AI@KI: Final Report[1]

Magnus Boman

27 January 2022

[1] The Tufte handout style used originated with Edward Tufte. This and most of the other margin notes have hyperlinks to sources.

Artificial Intelligence (AI) means inductively jumping to conclusions through clever guesswork and then over time learning how to automatically correct for error. Karolinska Institutet (KI) is considering the strategic importance of AI and becoming informed about the current status of AI interest, competence and deployment is therefore crucial. All interviewees and discussants were from KI or its surrounding ecosystem.[2] The knowledge elicited allows for a vision in which the resources necessary for using AI widely at KI are in place, with smooth processes for the sharing of data, models and results. The move to wider AI use within health needs to be fast because there are burning issues that AI can solve using unique Swedish opportunities for data-driven research and even automation, but the process also needs to be slow enough to build competence and trust in AI methods, in a sustainable way. KI is in this future an international top player in fair and efficient AI deployment. Processes and insights are ready to be exported, in an ethical and respectful way. The validated and evidence-based AI systems have reached clinicians and their patients, with AI having moved from successful pilots to wide implementation. This future will be realised chiefly through precision diagnostics and care. Precision medicine requires multimodal patient stratification, and AI excels in fusing different modalities for improved performance on prediction and classification tasks. For significant health-related problems for which AI has been scientifically proven to make a difference, it could be considered unethical not to at least test and validate AI methods in realistic circumstances. At the same time, it is important to take a critical stance on the value added, asking hard questions about costs and methodologically immature parts of AI. The project focus was on the impactful implementation of AI, not on visions, and so the people whose work has been scrutinised are self-motivated and driven. Because they range from established principal investigators heading large groups or clinics to individuals so far without any local AI-support, the perspective is bottom-up. Only by grounding findings in this manner can a top-down strategy feasible to implement and support be devised by the President of KI, the main stakeholder. This final report, after an executive summary, has the following disposition.

[2] For the sake of brevity, I will refer to the ecosystem and its inhabitants as simply *KI*, regardless of employment or legal status.

1. Introduction
2. Ethics
3. Laws and Regulations
4. The KI Ecosystem
5. Scientometrics
6. AI Deployment at KI
7. AI and Precision Medicine
8. The Way Forward

Figure 1: The project Web page has been a dynamic portal for AI@KI information since the spring of 2020. Graphics by Marie Lind.

Executive Summary

The main findings from the project can be split into neutral (*o*), positive (+) and negative (−) one-liners where the negative ones prompt actions to achieve organisational change.

+ There are more than 50 people conducting or leading health-related research at KI who use AI methods, with good and often published and disseminated results

+ Modalities studied for AI processing in KI research efforts are not limited to images, text and audio, but also include biomarkers, paving the way for advanced multimodal fusion models

+ KI and Karolinska jointly encourage AI methods and models for precision medicine

+ There is an active seminar culture on AI-related matters at KI

+ Methodological discussions on what AI can do for KI people are vivid and well informed by an overlap with biostatistics, bioinformatics and health informatics

+ The opportunities for research funding for AI applied to health are diverse and relatively many, including the massive support for data-driven reasoning offered within the Data-Driven Life Sciences initiative

+ Results from AI use at KI are in some cases of such high quality and technically innovating that they could be published as computer science research and not only as research in the life sciences

+ Many interdisciplinary collaborations are ongoing and some constellations have sustainable long-term AI employment as a goal, including successful people migration in some cases, in which another university pays for a researcher spending time at KI

+ Strong researchers at KI are interested in and are making efforts to deploy AI for serving the underprivileged parts of the global population

+ The sentiment towards AI is very positive at KI, not only among researchers

+ There is full management support for AI use, now and in the future

This report addresses what must be true in order to maintain the above positive aspects of AI, as well as to address the issues that lie behind the following negative aspects.

– There are successful concluded pilot research projects involving AI methods that have still to develop into full projects with the potential for future use at the clinic

– Like the rest of the world, KI is affected by the talent problem, and long-term interdisciplinary collaborations to secure competence in AI and data science are therefore needed

– There is considerable myopia among different people developing models and implementing AI-related systems at KI

– There is currently no single unit or centre at KI for AI-related questions and support

– While individual researchers have developed and sometimes published state-of-the-art AI solutions to problems in health and medicine, KI is yet to be internationally recognised as having a strong AI profile

– The spectrum of competence and maturity on AI is very wide at KI, forcing efforts towards further education and hands-on AI experience to be tailored to individual or small group needs

– There are research leaders at KI that consider AI methods to be something they have tested and failed to achieve satisfying results with, prompting skepticism towards AI on their part

There are also neutral one-liners that have positive as well as negative interpretations, and which could affect future-proofing:

o There are a few cases of individuals and groups diving deep into technology, specifying and building their own hardware devices for data-driven reasoning, furthering their own knowledge on IT but without direct cooperation with other groups or IT departments

o New life science innovations, such as new sensors or scanners, are in a few cases being connected to machine learning methods for output data processing and understanding

o There are two distinct groups of people employing AI methods at KI, the first consisting of established principle investigators or group leaders, and the second consisting of young researchers at the beginning of their first project or in education

1 Introduction

1.1 Method

With one day per week at my disposal, I settled on semi-structured interviews with principle investigators and young researchers.[3] To the last days of the project, a shortlist of people to interview and advise has been in hand, with people continuously added to the list whom only recently picked up on AI methods and techniques. This report covers more than 100 researchers and practitioners, with more than 50 people driving AI use at KI, albeit varying immensely in their respective experience of AI. Their interests come in two main (overlapping) categories: data-driven reasoning and machine learning. The former allows for exploratory and hypothesis-less data mining, including finding support for causal relationships or correlations. The latter covers prediction and classification (usually with supervised methods), as well as clustering (usually unsupervised).[4]. Current users of machine learning have been placed on a ladder I developed to assess maturity (Figure 2). The assessments pertaining to individuals or groups at KI are not part of this report, but the general lines of my observations frequently generalise or anecdotally refer to those assessments. The average number of steps taken on the ladder is three, and a step or two is sometimes skipped. A small minority are at the bottom rung, close to the firm ground of stable AI-augmented research and development, with at least five steps fully explored. The AI@KI project is very much studying a moving target, however, and after my first-year reporting back to the stakeholders, I have sometimes checked back on individuals or groups, noting that they have recently climbed down another step.

[3] I have consciously tried to avoid the senior/junior dichotomy, but I have found myself still using it at times, it is as if I have been bitten by a bug.

