibrosis, and these changes could provide an early indicator of artery inflammation that may lead to a heart attack. However, current imaging strategies cannot detect this.

Fat Attenuation Indexing has been developed as a means of detecting this. It is possible to use minor variations in the texture of the fat surrounding the artery to detect changes in the fat layer that are indicative of inflammation. But how can we label this data and train a model?

This was achieved by taking biopsies from 1500 patients scheduled for cardiac surgery and gathering extensive data from the samples. CT images were then taken of the same patients, giving a ground truth (from biopsy) linked to CT images. This allows for training of "radio-transcriptomic biomarkers" i.e., biomarkers typically only available from biopsy that can be capture from CT images using machine learning.

Validation was performed on a prospective study of 4000 patients with a 10-year follow-up, including data from Germany and the US. Hazard ratios using the Fat Attenuation Index were between 5 and 9, showing these radio-transcriptomic biomarkers can be used to predict patient outcomes.

Further validation on a large international dataset (target 250,000 individuals, current 135,000 gathered so far) with 10 years follow up is ongoing. During follow up, there are several thousand examples of various adverse cardiac events, meaning the data set can be used to validate the models across a large, diverse cohort.

In the UK, there is an extensive process of cross-checking and linking a variety of different datasets resulting in a composite NHS digital dataset – ORFAN. Coronary inflammation (FAI) on this cohort shows a good stratification of risk, with the upper quartile having a hazard ratio of 20 compared to the lowest quartile.

Work is ongoing on combining EHR data and biomarkers derived from CT images to build a single coherent model that can be used to predict risk of adverse cardiovascular events. On ORFAN, this model has an AUROC of 0.844 of predicting coronary artery disease. This model predicts adverse events well over a 10-year follow-up period.

Importantly, there are interventions available to help those with elevated risk, such as statin treatments, improving patient outcomes. The prediction model has been adopted in 5 NHS trusts and has potential health economic benefits in the UK. In summary, we have the tools to treat CV risk, we can measure coronary inflammation using images, and these approaches can be implemented and have cost benefits.

Prof. Raphaël Porcher (Professor of Biostatistics at Paris Cité University - PR[AI]RIE (PaRis Artificial Intelligence Research Institute))

Title: *Potential of artificial intelligence for healthcare: from digital therapeutics to the development of new drugs*

Abstract: *Artificial intelligence (AI) has many potential applications for healthcare. For instance, devices incorporating AI algorithms have shown comparable performance as professionals to interpret pathology slides, dermoscopic images, MRIs or CT scans. Some of those algorithms are already incorporated in commercial imaging softwares. Beside those prominent examples, AI can also have important applications for digital therapeutics, and to help drug development, from very early phase to clinical trials. In this talk, we will review the specific issues raised by digital therapeutics and their evaluation, as well as how AI can enhance drug development and the related research perspectives.*

Summary:
AI in health is an extremely wide topic, with a variety of potential applications. The applications can include decision support tools, software as medical devices, and AI for drug development. Each of these raise particular questions with regards to evaluation.

Software as medical devices cover a variety of potential implementations. These can include helping prescription/delivery of drugs, telemedicine, or screening, diagnosis, and medical decision making. Importantly, with an aging population and increase in chronic illnesses, without these tools society will not be able to support the management of patients in 20 years' time. 42% of adults have a chronic illness, and 25% multiple chronic illnesses.

Connected devices could enable remote monitoring of patients in real time, passively and at ahigh frequency. This can provide the ability to refine diagnoses and intervene at the 'right' time. A major example of this is implantable devices for closed-loop insulin delivery at home.

However, digital medicine does raise new challenges; it adds a new burden of treatment (patients must undergo additional monitoring, healthcare professional may need to respond to additional alerts) and must take account of multimorbidity. Patients may also feel that digital tools are intrusive.

A systematic review looked at 38 devices aimed at delivering "just in time intervention" (JITI). Most of these were focused on monitoring for cardiovascular diseases or diabetes. Only 26.3% of products provided a full summary of how the device operated to deliver a JITI. 13.2% of product summaries provided the results of clinical studies. None of the product summaries reported any information on data ownership.

The regulatory evaluation of these digital therapeutics needs to be re-thought. The current system is based on what has been developed for drugs, with few new compounds each year, and these do not change once developed and approved. For digital therapeutics, this is very different. 1000s of devices are developed every year, and software updates can mean the device needs to undergo approval again.

AI can also be applied in the area of drug development. This can enable you to do *in silico* "virtual" trials. These can be performed using mechanistic models, which use physiological

knowledge, or phenomenological models, which are based on prior observations. However, there are limited guidelines for this type of virtual *in silico* trial, making it unclear how these trials can be made part of the regulatory process.

Even when clinical trials are undertaken, AI can also be used to enhance those, for example as part of the recruitment process or creating virtual 'control' arms. For example, in an Alzheimer's trial, AI was used to select and recruit patients who were at a 'high risk' of progression of Alzheimer's for their clinical trial, so that drug effect on trajectory could be more clearly observed. Another possible application is synthetic data for a control arm – for example from the Medidata Synthetic Control Arm (SCA) which is based on data from 22 000 clinical trials. A further potential use is to use generative AI (e.g., GANs or VAEs) to generate an entirely virtual cohort of patients, for instance to be used as controls or to enrich a control group in a trial.