[4] Under the lead of my KTH colleague Erik Aurell, I will be co-arranging an on topic Nobel Symposium called *Predictability in Science in the Age of AI* in October 2022, in South Africa

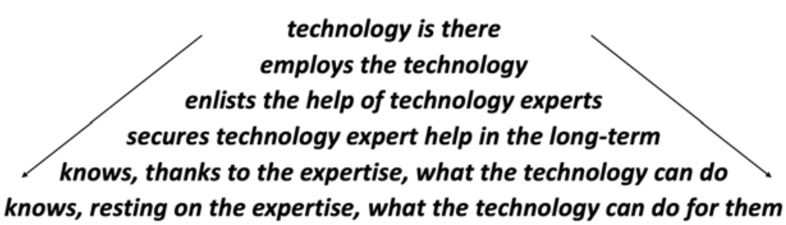


Figure 2: The machine learning maturity ladder. I use the ladder metaphor because some people skip a step or two when climbing down. This is fine in some cases, but in other cases it points to possible improvement. The maturity ladder is merely a handle on interest, competence, maturity, and future prospects. It provides no substitute for deeper structured analysis, but like all handles it is at times convenient.

A concerted effort on AI use rhymes well with the 2030 Strategy of KI.[5] Because the project goals have long-term strategic implications, my work has not analysed current or short-term risks with AI employment. What I can observe and assess today will only constitute pieces of a large puzzle. A thorough risk analysis requires co-creation and full stakeholder involvement. What I have done is to engage with experts on ethics and on law,[6] and I have also engaged directly

[5] On page 18 in the strategy document, it is stated that "KI employees must be given the opportunity to use data management and programming tools and advanced quantitative methods."

[6] Besides Gert Helgesson and Niklas Juth at KI/LIME, I have had enlightening discussions with professor Frantzeska Papadopoulou Skarp at SU/Law.

with the communities of practice in the KI ecosystem. Many of my conclusions rest on direct and situated observation. The largest risk—that AI remains a success case in research projects and pilot tests, but is not adopted at the clinic—was addressed in a master project in health informatics that ran through the first half of 2021. Another important AI@KI object was a synchronization of barriers and opportunities with those of project Clinicum, with a planned roll-out in 2023.[7] Clinicum will support clinical research on study design, biostatistics, bioinformatics, and AI should now be added to this list, since a forthcoming report on Clinicum shows expectations from KI scientists for this to happen. From a questionnaire ($n = 542$) sent out to the Clinicum network, in which respondents could indicate their need for support, a recently completed analysis of the free text fields showed that AI and machine learning, as well as data wrangling, management and cleaning were important. The results of the efforts on targeting the main risk included deliverables in the form of a master thesis and a journal paper, detailed in Section 6 below. The next subsection describes some KI-relevant methodological perspectives not covered by those deliverables.

[7] Project Clinicum is a collaboration between KI and Region Stockholm to support researchers who need access to clinical data or data management and methodology support. It is led by Erik Melén and Karin Ekström-Smedby, jointly. A report detailing Clinicum will be presented by Sandra Eloranta (KI/KEP) in February 2022.

1.2 Data

Imaging is the most important modality today for AI analysis of health data. A good indication of what research efforts are required to proceed with AI for medical imaging was published as a roadmap by a large radiology consortium, three years ago:[8]

[8] Key Points from page 782 of Langlotz, C.P. *et al.* (2019) A Roadmap for Foundational Research on Artificial Intelligence in Medical Imaging, *Radiology* 291:781–791.

- New image reconstruction and enhancement methods are needed to produce images suitable for human interpretation from the source data produced by the imaging device.
- Automated labeling methods are needed to rapidly produce training data for machine learning research by extracting information from narrative reports and clinical notes.
- Novel machine learning algorithms are needed that are tailored for the complexity of clinical imaging data, which are often high resolution, 3D, 4D, multimodality, and multichannel.
- Machine learning systems must be capable of explaining or illustrating the advice they provide to human users (so-called explainable artificial intelligence).
- Aggregation methods for clinical imaging data are needed to produce the large volume of data necessary to train machine learning algorithms.

Because of its multidisciplinary nature—involving physics and computer science, as well as the life sciences—AI for medical imaging has spawned new forms of collaboration at KI. These include MedTechLabs for KTH, Karolinska university hospital and Region Stockholm collaborations, and SciLifeLab for KTH, Stockholm University and Uppsala University collaborations. The SciLifeLab Bioinformatics Platform (NBIS) is also involved with the Analytic Imaging Diagnostics Arena (AIDA) Data Hub for AI applied to sensitive image data. Such centres and initiatives also allow for social networks to form. Since there is no AI centre at KI, some of my time has been taken up by mapping out such networks, spread over the whole KI ecosystem.[9] In Section 5, bibliometric networks involving AI are highlighted.

At Neuroradiology, six MRI physicists are employed—stressing the interdisciplinary nature of the work—who use deep learning filters available on some of their MR systems, leading to massively reduced image noise. There is considerable programming experience in the group and also an interest in Radiology Information System (RIS) data. Like the Picture Archiving and Communication System (PACS) and the Digital Imaging and COmmunications in Medicine (DICOM) standard for digital image exchange, RIS can help measure the sensitivity and specificity of signals in x-rays, which is of great interest to natural language processing methods. A new CT lab recently opened at BioClinicum in Solna, offering next-generation computed tomography, thanks to photon-counting technology. This means much higher resolution and greater detail at a lower dose of radiation. Next-generation on-scalp sensors are also being tested at the MEG lab nearby, where AI has been used to process sensor output, details are in the applied AI examples appendix to this report. At this scale, quantum interference techniques are used, and there are also efforts to look into spin models of raw data. For example, MR images are Fourier transforms of raw data. If a machine learning algorithm works directly on the so-called *k-space*, consisting of an array of numbers representing spatial frequencies in the MR image, AI-driven raw data processing might allow for shorter scanning time, leading to faster examination time. In 2021, a quantum life science interdisciplinary hub was created, under the lead of Ebba Carbonnier, SWElife and KI. Several KI researchers participated in the first Nordic Quantum Life Science Round Table, which I helped Ebba arrange in November 2021,[10] such as Per-Olof Berggren, who in his talk demonstrated the importance of quantum microscopy for signalling pathways. Currently, AI for sensing and neuromorphic computing are the two areas of technology development that have obvious life science applications.

[9] Just to give a flavour here of the cross-organisational interest in AI in the context of imaging: among the scientific directors for MedTechLabs, at least one (Kevin Smith) is also at SciLifeLab, and another researcher (Johan Hartman) was among the interviewees for AI@KI. Its vice director (Staffan Holmin) is also on the Precision Medicine task force, in Diagnostic Development, to which I too belong. On the board of MedTechLabs is the vice resident of KI (Anders Gustafsson), one of the main stakeholders of AI@KI, as is Birgitta Janerot Sjöberg, who is also on the board of AI Sweden.

[10] The roundtable was an IRL meeting in Solna drawing participants from the Nordic countries. Many WACQT (Wallenberg Centre for Quantum Technology) researchers participated, including Göran Johansson (Head of the Applied Quantum Physics Laboratory at Chalmers university, who leads the theory efforts in quantum computing and simulation in WACQT) leading to discussions on new solutions for life science problems coupling quantum technology with AI. Next November, the second roundtable is planned for Copenhagen.

Folklore has it that 30-45 per cent of the work in any AI project (imaging or not) is spent on pre-processing digital data, getting it ready for learning. Different machine learning methods have different tolerances for missing values and noise, which affects the time spent in pre-processing. Imputation is sometimes necessary and at other times forbidden. Part of working in a data-driven fashion means committing to not destroying data, meaning risk scores and other simplifications of data properties are ideally not used by learning algorithms. Labels and a so-called gold standard to be emulated or even outperformed for supervised learning are often used. Human annotations can also be disregarded, such as when clustering data points using unsupervised machine learning, even if choosing clustering methods and visualization techniques still require human input. To realistically estimate the time required is important, for many reasons, including the following.

- Most researchers consider pre-processing a tedious task and would prefer to "get on with the work"
- AI is marketed as a means to data processing that simply slurps as much data as possible into the mix and returns wisdom and insight, without much concern for the nature of the input
- There is a risk that unnecessary digitalisation efforts, such as digitizing video material, are given priority before considering what kind of data is lacking and what is already present[11]
- The sensitivity of health data means that a data policy and ethical permits must be in place, and the extent to which this affects the project usually becomes clear only after pre-processing has started (i.e., when the first email reply from the judicial department arrives)

[11] It is unfortunate that AI is often placed under the banner of Digitalisation in the health domain, since it is not necessary for AI methods to exclusively process digital data. Besides analogue computing, e.g. using physical reservoirs as in neuromorphic computing, it is often fruitful to look at metadata for analogue material. It is sometimes just as interesting when and how a video came into existence (meta-level data) as its content (object-level data). The latter can then be digitalised later, as necessary.

To understand what an AI system is doing with your data has been in and out of vogue since the first expert systems were applied to medical data, in the 1970s. MYCIN, developed to help identify bacteria causing severe infections, gave its users the opportunity to ask why a particular rule had been triggered in a chain of reasoning.[12] Arguably, the hype around such systems and what they could support the clinic with, helped create the 'AI winter' that followed.[13] As AI slowly crept back into organisations and companies, it often returned under different monikers, taking the 'Why?' questions from the users more seriously again, not least to provide better customer service, which must avoid the "Computer says NO" message at all costs. A few years ago, AI researchers seemed more determined than ever before to open up their black boxes, motivated by requirements

[12] Shortliffe, E.H.; Buchanan, B.G. (1975). A model of inexact reasoning in medicine. *Mathematical Biosciences*. 23 (3–4): 351–379.

[13] IBM has been struggling with bringing back expert systems to health applications through its Watson system, notably in oncology. Several large collaborations produced *Watson for Genomics* and *Oncology Expert Advisor*, now discontinued, within Watson Health.

on transparency from politicians and grant providers. This was in part due to the ubiquity of deep learning approaches in many new domains, meaning that in the life sciences, many researchers were exposed to impressive results for many tasks. With clinical guidelines and ethical codes to adhere to, the explainability of AI algorithms (XAI for short) became a judicial matter. Fears around what roles GDPR and MDR would play further increased the need for judicial support. This, in turn, made researchers sometimes abstain from computing optima—and so from using the tool that had the lowest computational complexity—if it obscured its processing unintentionally. Tools like SHAP—a clever way of illustrating the importance of each feature in a learning model, and hence explaining the model, albeit in a weak way—became more widely used.[14] When it was recently demonstrated how easy it was to corrupt such explanations,[15] we found ourselves right back at the start: what counts is *trust* in a model. Since it is so hard to explain how learning works even for a shallow neural network, the pendulum now seems to slowly swing back towards less focus on explainability.

[14] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 4765-4774.

[15] Slack, D., Hilgard, S., Jia, E., Singh, S., & Lakkaraju, H. (2019). Fooling LIME and SHAP: Adversarial Attacks on Post hoc Explanation Methods. Proc AAAI/ACM Conf on AI, Ethics, and Society, Feb 2020, 180–186.

After the obligatory pre-processing and model family selection steps, an actual model can be designed, implemented and tested, with some internal validation (always) and some external validation (only if we are lucky enough to get to run an RCT or similar). The internal validation is necessary for replicability of results and will also determine if the model can be generalised. Many health applications suffer from overfitted models that do handle new data well and could not be transferred to another patient population, for instance. Overfitting is in theory easy to avoid but in practice, a small n is particularly hard to handle well. What we should do, assuming a train-test-validate loop, is compare our results from training to our test results. If the difference is large (in the favour of the training results), we are overfitted. In practice, we cross-validate five or ten times but this shrinks our dataset, since the holdout sample takes away training data. It is therefore tempting to use the sample also for training, thus testing on a subsample of already seen data.

Another temptation is to use hyperparameter tuning to get better quantitative results, as measured with e.g. F1-score, balanced accuracy, or ROC-AUC.[16] This sometimes only replaces points by intervals, to give our variables some slack, which is not a bad idea if you have the computational power to test for all points inside all the intervals. As this is often achieved by brute force search, the computational complexity can be forbidding. But there are also other kinds of hyperparameters, which require more from the modeller in terms of methodological skills. An example would be parameters that dictate how to branch decision trees, another would be how to

[16] Luckily, the mapping between old Fisher-style statistics and machine learning lingo is beautifully summarised in one extremely dense but useful diagram, part of many a Wikipedia entry.

regularise, or how to mix different penalties in regression. The language of choice usually sports libraries of code for such things, all of which are almost too easy to employ. Python in particular has an extensive and well-documented collection of scripts for hyperparameterisation. Other languages, like R or MATLAB/Octave, have other selling points but both enjoy a large community of users, so that there is always someone ready to send you snippets of code to fix (*sic*) any problem. Healthcare is particular in its use of SAS, while general programming languages like Java and C in particular are used to a relatively small extent.[17]

[17] A nerdy note is that SAS itself is implemented in C.

So-called data dredging—abusing data in search of significance via brute force correlation or p-value hacking, for example—has no corresponding term in AI, since much of data-driven reasoning requires mere association studies for getting results. The exploratory nature of AI methods may produce spurious and irreplicable results if the methodology is incorrect, but there are no hypotheses to identify via exhaustive search or reverse engineering. To save all random seeds used for stochastic variables is considered good practice as it supports replication studies.