Prof. Nadia Lahrichi (Polytechnique Montréal)
Title: AI Tools to Optimize Planning in the Operating Room

Abstract: *In this talk, I will introduce a novel approach to scheduling specialties to operating rooms, a problem commonly known as master surgical planning. To optimize resources utilization, we integrate the patient case mix, which involves selecting patients from the wait list, assigning them to the operating list, and scheduling their surgery dates. Each patient has unique requirements, such as ICU beds, surgical beds and surgery time. Our findings demonstrate that, contrary to existing literature, integrating these elements leads to superior outcomes from a systems perspective. We achieve optimal use of operating rooms and beds while improving surgery wait times. Our methodology uses a clustering algorithm to group patients for each specialty and a mathematical programming approach to solve the optimization model. The results we present are based on data from a partner hospital involved in the project.*

Summary:
Operations research aims to develop tools to improve decision making. About 40% of hospital budgets are spent in the operating room, so improving decision making here can save significant costs. Major potential areas of improvement are start-time tardiness, case cancellation, time between subsequent operations, etc. The focus of this research is on case cancellation rate.

There are multiple interacting systems. At the top level there is a master surgery schedule (built for 28 days assigning days or blocks of days to specialities. Then, there are surgeons selected patients from a wait list to be on an operating list for a given surgeon. At the next level there is the sequencing of cases in the operating room and the scheduling of personnel. These operate on a variety of different time scales.

This research encompasses 11 – 16 specialities, 75 surgeons and 2 different hospitals. It encompasses 10 – 15,000 operations per year. In this dataset, the wait times for operations have a significant right tail which suggests the 'first in first out' system is not operating properly, with wait times of up to more than two years. In general, the master surgical

schedule (set for 28 days) is not aligned with wait times or wait lists, which is part of what drives the issue.

How can we reduce wait times while avoiding cancelations? This can be achieved by integrating multiple decisions and characteristics to optimise bed usage. To do this, it is important to predict whether patients will need hospital beds after surgery, as well as how long they are likely to require this bed for (i.e., length of stay).

Another important element to optimise is the use of the operating room itself by accurately predicting how much time the operating room will need to be used for to perform a particular procedure. To help with this, we can investigate whether there are clusters of association between time spent in the operating room and hospital length of stay for different specialties. When this is investigated, up to 11 clusters of patients per specialty are observed.

We can now attempt to build a model to build a master surgical schedule to maximise operating room utilisation and minimise wait times. The model also incorporates a custom weighting factor that can be used to account for different priorities or relative importance of operations by different specialities. This is a decision support tool that is designed to provide a variety of different solutions that the user can select from, and so was applied to a variety of different scenarios. Incorporating predicted length-of-stay into creating the master surgical schedule saw the same number of operations occurring but notable improvements in wait time (68%) and utilization of the operating room (36%).

A second case study focused on maximising room occupancy by prioritising the oldest patients on the wait list. Comparing specialties, 2% of paediatric patients need an ICU but 91.4% of cardiovascular patients require this, so there is significant heterogeneity between specialities. Interestingly, there are also notable seasonal variations in the cancellation rate of operations, though the dataset doesn't give causes for these. Additionally, 75% of capacity in the ICU is occupied by emergency patients, and there is a reasonably consistent discharge rate of 2 patients per day.

Given this information, we can use a Markov Chain to predict availability of beds in the ICU. This model ingests wait list, availability, and can produce a 28-day master schedule. This yields similar performance to the earlier model incorporating length-of-stay into decision making. This model can, again, be varied based on priorities. For example, you can tell the model you don't want to ever cancel an operation due to emergency admissions (e.g., in a paediatric context). Overall, these results show that prioritising patients using this type of modelling strategy can improve resource utilisation and reduce wait times.

Second session continued and wrap-up of Day One

Summaries prepared by **Dr Peter Bannister, Managing Director** of Romilly Life Sciences Ltd, and former member of Tarassenko group (Papadakis and Pomey) and by **Bernardo Lustrini**, **PhD student** in the Biomedical Signal Processing & Machine Learning Group (Tarassenko group), for Guyet talk.

Dr Michalis Papadakis (CEO and co-founder of Brainomix)
Title: Transforming stroke treatment with AI-powered medical imaging

Abstract: *Originating as a spin-out from the University of Oxford, Brainomix specializes in the creation of AI-powered software technologies to enable precision medicine, delivering best-in-class solutions to drive better treatment decisions for conditions where imaging defines the disease profile, including stroke and lung fibrosis.*

Up to 80% of eligible stroke patients remain untreated. Brainomix has developed the Brainomix 360 Stroke platform to ensure more patients receive the right treatment in a timely manner. The Brainomix 360 Stroke platform seamlessly integrates withing the hospital system and uses AI to automatically analyze brain CT scans of stroke patients in real time, helping front-line physicians with diagnosis and treatment decisions.

Brainomix is the market leader in stroke AI in Europe, with widespread clinical adoption in more than 30 countries worldwide and nationwide installations in Hungary, Wales and Poland. More than 60 publications have validated the performance and clinical value of the Brainomix 360 Stroke platform, including studies that have shown stroke networks using the technology treat 50 more patients with mechanical thrombectomy. This growing body of studies has led the National Institute for Health and Care Excellence (NICE) to endorse the clinical use of Brainomix 360 Stroke in NHS England.

Summary:
Reflecting it was 1956 when the first AI workshop took place in Dartmouth; 250 AI companies at RSNA in 2023. We can see real value of AI in radiology.

Solutions range in (increasing) complexity

- Automated Detection
- Quantitative Imaging
- Decision Support - this is where we are now
- Automated Diagnosis

AI in radiology today means Augmented Intelligence. Brainomix focuses on diseases where imaging defines disease profile - stroke is one of these. A stroke ages the brain by 3.6 years every hour and 80% of patients are missing out on treatment.

Acute ischaemic stroke treatments: either thrombolytic drug to dissolve clot or fitting a stent retriever (mechanical thrombectomy). Choice of treatment relies on a non-contrast CT scan on admission with a treatment window of a few hours - very few visible features in this scan.