For AI programming, a dedicated AI software platform is employed, such as TensorFlow, Keras or PyTorch. Such platforms can be used together with other extensive open source and often Python-based software packages. Integrating these system components gets easier with programming experience, and at KI such experience is definitely there, albeit in spots. There are individuals and teams that require no basic training in any component, as they move between them with ease and are also capable of switching between combinations quickly. For those with interest and needs, but without solid basic training and support, a community was built at KI around the so-called Falafel seminar series, detailed in the appendix.

Figure 3: A very ambitious reference volume on AI in medicine was co-edited by Niklas Lidströmer, who belongs to Eric Herlenius' group at KI, in clinical pediatrics.

2 Ethics

There are two concise problems related to ethics and AI that have a direct bearing on KI work.[18] In short, one is about bias and the other is about equal access to AI solutions. The most pertinent bias problem leads to skewed training of machine learning systems, leading to problems in testing and validating an AI system. I have myself reported on what I called "an embarrassment" in the form of a pipeline for facial feature recognition in old video material taking longer to identify non-white faces than white ones.[19] To me, the most embarrassing element was when the clinicians apologised for having caused the problem by not providing us with a balanced set of examples. I then had to explain that their material—its substantial size

[18] For reference, the comprehensive *AI Index Report 2021* reports the following (p.128). "The five news topics that got the most attention in 2020 related to the ethical use of AI were the release of the European Commission's white paper on AI, Google's dismissal of ethics researcher Timnit Gebru, the AI ethics committee formed by the United Nations, the Vatican's AI ethics plan, and IBM's exiting the facial-recognition businesses."

[19] See Figure 3 and 4, in particular, in Boman, M., Downs, J., Karali, A., and Pawlby, S. (2020). Toward Learning Machines at a Mother and Baby Unit. *Frontiers in Psychology* 11.

notwithstanding—was not enough to even in the slightest affect the performance of the underlying pipeline, which was an open source software trained on millions of images. Since this work was not carried out in Sweden, an example closer to home could involve almost any deep learning system.[20] As mentioned earlier, a deep learning filter can be acquired, at no small cost, for improving the precision of imaging in an magnetic resonance scanner. As in the case of a software patch offered to the owner of a modern car or a digital assistant in the home, software updating via patching is optional. However, as the precision is easily seen to increase, many image modality owners would opt in. Even if the operators have substantial knowledge in physics, as they do at the Neuroradiology unit at Karolinska, it is extremely hard to know whether the increased precision is fairly distributed over all kinds of brains imaged or not. Perhaps imaging of one part of the brain, or at some contrasts, or of some pathologies is improved, but not others. And whether or not differences matter in practice, the fairness issue is ethical in its nature, and as long as there is at least one patient that would be less served by the underlying deep learning system, it is important, not least for future consideration when deep learning patches might be ubiquitous.[21] Fairness should also be compared to what went before, when there were no AI tools at hand, with an entirely human-devised process flow in place.

[20] In an infamous example of racial bias in healthcare algorithms, a quote from Rayid Ghani, a computer scientist from Carnegie Mellon University, is worth remembering: "We are still using these algorithms called humans that are really biased. We've tested them and known that they're horrible, but we still use them to make really important decisions every day." (p.609).

[21] See, e.g. Matthew, D. A new type of powerful artificial intelligence could make EU's new law obsolete, *Science | Business*, 21 Dec 2021.

The MRI example also serves to illustrate the equal access problem. If one patient has brain images taken by a patched MRI system and another has theirs taken by a non-patched system, this might mean little. But if the latter patient finds out, the positive sentiment towards AI in popular media and in society in general might affect this person to feel that they were underserved in the care they received. Why would they too not benefit from AI enhancements? Again, these are early days, with few examples, but if AI does prove to the quality of care, possibly at lower cost, we will beyond doubt see a gradual increase in its use. A related example would be an AI-augmented system for reassessment of mammography screening: letting human experts and AI algorithms in tandem revisit retrospective material, as done in the ScreenTrust research project at KI. The Breast Imaging unit at Karolinska will likely benefit from deep learning systems trained on more than two million mammography images, and for any individual patient this might mean improved chances of finding previously undetected problems in their images. The ethical problem then arguably arises for a patient whose images are not subject to AI-enhanced reassessment.[22] In particular, if good research results are reported on, such a patient might jump to the conclusion that the system has successfully passed clinical trials.

[22] What is arguable is whether or not this is merely a matter of fully informing the patient. A related point is how to best inform patient about technically complicated advances in medical technology.

None of the problems listed above should make researchers abstain from AI use. Instead, there are ethics resources at KI that can help minimise risks and future-proof AI research. For deep learning, there are no quick fixes, but substantial efforts are being put into methodological development in learning systems for both industrial and governmental use, in anticipation of stricter regulations in the future.[23]

[23] This was a discussion theme on the AI Sweden ethics council, which met three times in 2021, and on which I represented KI.

3 *Laws and Regulations*

Since the inception of AI@KI, a wealth of relevant policy regulation documents have seen the light of day, a select list being:[24]

[24] *White Paper on Artificial Intelligence: a European approach to excellence and trust*; *Ethics Guidelines for Trustworthy AI*; *Draft Recommendation on the Ethics of Artificial Intelligence*; *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts*.

1. The European Commission's white paper on AI
2. The EU Ethics Guidelines for Trustworthy AI
3. A UNESCO draft recommendation on the ethics of AI
4. The EU Regulatory Framework Proposal on AI (The AI Act)

A regulating document can be evaluated on what it forbids and prevents—or on the environment of self-regulation it helps create—and not on how many times it tells people to be nice or careful. The latter is covered by ethics and not necessarily improved via regulation, as it in part is governed by norms. To formally regulate means to judicially regulate, which is not what every seemingly relevant document does.[25] We are then left with recommendations on informal regulations. This paints a bleak picture in a world where private companies are free to develop autonomous war machines that rest on AI technology:[26] surely, health is an area where we should be able to do better? One may even consider non-constructive informal regulations as cynical, especially those documents written by people ill-informed about the technological ramifications of AI and about legistics. At KI, the Swedish biobanking law (2002:97) has largely affected the working processes at the KI Biobank, and will continue to do so in the future, since the law will be sharpened in 2022 in some respects, not loosened.[27] If AI in the future is to be employed to fuse multimodal signals from various modalities, including those signals detectable from samples held at the biobank, then a law may or may not restrict the future of precision medicine, but at least formal regulations can be addressed paragraph by paragraph. The intended interpretation of the lawmakers can also be discussed openly and advocates of AI for precision medicine, myself included, can provide crisp practical cases for reformulating or reinterpreting the law, as necessary, to the benefit of Swedish citizens.

[25] I have blogged (in Swedish) about the lack of bite in general of the EU Ethics Guidelines for Trustworthy AI, so I will not comment further on it here. Suffice to say it was prepared by the High-Level Expert Group on Artificial Intelligence, an independent expert group set up by the EC in June 2018, as part of the EU AI strategy.

[26] A concise analysis of the collision course between humans and machines in armed conflict was published already in 1991, and it has not been improved upon by AI naysayers since then: Manuel De Landa, *War in the Age of Intelligent Machines*, Zone Books, New York.

[27] I made a site visit to the KI Biobank, which was illuminating; not least the wide variety of material kept there was remarkable, but also how laws affected the daily work. Details on relevant laws are available (in Swedish) from Sveriges Riksdag, *Lag (2002:297) om biobanker i hälso- och sjukvården m.m.*.

The EU White Paper starts with the words "Artificial Intelligence is developing fast. It will change our lives by improving healthcare (e.g. making diagnosis more precise, enabling better prevention of diseases)..." so it is definitely relevant. Health is named as one of the sectors within which "Europe has the potential to become a global champion" (p.6) and "It is essential that public administrations, hospitals, utility and transport services, financial supervisors, and other areas of public interest rapidly begin to deploy products and services that rely on AI in their activities. A specific focus will be in the areas of healthcare and transport where technology is mature for large-scale deployment." (p.8). A very succinct point is made on AI in the context of improving current legislation (p.18):

> *Changing functionality of AI systems*: the integration of software, including AI, into products can modify the functioning of such products and systems during their lifecycle. This is particularly true for systems that require frequent software updates or which rely on machine learning. These features can give rise to new risks that were not present when the system was placed on the market. These risks are not adequately addressed in the existing legislation which predominantly focuses on safety risks present at the time of placing on the market.

By now, every health professional will know basic facts about GDPR and how it by large does not impact research as much as originally feared. What is less known is that the consequences of violations, largely covered by the *ePrivacy regulation*, are yet to be implemented in practice. Much confusion and delay were caused by one forward-thinking lawyer by the name of Maximilian Schrems. His judicial actions against data moving between Europe and the U.S. in particular has greatly impacted data retention laws, and in effect to the Court of Justice of the European Union deciding that U.S. law does not sufficiently protect the privacy of Europeans, now known as *Schrems II*. Even if GDPR and the Medical Device Regulation (MDR) were expected to help shape what lawful AI would be, it has turned out that the much anticipated Schrems III will have much more impact on AI. Schrems III is expected to explicitly address current deficiencies in legislation, as quoted above.

The UNESCO document, still in draft (dated 14 Sept 2021), is well researched but the policy area Health and Social Well-Being holds some unintended comical passages, like:[28]

> 123. Member States should pay particular attention in regulating prediction, detection and treatment solutions for health care in AI applications by:
> (a) ensuring oversight to minimize and mitigate bias;
> (b) ensuring that the professional, the patient, caregiver or service user is included as a "domain expert" in the team in all relevant steps when developing the algorithms; ...

[28] It seems to me that in 123(b), some of the rightful stakeholders are included, but for at least the patient and the service user sitting through algorithm development would be a pain, not mentioning the pressure of acting as a "domain expert" in that context. As for 124, I parse the sentence to mean that AI systems could propagate the spread of anti-social information, negatively affecting the mental health of whomever is listening, but the semantic path linking this interpretation to trafficking remains a mystery to me.

> 124. Member States should establish research on the effects and regulation of potential harms to mental health related to AI systems, such as higher degrees of depression, anxiety, social isolation, developing addiction, trafficking, radicalization and misinformation, among others.

Later (paragraph 129), however, an important point about long-term use is made. In particular, norm changes over time is understudied in AI research in general. It is key to acceptance and trust, and therefore bars deployment if not well understood.

> Member States should encourage and promote collaborative research into the effects of long-term interaction of people with AI systems, paying particular attention to the psychological and cognitive impact that these systems can have on children and young people. This should be done using multiple norms, principles, protocols, disciplinary approaches, and assessment of the modification of behaviours and habits, as well as careful evaluation of the downstream cultural and societal impacts. Furthermore, Member States should encourage research on the effect of AI technologies on health system performance and health outcomes.

For AI@KI, the most relevant and useful document by far is the AI Act. For a start, it holds a useful classification of AI techniques and approaches (ANNEX I):

> (a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems; (c) Statistical approaches, Bayesian estimation, search and optimization methods.

The 14 months that passed after the white paper on AI was published in February 2020 saw a sharp increase in the use of astonishingly large deep learning models, so-called transformers (sometimes referred to as *foundation models)*.[29] While transformers are costly to train and maintain, they are reusable and can improve their usefulness over time, in many domains. Because they learn, they also potentially change their predictions and recommendations over time, making them harder to regulate than traditional machine learning models. The AI Act does not explicitly discuss transformers, but "deep fakes" are mentioned several times as in need of oversight. A transformer can—for good as well as for evil purposes—generate synthetic but credible images or documents, such as a complete electronic health record of a non-existing person. If the purpose is to fool the viewer into believing the fake is real, deep fake detection software can be used to reveal this fact. If provenance and veracity is important and the data is valuable, this could lead to an arms race

[29] While such models were developed for semantic understanding of text, they have shown promise in imaging and are now proposed for multimodal prediction. Recently, DeepMind presented an architecture that purportedly "advances genetic research by improving the ability to predict how DNA sequence influences gene expression" using transformer models.

of sorts in the future, between generating and detecting intelligent software. For this reason, the transformer-related software developed by AI Sweden will not be made available as open source, but only through a third-party service (i.e., an API).[30] Unlike the UNESCO document, the act gives surprisingly little attention to the most sensitive data there is, namely health data. This is particular, since health is identified in the act as a high-impact area for AI. That said, its 28th paragraph does state that[31]

> ...in the health sector where the stakes for life and health are particularly high, increasingly sophisticated diagnostics systems and systems supporting human decisions should be reliable and accurate. The extent of the adverse impact caused by the AI system on the fundamental rights protected by the Charter is of particular relevance when classifying an AI system as high-risk. Those rights include the right to human dignity, respect for private and family life, protection of personal data, freedom of expression and information, freedom of assembly and of association, and non-discrimination, consumer protection, workers' rights, rights of persons with disabilities, right to an effective remedy and to a fair trial, right of defence and the presumption of innocence, right to good administration.

[30] The 3.5 billion parameter GPT-SWE model for generating text in Swedish was presented by AI Sweden in December, 2021. Since KI is a member of AI Sweden, this unique model is accessible for research experiments at KI, as is in some sense the Swedish Medical Language Data Lab. At KTH, I have an industrial Ph D student employed at RISE, Evangelia Gogoulou, working on this model.

[31] Personally, I really like the last point made. I always felt bureaucracy was the killer app for AI. No, seriously.