Brainomix was founded to deliver a vision of enabling the best treatment decisions in all hospitals. Results to date:

- 2m patient scans in clinical care
- Increase in patients with functional recovery from 16% to 48%

- 55% treatment uplift in NHS hospitals - largest prospective evaluation of stroke imaging AI
- Deployment across 300 European sites through 2024.

Most strokes occur in community; patients will receive a scan in a community hospital, and this will determine whether they will receive treatment. However, the clinicians are not stroke experts, so Brainomix enables the correct patients to be transferred to larger hospitals with stroke specialists. Workflow is completely automated and deployed in both hub (specialist) and spoke (community) hospitals. Results are provided within a few minutes.

Products are CE-marked and FDA-approved. Integrates own platform with PACS, multi-vendor scanner environment.

1. e-ASPECTS module analyses non-contrast CT scans to measure degree of ischemia. Heatmap overlay explains to clinician how the algorithm derived the ASPECTS score. From a regulatory aspect, its intended use is to support the decision making of a physician alongside other diagnostic information.
2. e-CTA uses contrast CT and can measure degree of collateral symptoms
3. e-MRI allows analysis of DWI and DSC (functional MR)
4. e-CTP provides analysis of CT perfusion
5. Triage LVO and ICH provide real-time alerts to clinicians when a large vessel occlusion or a bleed is suspected.

User design makes or breaks adoption – it needs to be intuitive. Superior clinical performance alone is not enough when there are incumbent solutions.

Comprehensive clinical validation evidence has been generated, comparing independent gold standard clinical review with automated analysis.

NHSX AI Award deployment over 26 sites with 83,000 patients. Able to direct over 50% more patients for treatment (mechanical thrombectomy).

Case Study at Royal Berkshire Hospital in Reading (referrals to Oxford for treatment):
- >1 hour time saves in assessment -> referral
- 3x increase in functional independence

NICE approval obtained: 8 years from initial launch in 2016.

In France, primary imaging modality is MRI which has meant Brainomix has had to follow other markets that use CT.

Platform approach will enable solutions for other conditions where imaging is central to treatment decision making, where there is an opportunity to positively impact patient treatment:

- Pulmonary lung fibrosis
- Oncology

Longer-term vision to enable AI driven personalized medicine. It is still not established how you scale up a revenue model for imaging AI; currently based around (multi-year) annual software licences. Strong clinical evidence has followed commercial market deployment, so not yet capturing the full value being delivered.

Prof. Marie-Pascale Pomey (Université de Montréal)
Title: How to Integrate Patient Partners in AI Development and Patients in the Clinic When AI is Mobilized?

Abstract: *In recent years, numerous patient partner engagement initiatives have been carried out in Quebec in support of various digital health initiatives. This presentation will recount different projects and present assessments of the added value of co-constructing with patient partners in the fields of AI and digital health.*

One of the areas that has been most investigated is the importance of involving patients at the clinical level when AI is used in decision-making. Indeed, in practice, it is necessary on the one hand to develop clinical algorithms with patient partners and on the other hand to establish a discussion with patients so that they understand how AI is mobilized to help them make informed choices that meet their values.

Summary:
Experiential knowledge:

- All know-how acquired from lived experience - self-care, interaction with care providers, utilization of health and social care services
- In living with a condition, the person finds an opportunity to develop new skills, move toward self-transformation through the experience and move the lines of knowledge sharing
- For people with a chronic illness or disability, this is a lifelong experience.

Patient Knowledge:

- Embodied - sensory knowledge
- Monitoring - self-management via devices and/ or self-awareness
- Relational - role of caregivers
- Cultural - how to influence preferences, needs and care relationships
- Navigation re. the healthcare system
- Medical - physiological, pharmacological

From centrism to partnership: a difficult transition
Paternalistic → patient centred-case → (?) partnership with the patient

In the partnership model, HCPs are experts in disease, but patients are experts in living with an illness and in utilization of the healthcare system.

What makes involving patients in the development and use of AI medicine important?

1. Improved clinical relevance
2. Improved personalization and patient partnership
3. Building trust and acceptance
4. Ethical considerations and fairness
5. Quality and safety improvements
6. Strengthening patient-provider relationships
7. Promotion of innovation

How to involve patients? Framework published in 2014: considers level of participation and factors affecting this.

1. Patient education and awareness:
    - Clear communication using simple language and avoiding jargon
    - Educational resources - accessible material that provide balanced view
    - Workshops and seminars to allow concerns to be raised and questions asked

2. Transparent Communication
    - Explain how Ai is being used in care today - how they influence decisions and what data is being used
    - Open dialogue - encourage patients asking questions/ voicing concerns
    - Share results - allow patients to access results and understand how they were generated

3. Training clinicians
    - How to use AI
    - Patient communications
    - Sensitivity to patient concerns

(In 2021, 57% (294) of family physicians in Quebec said they were interested in learning about AI and 34% didn't know. In France, an online information campaign was recently launched, including 4 short videos focused on the ethics of using digital tools in health.)

4. Consultation and involvement
    - Patient involvement in consultation
    - Advisory panels

(The Digital Health Committee of the Centre of Excellence for Patient & Public Partnership focuses on digital health issues; created just before Covid, it was very influential in responding to ministry questions during the pandemic to support decision making.)

5. Co-construction
    - Co-design workshops

- Pilot trials with patients

6. Personalised use of AI
   - Tailored AI solutions
   - Monitor and adjust.

(Case study of missing information resulting from a scoping review of the Covid contact tracing app. Value added by co-designing the questionnaire included better integration of needs of vulnerable population.

Case study of Transplant Action Connected: telehealth-delivered program and accompanying patients to enhance the clinical condition of patients throughout a liver transplant. Use of connected devices, incorporating accompanying patients into the care team. Use time waiting for diagnostic tests (average 232 days on waiting list) to better prepare for the transplant procedure itself. Similarly for post-transplant recovery.)