4 *The KI Ecosystem*

Contrary to what has been claimed in many a research strategy or project proposal, one does not build ecosystems, they emerge. Thus, KI lives and thrives in an ecosystem in which KI people can control only smaller and local parts. The funding agencies and other benefactors likewise cannot control how this ecosystem evolves, but they can nudge people in certain directions. In 2015, the Knut and Alice Wallenberg Foundation launched a ten-year grant program initially funded by SEK 1.3B and later substantially increased: the Wallenberg Autonomous Systems and Software Program (WASP) is now up to 5.5B until the year 2030, with over 300 Ph D students (of which I have one). In 2020, that same foundation put SEK3.1B into Data-Driven Life Sciences (DDLS) over the next 12 years. The WASP main programme is also directly related to the Wallenberg Initiative on Networks and Quantum information (WINQ), with many AI connections and with some innovative applications within the life sciences. Possibly as a consequence of political efforts to increase the level of digitalisation in Sweden, politicians have in the last few years asked for more AI research and development. That political goal has been met partly by money ear-marked for AI research and innovation. The following partial list includes efforts that have bearing on health and medicine and which all involve AI research:

- SciLifeLab has a geographical and conceptual adjacency to almost all the work going on at KI, conducting research in clinical genomics and proteomics and providing technical services and assistance, with staff that include employees at KI and surrounding universities[32]

- DDLS was launched with its four priority areas (i) cell and molecular biology, (ii) evolution and biodiversity, (iii) precision medicine and diagnostics, and (iv) epidemiology and infection biology, the last two of which are already engaging KI and SciLifeLab researchers

- WASP split into Autonomous Systems and Software (WASP-AS) and WASP-AI, with the latter having two parts: (i) Machine Learning and (ii) Mathematical Foundations of AI

- WASP-HS for the humanities and the social sciences was launched, with several initial projects devoted to ethics for data processing by humans or machines[33]

- The national Strategic Innovation Programmes (SIPs) got additional funds for AI activities, spawning a range of small AI projects in 2019-20[34]

- AI Sweden was started in 2019, boosted by a SEK 100M grant from Vinnova for 2020-24, with a 2020 addition of a Stockholm node directed towards climate and health[35]

- EIT Health counts KI among its members, and several individuals at KI has had a great impact on developments; among the innovation projects and many KIC-led activities, the *Transforming healthcare with AI* Hub is particularly noteworthy

- AIMES: Center for Advancement of Integrated Medical and Engineering Sciences was inaugurated in September 2020 as a collaborative effort by KI and KTH to promote interdisciplinary research and its translation to societal use

- MedTechLabs is run by KI, KTH and Region Stockholm as a centre for medical technology research, imaging in particular, with the mission of providing patients with faster diagnosis and better treatment

- The Centre for Bioinformatics and Biostatistics (CBB) is a recently launched virtual centre, based in Campus Flemingsberg, for such methods, which are increasingly used in pre-clinical and clinical research[36]

[32] I had the pleasure of co-organising the 6th and 7th KI/SciLifeLab/RIKEN symposia which produced a White Paper on reference datasets for biomedical research with AI methods, included in the digital repository attached to this report, see the appendix. In these symposia it became evident to me how much could be achieved by leveraging on the cross-cultural interdisciplinary tripod that the three organisations have mutually constructed over the last decade or so.

[33] One such project is particularly directed towards the role of AI in establishing new scientific results in biology and medicine: *The new scientific revolution? AI and big data in biomedicine* with Francis Lee from Chalmers University as PI. Francis co-led with me a roundtable discussion on March 25, 2021 that led to a paper in preparation titled *Data work in biomedical AI: the hidden challenges of data, pretraining, and ground truths*.

[34] One of these SIPs is SWElife, which started in 2014 and "supports collaboration within academia, industry and healthcare, with the goal to strengthen Life Science in Sweden and to improve public health." They are currently running a project to capture Swedish activities with AI for the life sciences.

[35] KI is a partner, represented on the research side by myself and Sabine Koch (KI/LIME). With SKR (the Swedish Association of Local Authorities and Regions), AI Sweden started an AI network in the autumn of 2021. There are other national membership organisations that directly affect AI-interested researchers at KI through their activities, such as Forska!Sverige (Research!Sweden) and the strategic innovation programs, but AI Sweden is the only national fully AI-centred effort.

[36] "Understanding and managing the large-scale data that today's research often creates is a major challenge. CBB offers a holistic perspective with the aim of giving more researchers and research groups their own knowledge in the field," says Carsten Daub, CBB director. Carsten is an example of a researcher that actively involves machine learning in his own studies, see the applied AI examples in the appendix.

The above constellations support innovation to a varying extent and many KI researchers could rightfully be called innovators. As a case in point, Johan Lundin is a successful researcher and innovator in AI for image analysis in geographically distributed locations. His point-of-care solutions for resource-limited settings (Figure 4) has moved his work to field use in developing countries, like Kenya and Tanzania, where there is a severe shortage of pathologists and medical equipment. With his team, Johan has since 2016 worked there on malaria, helminth and cervical cancer diagnostics. This is a case of early adoption, in the sense that as late as April 2018, FDA approved the first AI diagnostics tool in the field of medicine (IDx-DR, from Digital Diagnostics). At KI, Johan is a professor of Medical Technology at the department of Global Public Health. He is also a research director at the Institute for Molecular Medicine Finland, at the university of Helsinki, from where he came to Stockholm first as a guest professor. Johan co-founded Aiforia, a company based in Helsinki and Boston, with a staff of 40.[37] The company engagement has allowed for concrete tools to be developed, like the WebMicroscope Platform for cloud-based microscopy, making possible mobile microscopes for cancer diagnostics. Johan has one of the best track records of AI research publications in all of KI, with many application- and use-oriented publications, but also more technical AI-related papers. In the next section, the scientometry of these efforts will be visualised.

[37] The company just released news of a collaboration on digital pathology and deep learning with the Mayo Clinic.

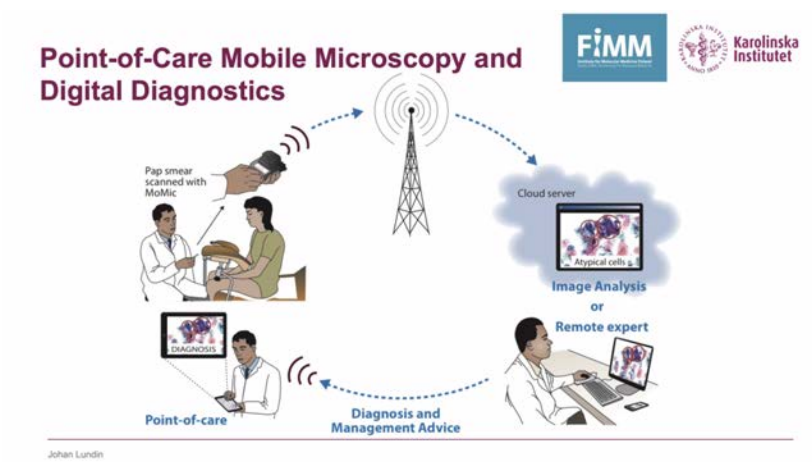


Figure 4: From Johan's Falafel seminar presentation on 23 April 2021, the mobile microscopy and digital diagnostics as envisioned and now financed by the Erling-Persson family foundation in the project *Artificial intelligence for diagnostics of cancer and infectious diseases in resource-limited settings - the MoMic Project*. Tools are meant to constitute end-to-end solutions for Kenya and other countries, in a five-year perspective. Applications that involve AI also include breast cancer target-seeking treatments and pneumonia diagnoses.

Fredrik Strand has like Johan used company structures to make full deployment of AI software, with all of the extra costs that this entails, possible. Fredrik is a PI at Oncology-Pathology/KI and a breast radiologist at Karolinska. Since 2017, he has had a close collaboration with Kevin Smith and Hossein Azizpour at KTH and SciLifeLab. In 2021, I was the examiner of a KTH bachelor thesis on technology acceptance of a large-scale AI-aided detection system for mammography in the breast screening unit at Capio S:t Göran. In addition to the two radiologists that assess each screening mammogram, AI was introduced as a third reader, whereafter the mammogram was passed on to a consensus discussion if cancer was suspected by any of the three. Fredrik's group had previously published two retrospective studies on the possible impact of AI implementation in screening. One study showed that an AI system could on its own assess up to 60 per cent of the mammograms without missing any screen-detected cancer[38] A study by researchers outside KI had shown that if an AI system could serve as an independent second reader of mammograms, such an approach would reduce the workload of radiologists by 44 per cent and also reduce the number of false positives.[39] In addition, Fredrik had led a study with the following key points.[40]

> **Question** Are there currently commercially available artificial intelligence (AI) algorithms that perform as well as or above the level of radiologists in mammography screening assessment?
>
> **Findings** In this case-control study that included 8805 women, 1 of the 3 externally evaluated AI computer-aided detection algorithms was more accurate than first-reader radiologists in assessing screening mammograms. However, the highest number of cases positive for breast cancer was detected by combining this best algorithm with first-reader radiologists.
>
> **Meaning** One commercially available AI algorithm performed independent reading of screening mammograms with sufficient diagnostic performance to act as an independent reader in prospective clinical studies.

This study had several aspects worth commenting on from the AI perspective. First, one of the limitations were cited as: "A weakness of our study is that the AI CAD algorithms did not consider prior mammograms, hormonal medication, or breast symptoms—which puts AI CAD algorithms at a disadvantage compared with radiologists." When measuring performance, here done in terms of AUC, researchers tend to attach much importance to quantitative differences. Conclusions then might be about the prospects of replacing radiologists by a learning algorithm, for instance. But the background knowledge and the gold standard of evaluation is often assumed possible for AI algorithms to somehow pick up; magically, or through osmosis. While there are numerous ways for people to be treated unfairly by algorithms, the reverse situation sometimes manifests too. I also asked Fredrik why he picked commercial AI algorithms, rather than open source and published research algorithms. He replied that it was simply a matter of algorithm performance and system quality, which surprised me but is nevertheless certainly true; AI researchers

[38] Dembrower, K., Wåhlin, E., Liu, Y., Salim, M., Smith, K., Lindholm, P., Eklund, M. and Strand, F. (2020) Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study. *The Lancet Digital Health* 2(9):e468-e474.

[39] Details and references are in the thesis: M. Kloub and A. Gerigoorian: A Cross-Sectional Technology Acceptance Study of an AI CAD System in a Breast Screening Unit, TRITA-EECS-EX;2021:265, KTH.

[40] Salim, M., Wåhlin, E., Dembrower, K., Azavedo, E., Foukakis, T., Liu, Y., Smith, K., Eklund, M. and Strand, F. (2020) External evaluation of 3 commercial artificial intelligence algorithms for independent assessment of screening mammograms. *JAMA oncology* 6(10):1581.

like myself would have the opportunity here to attempt to replicate the study findings with purportedly better AI software. The same holds true for a stacked model, which in the study findings was reported not to outperform the best AI algorithm. A voting protocol between the three algorithms were considered, but in AI modelling, we often look at other means to stacking that in some cases are more powerful than voting.

These research challenges notwithstanding, the results of the two retrospective studies allowed for Fredrik's group to start the clinical study ScreenTrust CAD at Capio S:t Göran. Another type of AI implementation is being tested at Karolinska, funded mainly by MedTechLabs, and is based on a pipeline of three AI algorithms developed in collaboration with KTH and one commercial AI algorithm. In the on-going ScreenTrust MRI study, the top 8 per cent of AI-ranked cases of possible false negatives, where the screening radiologists found no reason to suspect cancer, are invited to the study. Half are randomised to have supplementary screening MRI and the other half to the control group with no further examination. Fredrik has also considered amassing case numbers large enough for training deep learning systems to predict breast cancer risk, estimate mammographic sensitivity, and detect tumours, for which more than two million mammography images have already been collected.[41] In the autumn of 2021, RCC and Vinnova funded a project where Fredrik is the scientific leader in the creation of a national validation platform connecting retrospective data from various hospitals with AI algorithms to perform evaluation of accuracy and robustness.

[41] The project MammoAI is investigating deep learning together with experienced KTH researchers Kevin Smith, Hossein Azizpour and Mats Danielsson. Dembrower, K., Lindholm, P. and Strand, F., 2020. A multi-million mammography image dataset and population-based screening cohort for the training and evaluation of deep neural networks—the cohort of screen-aged women (CSAW). *Journal of digital imaging* 33(2): 408-413.

5 *Scientometrics*

I have had the good fortune of collaborating with Peter Sjögårde, using his tools for bibliometric analysis for the benefit of AI@KI.[42] Peter has helped me with community detection by means of grouping KI researchers via their publications, using subject mapping. Peter prepared a list of AI-related keywords, which was then pruned and adjusted by myself and Peter's Ph D supervisor Sabine Koch. It finally consisted of about 300 words, which were used to search PubMed.[43] Every KI researcher involved with one or more topics from that list were mapped out in graphs, based on different criteria on the relationship to other KI researchers. To fully appreciate Peter's tools, one must spend time zooming in and out of graphs, but to give a flavour of the information available, I have included a few static views. The dynamics of such graphs over time I consider an important handle on the strategic development of AI at KI. I would expect the graphs to grow in both size and density as the number of AI-related publi-

[42] Peter has an amazing map of PubMed open sourced on GitHub, but the maps of KI, like the one on AI-related KI topics are naturally extra relevant here.