7. Citizen and community service
   - Peer product review
   - Personal science
   - Interdisciplinary

(Case study of artificial pancreas systems - experiences of diabetic who share open-source, unregulated "recipes" for automating insulin delivery using digital devices. Tidepool Loop received FDA clearance, originating as a patient-led initiative.)

8. Ethical Issues
   - Addressing bias and fairness
   - Data Privacy

In Quebec, there is someone in all healthcare organizations whose job is to recruit patients for participation in development projects. It can be difficult to rely on long-term contribution from younger and/ or working adult patients; scheduling meeting outside of normal working hours for HCPs can help.

Dr Thomas GUYET (INRIA, Lyon)

Title: Optimizing Care Pathways using AI tools

Abstract: *One objective of health regulatory agencies is to continuously improve care and its organization. France is engaged in transforming its health system into a pathway-based system, where care is organized according to treatment guidelines as care pathways. This raises the challenge of designing such care pathways. Medical data warehouses contain information about care delivered to patients, providing a longitudinal view of patients' care trajectories. The development of artificial intelligence techniques, particularly machine learning (ML) and natural language processing (NLP), offers an opportunity to leverage these massive datasets to propose evidence-based guidelines. In this presentation, we will introduce a methodology based on AI tools to support clinicians in describing, analyzing, and*

optimizing care pathways. We will then discuss some limitations related to this methodology. An ongoing study on lung cancer surgery conducted at the Greater Paris University Hospitals will illustrate our methodology.

Summary:
Care pathways are management tools, made to help different care stakeholders take decisions and organise care, hospitals and the healthcare system. The objective is not onlyt to improve the quality of the care, but also to optimise physical and human resource usage, and standardise processes.

Since the Cordier report (2013), the notion of "pathway medicine" has been promoted in France for the organisation of the healthcare system. Health authorities adopted this for the management of chronic diseases. For other cares, Hospitals/health professionals are currently funded through fee-for-service (T2A in French), but some people want to change this to funding based on the handling of the entire patient care pathway. To achieve this, we first need to design care pathways - these do not always exist and a methodology to design them is still missing.

Case Study: Bronchopulmonary Cancers (BC). 50k cases/year in France; only treatment is surgery. The time to treatment is very important for prognosis. The care pathway according to experts is a list of 5 steps. This is very simplistic and does not address the main question of time-to-surgery.

We need to provide a more specific pathway. To design pathways, we need to do epidemiological studies. Either we acquire new data (expensive and long) or re-use readily available data (preferable) such as the SNDS dataset (SNDS is a database of care reimbursements) or hospital data warehouses.

We use AI to take existing data and create care pathways. Importantly, the observation data that is readily available has been collected in real-world situations and we cannot control the measured variables. This is often a limitation. The data has not necessarily been collected for epidemiological surveys but for administrative purposes. For instance, we have temporal drug administration data but not drug exposure data.

How do we exploit these databases? Two main steps: first, transform real data into a set of care trajectories; use data blending to sum up row data into care trajectories. Then, do analysis to extract typical observed care pathways.

There are two main issues in these processes. First, the semantic gap problem: difference in raw administrative data and medical data. We have to convert from administrative semantics to medical semantics. This is not easy. We use automatic reasoning and NLP for this. Second, unveiling "typical" care pathways from massive trajectories is hard. This is a data analysis task, and we do this with classical machine learning.

Data Processing: How to process raw databases? Typical database schemas and SQL queries for SNDS databases are very complicated. We want to transform a complex dataset into a much simpler one where we have a collection of trajectories. Typical approach is to make a Data Management Book (DMB): write down everything you are going to do to the raw data and create large SQL queries to transform the dataset. This is a very long process.

It is often hard to find when specific patients had specific interventions, as we have to dig into the textual data. Here we use NLP tools, which are very useful, for example Named Entity Recognition. Challenges arise from semantic gaps and the temporal dimension.

We also use semantic web tools (automated reasoning). We do not hear a lot about this today, but it is very important in healthcare. We have a lot of structured data and nearly 30 years of research on different ontologies of medicine. We can use these techniques in machine learning, and we should promote their usage.

Extracting Pathways: We then want to figure out from individual patient pathways if there are typical pathways. We do some clustering to extract typical care pathways; previous work with AIRacles to develop techniques on extracting care pathways from longitudinal trajectories. We developed two main approaches, clustering and tensor decomposition. (Incidentally, you can have a patient undergoing multiple pathways at the same time.)

Basic pathway modelling is done by clustering to identify groups of similar trajectories. Once we have a group, we identify a representative within a group. This is not as simple as K-means as the dataset is too complicated; for instance, not all patients have the same number of events. We need to adapt the notion of similarity, by using for instance Dynamic Time Warping. This is semantically interesting but technically very complex. There is no longer a distance metric, so you cannot apply standard algorithms, e.g. you need to develop new techniques to find representative samples from groups. We do not believe that there is a unique similarity metric for all problems. We are creating a Python library for clinicians to develop their own similarity metrics between care trajectories.

AI solutions are interesting, especially for bridging the semantic gap and constructing care trajectories. Focusing on 3 viewpoints:

- patient view (life pathway)
- regulator view (care pathway)
- healthcare provider view

We identify the notion of view. These views represent different ways of viewing a sequence of caretaking events.

It is important, when optimising a care pathway, to know for whom you are optimising the pathway. Often, we optimise the regulator view. e.g. medico-economic objective, but this is not necessarily always what you want. Maybe we need to take all three views into balance when optimising.

Regarding the future of AI and Health, we can observe that AI is transforming epidemiology. We see more and more processing pipelines, and Python libraries that aid with designing epidemiological studies. This can increase reproducibility and automate the process. It is important to understand that the design and implementation of these pipelines will ultimately be dependent on the design and implementation of these platforms. Therefore, the design of the platforms can have long-term impacts on the field of epidemiology.

Perhaps "care pathways" are already old-fashioned, too static and guidelines driven. In practice, AI may be used to make personal decisions for each individual patient. There are new ML techniques, mostly RL, that try to provide personalised and situated decision support along the course of patient trajectory. Some fundamental questions at the crossroad of AI, health and social science remain open: Do we really need interpretable care pathways, or do we need an individual recommendation system? Is a well-founded automated recommendation system enough?

***Third session (Day 2)*: Challenges and Opportunities for Responsible AI**
Session chaired by Prof. Catherine Régis (Université de Montréal and MILA)

Summaries prepared by **Revd Dr Lyndon Drake,** Research Fellow in Theological Ethics and Artificial Intelligence (Faculty of Theology, University of Oxford)

Prof. Pierre-Luc Déziel (Université Laval)
Title: Designing Responsible AI Tools for Medical Imaging: Introducing PACS AI

Abstract: *This presentation will use PACS AI as a case study to assess the challenges and opportunities of developing responsible AI for medical imaging. PACS AI is a custom platform developed by a team of physicians, data scientists, and legal experts at the Canadian Institute for Advanced Research (CIFAR) to deploy AI models on images stored in a hospital's Picture Archiving and Communication System (PACS). Specifically, the presentation will focus on the privacy impact assessment conducted in recent months to discuss key components and design features of the platform related to the use and communication of personal health information. We will identify solutions and best practices that will be implemented to ensure PACS AI respects patients' privacy expectations and rights.*

This presentation is ***partially*** *based on the following article: Theriault-Lauzier P, Cobin D, Tastet O, Langlais EL, Taji B, Kang G, Chong A-Y, So D, Tang A, Gichoya JW, Chandar S, Déziel P-L, Hussin JG, Kadoury S, Avram R, A responsible framework for applying artificial intelligence on medical images and signals at the point-of-care: the PACS-AI platform, Canadian Journal of Cardiology (2024), doi: https://doi.org/10.1016/ j.cjca.2024.05.025.*

Summary:
Developed initially for cardiology, PACS AI is a decision-support tool that enables medical practitioners to use AI for analysing images stored within clinical systems. The central theme

of the presentation was privacy, emphasising that legal and ethical considerations are as important as the technical aspects of AI deployment.

The implementation of Privacy Impact Assessments (PIAs) has been mandatory for Canadian researchers since July 2024. This new requirement has prompted a first round of drafting within the team. PIAs should not be static documents but require regular updates, and their development must involve both developers and decision-makers. Although PIAs are often seen as administrative tasks, they provide opportunities for research and consultation, such as gathering feedback from patients on privacy concerns. The goal is to ensure that privacy rights and expectations are effectively balanced.

The PACS AI system incorporates various privacy safeguards, including encrypted transfers between clinical and local PACS data stores, and the deployment of localised compute to build models and run them per request, storing only the result. Under Canadian law, AI-generated inferences are considered sensitive personal information, just like the medical images themselves. A significant challenge lies in adhering to legal restrictions that require using only the data necessary for the task, such as limiting metadata retrieval. The issue of profiling, now regulated under Quebec law, also presents legal complexities. Explicit patient consent must be obtained if the system is used to analyse or predict personal characteristics such as health, behaviour, or preferences.

The discussion also touched on the importance of addressing patient trust in AI tools, which extends beyond compliance with privacy laws. Effective communication is needed to ensure patients feel comfortable with AI technologies. In the future, PACS AI aims to adopt federated learning, allowing AI models to learn from decentralised datasets without centralising patient data. However, this shift involves technical and legal hurdles, particularly in handling data transfers across hospitals and provincial boundaries.

PIAs have been criticised because they are a legally-driven exercise, and a risk-based or managerial approach which is not appropriate for what ought to be a fundamental right. Still, it is better to approach the issue constructively. Key lessons including involving developers as well as senior decision-makers and taking an active part in the process as a privacy specialist as part of building a multi-disciplinary team.

Discussion: A question about metadata noted that while PACS AI currently strips metadata during transfers, it was suggested that metadata could enhance model performance. Future iterations of the system may need to reconsider this approach. The discussion of risk emphasized that while zero-risk systems are impractical, identifying and managing risks allows informed decision-making. Consent for AI usage was clarified, with the individual acquiring the image also responsible for obtaining patient consent for AI processing. Concerns about profiling and the anonymity of imaging were addressed, with the clarification that internal systems can still connect images to patient identities. Lastly, the distinction between research and clinical applications was explored, with the acknowledgment that research frameworks allow more flexibility, especially regarding the use of metadata, compared to clinical settings.

Dr Catherine Bourgain, Research Director at Inserm (in sociology of sciences and biomedicine), Head of Cermes3 (Villejuif)

Title: *AI and the work, expertise, experience and meanings in/of healthcare*

Abstract: *The transformative potential of the multiple forms of AI is strong in a large number of biomedical contexts, including the analysis and processing of a variety of signals for diagnosis or prognosis (e.g. images, molecular data...), decision making (algorithms for risk estimation, treatment choice), optimization of care trajectories and biomedical work...*
In all these usages, AI technologies are part of pre-existing professional and social contexts, which they profoundly reshape. This talk argues that our reflections on responsible AI must include a consideration for these properly socio-technical dimensions. Taking into account the impacts of AI on the work of caregivers (reorganization, new practices, new professions, etc.) and patients, on their expertise (including the conditions for transmitting this expertise) and experience, and more generally on what it means to care or to be cared for, is crucial. These ideas will be illustrated with results from ongoing socio-ethnographic studies in the field of genomic medicine.

Summary*:*
Genomic data, produced in digital formats, requires substantial processing, much of which is automated through AI. Bioinformaticians, though relatively new to hospital settings and neither clinicians nor biologists, play a critical role in managing and interpreting this data. Their involvement has become central to producing reliable sequencing data, particularly in the context of genomic medicine, which is dynamic and continuously evolving. Knowledge in genomics is often not fully stabilised, and new cases and data sets present constant challenges for interpretation.

There is an ongoing need for human expertise in genomic data analysis. This expertise, however, varies depending on the context, such as rare or common diseases, and the status of knowledge in that specific genomic area. While AI tools help sort and classify large amounts of data, certain tasks still require human intervention due to the nuances and uncertainties involved in interpreting genomic results. As genomic medicine continues to expand, maintaining high-quality care and the capacity for innovation within care settings remain top priorities.

A large-scale project, "Trajectories of Genomic Innovations," studies the integration of genomic technologies in French hospitals from 2022 to 2026. This project, through ethnographic research, examines how these technologies are being normalised in various care settings. It covers a range of healthcare environments and departments, emphasising the differences and similarities in how genomic innovations are implemented. The study aims to identify both universalities and specificities in the application of AI and genomics across different medical contexts.

The presentation also examined the impacts of new software systems on laboratory work. A key example involved the introduction of new software for managing DNA samples. This system, which has improved efficiency, caused disruptions in workflows, particularly when

technicians struggled with new labels and robotic systems that failed to recognise the updated formats. The software's permission structure, restricting certain aspects of work to formal personnel responsibilities, did not fully align with the established work dynamics in the laboratory. The importance of collective work was underscored, as incorporating technological innovations into care practices requires significant discussion, problem-solving, and the balancing of local practices with new systems.

Discussion: One notable theme raised was the concept of "dirty work" (a term already in use) and the invisible labour performed by staff, particularly prescription assistants and technicians, who play vital roles. The integration of new technologies was seen as adding to the technicians' workload while making their work more mundane and rendering much of it invisible. The discussion raised the importance of ensuring AI is not just about technological and productivity advancements, but also about maintaining the social and human dimensions of work. Another point addressed was the need for flexibility in AI system design, particularly in ensuring that systems can adapt to the existing social structures within healthcare settings.

Prof. Angeliki Kerasidou (Associate Professor, Ethox Centre, and Research Fellow, Wellcome Centre for Ethics and Humanities, University of Oxford)
Title: Trust in AI in healthcare: is it important, is it relevant, is it necessary?

Abstract: *Trust is often discussed as a central issue in the context of data driven health research and innovation, including AI. From trust in the institutions and companies that develop AI tools, to trust in the technology itself, there has been a lot of attention on how to secure, engender and maintain trust. Others, on the other hand, maintain that trust is irrelevant when it comes to AI. AI tools are not the type of 'agents' that can participate in trust relationships. Instead of aiming for trust and trustworthiness, more attention needs to be given to building reliance.*

Drawing from theoretical literature on trust, and also from empirical studies investigating end-users' (patients and healthcare professionals) views and perceptions regarding trust in AI, I will try to clarify what trust and reliance mean in this context. I will suggest that being able to trust or rely on AI might mean different things for different end-users, which reflects their relationship and epistemic standing towards this technology. I will close by suggesting that understanding what trust in AI means and for whom has important practical and policy implications.

Summary:
Trust is often seen as a barrier to the widespread adoption of AI. Yet, despite growing attention, trust in AI is declining globally, particularly in Western countries. Trust, as discussed in the talk, is a relationship that involves more than reliance – it involves an expectation of goodwill and a willingness to accept vulnerability. When trust is placed in another human, there is an assumption of positive intent, which distinguishes human actors from machines like AI, which cannot possess or exhibit goodwill.

AI systems are tools, not moral agents, and therefore the concept of trust may not be the most appropriate framework for discussing AI. AI systems are more like cars, which bear no ill will towards humans when they break down, and consequently do not induce feelings of betrayal. Instead, the focus should be on reliance. In healthcare, professionals and patients prioritise whether AI is fair, unbiased, safe, and explainable. The concern lies not in AI's intentions but in its epistemic reliability – whether it will work effectively for specific populations and tasks. Reliable AI systems are key to success in healthcare, rather than systems that are "trusted" in the traditional sense.

However, certain contexts, such as mental healthcare, challenge this framework. This is true even though within the discipline of philosophy it is a category error to talk about trust in a machine, because machines are tools, not actors. AI chatbots, for example, are being designed to enhance the therapeutic alliance, offering accessible mental health support. In this context, it appears that the perception of care—rather than genuine human empathy—may be important for successful therapeutic outcomes. Patients may feel heard and supported by AI chatbots, and they may even attribute goodwill to these systems. This phenomenon raises questions about the sustainability of such trust, particularly if AI systems fail to meet expectations or users realise their trust is misplaced.

The question of whether trust in AI is necessary also touches on the broader issue of trust in the institutions and individuals behind AI development. Public institutions, such as universities, are often seen as more trustworthy than private companies because of their perceived commitment to the public good. However, healthcare professionals typically focus on the epistemic reliability of AI systems and do not distinguish between public or private sources. Strong regulatory frameworks are essential to ensure that AI systems are developed and deployed in ways that protect the public and ensure reliability.

Discussion: One question explored cultural differences in how trust is perceived, while another raised concerns about the dehumanising potential of AI, particularly when humans are treated as machines. A discussion ensued about whether public institutions are inherently more trustworthy than private companies. One scientist argued that private companies can align their interests with the public good, suggesting that trust in institutions should not be automatically withheld. The conversation also touched on whether AI is fundamentally different from other tools, such as cars or planes, with the proposal in response that AI, like other tools, should be relied upon based on its reliability and performance rather than trusted as a moral agent. The importance of undoing AI hype and building public awareness of AI's true capabilities and limitations was emphasised.

Third session continued and wrap-up of Day Two

Summaries prepared by **Nabila Puspakesuma, PhD student** in Ethox Centre, Nuffield Department of Population Health, University of Oxford

Dr Mireille Régnier, présidente du COERLE (Operational Committee for the Evaluation of Legal and Ethical Risks), le comité d'éthique d'Inria

Title: A few thoughts on Ethics and AI

Abstract: *As AI takes hold of our society, there is a growing demand for AI ethics. Discourses and recommendations are being built around shared values, in the Western world and beyond. The hierarchy of values, and consequently ethical choices, may vary from one field to another. Philosopher Thierry Ménissier distinguishes between Computer ethics, Robot ethics, Digital ethics and Usage ethics. We argue that a new ethical principle must emerge: that of differentiation between man and machine. We will deal with four examples: Bias of algorithms, chatbots, Information Warfare and AI and Justice.*

Summary*:*
This talk focused on the opportunities and challenges for responsible AI, with reflections from French philosophy and contexts. The challenges, Dr Régnier argued, mainly come from a contradiction between strain and enthusiasm in AI. For example, when looking at how innovations in AI are advancing rapidly, issues concerning reliability may arise. On the other hand, there is an increasing demand for ethics in academia, publication, and also from society when it comes to responsible AI, which can be an opportunity.

In AI, we delegate human tasks to a machine, prompting the question of responsibility. Dr Regnier presented four cases to expand the issues: computer ethics, robot ethics, digital ethics, and usage ethics. In each case, Dr Régnier identified participants, main challenges, primary and secondary values, issues of human vs machine control, and issues of responsibility attribution.

The first case, computer ethics, mainly deals with AI design. The participants in this case are co-designers, mathematicians, and computer scientists. Its primary and secondary values are explainability (prediction created from the data) and transparency (philosophy of enlightenment; science enlightens citizens). The main challenge of computer ethics is clarifying the technical system – not only in an epistemic sense, but also from the point of view of value (e.g. algorithmic bias). Dr Régnier believes that algorithms should not just be controlled by AI, but also by human beings.

The second case, robot ethics, operates in the context of artificial artefacts that interact with other artefacts or human society. These artefacts may include chatbots, autonomous vehicles and drones. The participants are specialists in robotics, engineers, and technicians. Its primary and secondary values should be safety and civility. Issues concerning human vs machine control in this area may include social learning, civil responsibility, and chain of command (e.g. military drones). Dr Régnier argued that the final responsibility in robot ethics should also rest with humans, but here is a question to consider: if we attribute legal personality to a robot, should it be responsible?

The third case presented was digital ethics, which aims to design and use network platforms to enable interactions (e.g. smart city, e-medicine). The participants are networks and platform designers and managers. Its primary and secondary values should be accessibility, privacy, and civility. Notable challenges highlighted by Dr Régnier included risks of modern information and communication, such as information warfare. In terms of assigning control

to human versus machine, issues such as confirmation bias, false information, training and epistemological reflection may be considered.

The last case, usage ethics (or UX AI ethics) deals with implementing and deploying AI in all areas of social activity. Usage ethics brings forth fundamental questions related to the nature of human, e.g. when tasks (writing reports, doing exams) are automated, what does it mean to be and remain human? How do we identify as a human in this era? What do we bring to society? The participants of this community include social scientists, with particular concerns regarding thoughts, vigilance, and regulation. The primary and secondary values should be autonomy, inclusion, and fairness, provoking further questions about access and distribution. Dr Régnier argued that the balance of power in usage ethics is now shifting towards the users more, and therefore appropriate regulation should be in place to mitigate potential harmful risks. For example, in the French context, there is a regulation that specifically requires companies deploying AI chatbots to explicitly state that users are interacting with a machine.

During the discussion session, some clarifications were made regarding the value of civility in the case of digital ethics. This value emerges from past occurrences when robots were found to make racist responses, and how the methods we choose to train our algorithms might lead us to under-civilised or over-civilised traits.

Dr Caroline Green (Early Career Research Fellow, Institute for Ethics in AI, University of Oxford)
Title: AI in the care of people with dementia: Benefits and ethical challenges

Abstract: *Trends in demographic ageing are leading to an increasing number of older people living with dementia. AI offers opportunities for people living with dementia and their carers to support independent living for longer and a higher quality of care. However, AI also poses unprecedented ethical challenges in this domain. Drawing on a social model of dementia and the UN Convention on the Rights of People with Disabilities, this talk will discuss implications of AI on pressing topics for dementia care, including mental capacity assessments, supporting independence and autonomy and beating social isolation and loneliness.*

Summary*:*
Dr Green's talk revolved around her work and reflections on integrating AI for the social care of people living with dementia. During the first section of her talk, she explored the human rights situation of people living with dementia, particularly in the UK. She discussed beliefs surrounding dementia, such as how dementia is seen as part of normal ageing, and how the public seems to think that health professions ignore people with dementia, with various factors (stigma, intersectionality, and gender) coming into play. People with dementia also report experiencing stigma and discrimination in at least one area of their life.

The UN Convention on the Rights of People with Disabilities provides useful guiding principles that are relevant for service provision for people with dementia. Through this framework, human rights are seen not only as a legal term, but also as a value. Nonetheless, it is also important to look beyond human rights and consider other elements such as

friendship, love, kindness, compassion, a sense of belonging, a sense of community, and a sense of purpose. These are what make us humans, and when it comes to incorporating AI into social care, the question of how we make AI work for humans becomes even more fundamental.

There also appears to be differences in how social care can be defined. One definition can refer to the practical and emotional support with activities of daily living with a view to maintain as much independence and autonomy – but also relational – as possible, offered by professionals and or informally by family members, relatives in venues such as care homes, at home, in hospitals, day centres, and so on. Defining 'high quality social care' in England & Scotland points to these characteristics: highly regulated (inspections carried out by professional sector such as Care Quality Commission or Care Inspectorate), dependent on value-based law and regulation (human rights, person-centred, safety (conflict of freedom and safety) and process-oriented (with high administrative load, writing care plans, and sometimes staff shortage).

This distinction becomes key when we intend to bring AI into the picture, as it can be designed to fulfil different purposes. For example, Chat GPT in formal social care settings can be used to respond to administrative, non-care related tasks (e.g. write email), administrative, care-related tasks (e.g. write care plans; quite empowering because it removes some of the workload) or direct care related tasks (e.g. carers asking clinically nuanced questions to the AI). There is also the case of DEDICATE, Dr Green's project, which is based on needs analysis and endeavours to provide accessible trustworthy information all day using Open AI GPT 3.5. This project also partners with organisations to offer high-quality content and hopefully build the trustworthiness of people towards the tool. There is 'Call the AI' and 'Lola', which use generative AI to take inputs from the carer and generate reports to send to families. The cases are not without challenges, however. For instance, during DEDICATE's initial trial, the algorithm occasionally came up with problematic responses towards some users.

AI becomes increasingly seen as a panacea and key to efficiency in social care, but everybody is also worried about the ethics of AI. Everyone wants to know more and have perspectives about it, and AI developers are desperate to grasp onto a legal and ethical guideline. Dr Green argues that it is important to question the big picture assumptions:

1. Is AI the best solution for a particular problem?
2. Whose values should be upheld?
3. What type of caring society do we want to create?
4. Who is going to benefit?

A multi-layered approach is also helpful to unpack potential ethical challenges when designing and deploying AI in social care settings through questioning the directionality of risk (to whom/by whom?), the source of the risks (is it from the technology? Human use? Or others?), whether the risk is direct or indirect, legal assessment and adjustment (what are the relevant laws/regulations? Mental Capacity is very complex, how can AI fit in?), ethical assessment (what rights and values are at stake? How do we catch them early before they play out?), and last but not least, who should answer all of these questions. Dr Green

illustrated the use of this multi-layered approach through the Oxford statement on the responsible use of generative AI in Adult Social Care—a collaborative effort between different agencies (25) and people living with dementia.

The discussion session brought up several questions. The participants and speaker discussed the problem of the skills gap amongst social carers who, coming from different social backgrounds and notions of social care, may not necessarily be equipped with appropriate competencies to navigate AI in their work. AI might be able to help with training, but currently there are no guidelines on best practice.

In terms of development process, there are concerns, that the quick and dynamic technology market may not tolerate the multi-layered approach, which may take time. It is therefore important to think outside the box to create a space where people involved can consider the questions iteratively at different points in time as the process goes along and have the courage to come back to the questions.

Prof. Catherine Régis (Université de Montréal, IVADO, Mila)
Title: Exploring the Role of Human-Centered AI (HCAI) Approaches in Healthcare

Abstract: *HCAI focuses on furthering human values and ensuring human usability. Combining these normative and technical desires must be done through a well-crafted interactive process of defining appropriate norms through regulation with a more context sensitive strategy that allows understanding, responding, and adapting to such desires at a practical level, that is within healthcare institutions. This presentation will explore what such interactive process could look like and why, more broadly, it is important to connect norms with context in AI good governance approaches in healthcare. Some concrete illustrations of Canadian experimentation with that respect will be given as well.*

This presentation is partially based on the following paper: M. Da Silva, J.-L. Denis and C. Régis, "Good Governance Strategies for Human-Centered AI in Healthcare: Connecting Norms and Context" in C. Régis and al. (editors), Human-Centered AI – A Multidisciplinary Perspective for Policy-Makers, Auditors, and Users, CRC Press (London), 2024, 342 p.

Summary:
The AI market is massive; we invest a lot of money in AI and healthcare. Still, many AI developments in healthcare are not implemented in clinical settings, which reduces the benefits we gain from the technology. T. This is why Prof. Régis focused on the alignment of aspiration with reality. She argued that there are various reasons for the clash between aspiration and reality in AI in health care. They include disconnection between IT system development and adaptation; ethical and legal issues (issues of responsibility, level of accuracy, and biases); lack of education and literacy of healthcare professionals about AI; and lack of time to experiment with and adapt to the tech, increasing fear and diminishing trust.

Prof. Régis has tried experimenting with alignment in some of her work, resulting in increased trust and more appropriate training when using AI. She argued that good governance

strategies are key to aligning inspiration and reality, and that it should also allow for a better connection between norms (regulation) and context (clinical/institutional).

Prof. Régis further explained what to consider in norms and context. To exercise risk-based approach in norms, we should ensure objective and proper performance of AI tools in health care (efficacy and security), developers and deployers accountability (answerability and responsibility, reporting mechanism) and sufficient information for appropriate usage for end users (intelligibility and agency). Within the context in which the AI is deployed, concerned actors should be able to influence the journey of innovation in practice, negotiate the success of AI innovation in the context of application, and engage in reflexive processes.

Echoing Dr Green's previous talk, Prof. Régis underlined that it takes time to align aspiration with reality, and that space and support mechanism should be encouraged through innovation labs, education strategies, ongoing access to developers, and sufficient resources for the implementation phase. It is also fundamental to create normative components and context opportunities where information can proceed from the world of end-users to regulators and back, allowing feedback loops to continue and more research to develop.