[43] A commented subject list is part of the digital repository for this report, see the appendix. More technical publications from KI researchers might lie outside PubMed, so finding every single relevant publication is not to be expected.

cations increases. Since publication year is one possible criterion for search, this growth from, say, 2018 to 2028 could inform strategy in many ways, but also form an important part of dissemination efforts: to tell the world how much KI researchers are doing and naturally also about who they cooperate with in the rest of the world.

Figure 5: Interacting with the map, you quickly find for example that Peter Ström in 2020 defended his Ph D thesis on *AI for streamlining prostate cancer diagnosis* and with the group published on *AI for diagnosis and grading of prostate cancer in biopsies* in *Lancet Oncology*. Facts like this are in the map easily found by clicking the name of the researcher and next clicking their PubMed link. One more click and the paper is in your browser: highly recommended as a tool, which naturally allows you more readability than from this screen dump. Note that edge thickness indicates the number of co-authorships.

Peter's *Map of Science for AI research at KI* runs from 2016 until mid-2021 and holds 353 publications. The corresponding map of global AI research based on the NIH Open Citation Collection holds 78519 publications. The less than half a per cent that involves KI researchers contains 144 researchers with two or more AI-related publications. Johan Lundin and Fredrik Strand, highlighted in the previous section, have twelve and nine, respectively, in the time period. The AI paper co-authorship graph for Fredrik Strand (Figure 5) shows a link to Martin Eklund (KI/MEB), also with nine publications. This link constitutes a bridge to a large number of researchers at MEB (and OncPat). Two of these, Mattias Rantalainen and Johan Hartman, were interviewed for AI@KI and they have co-written a most relevant paper, as well as formed a company, *Stratipath*, for "AI-based precision diagnostics to improve cancer treatment decisions and patient outcomes".[44]

[44] Acs, B., Rantalainen, M. and Hartman, J. (2020) Artificial intelligence as the next step towards precision pathology. *Journal of internal medicine* 288(1): 62-81. The company Stratipath is an interesting case story in itself for KI innovations on AI-related technology, and can be compared in some way to that of PathFX, as detailed in Sophie's master thesis (see the appendix).

In this way, the personal and inter-departmental collaborations can be firstly approximated, the second step being interviews where one might find that there is a lot of collaborations that have not yet resulted in PubMed publications. In exploratory mode, the co-author map can also identify prolific AI-interested researchers easily: I am puzzled as to why I have not interviewed someone who did his post-doc at a department where I have myself been a guest researcher

(Erik Westman, at Institute of Psychiatry, King's College London, now at Neurobiology, Care Sciences and Society) and who has twelve publications on the map, for example. We can also note high-impact publications in medical journals, as well as publications on health-related topics in technical journals. In the former category, there are papers relating AI to precision medicine in prestigious journals like *Nature Medicine* ($IF > 53$) and *Lancet Oncology* ($IF > 41$). In the latter category, examples include papers in top engineering conference proceedings, involving KTH and KI researchers.[45]

[45] Honoré, A., Liu, D., Forsberg, D., Coste, K., Herlenius, E., Chatterjee, S. and Skoglund, M. (2020) Hidden Markov Models for sepsis detection in preterm infants. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1130-1134, IEEE. For more on Herlenius lab, see the applied AI examples in the appendix.

For a topic-oriented view of research, we can turn to MeSH terms, 149 of which were included because they were used for at least two publications. Even if this graph (Figure 6) holds many generic terms, I find it immensely useful for its indications of methods, application areas and technologies. And we are looking only at the highest level of abstraction here, zooming in allows for even more careful deliberation on links and dependencies. Clicking a term opens up the PubMed list, just as for author names. We can also unify the two graphs into a blended researcher and term graph, browsing the two in tandem. In more coarse trend terms, the number of AI-related publications at KI has monotonously increased since 2016, and since 2018 by more than 50 per cent per year.

6 *AI Deployment at KI*

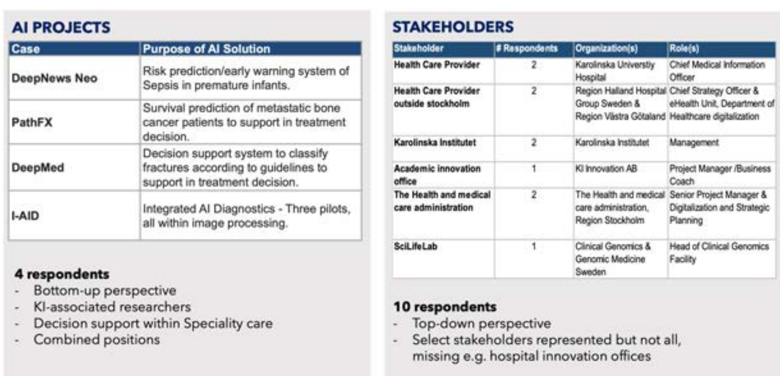
To secure competence with adequate skills long-term is a global issue for AI deployment, the so-called talent problem (cf. Figure 2). The light in the tunnel for KI when it comes to the talent problem is that an increasing number of data scientists and data wranglers are considering health as the top application area, more important than money or free soda from the company fridge. Besides its obvious uses and impact, some people find health applications attracting them for the same reason as others find it repelling them, namely data sensitivity. Federated architectures, data security and privacy, data retention, judicial aspects of data sharing are examples of IT issues that have special weight within health. With health informatics professor Sabine Koch, I will myself participate in a new project from 2022 on scalable federated learning, in which KI and Karolinska are data providers and stakeholders.[46] Before considering AI technology and the human competence required to make it useful, however, clinicians need to know it is there. Staying informed on AI developments is a different thing entirely from keeping abreast on medical developments, and the overlap is still very small. Successful AI applications in research projects that do not affect treatment of patients is likewise a global problem, in no way specific to KI. When

[46] VR 2022-25, with PI Dejan Kostic, professor at KTH.

Figure 6: The 149 MeSH terms that fulfilled the inclusion criterion of having been used in at least two publications.

AI@KI started, I was at the end of two large research projects at KI led by professor Viktor Kaldo at Clinical Neuroscience and the Internet Psychiatry unit at Psykiatri Sydväst, Region Stockholm. We had planned for and implemented AI supporting adaptive treatment, to be subjected to an RCT in 2021 and then rolled out, if the trial (still ongoing) proved successful. That meant that I had learned of many barriers to AI full deployment personally, but also that I—thanks to a tremendous team effort—could see that full deployment is possible. The global need for a solution to the problem and my personal experience of chiefly local obstacles and opportunities led to a dedicated effort within AI@KI to address this risk of failure. Practical use of AI in the clinic and the problem of how to best inform clinicians on AI are each given a subsection below.

6.1 From Successful Pilot to Deployment



AI PROJECTS

Case	Purpose of AI Solution
DeepNews Neo	Risk prediction/early warning system of Sepsis in premature infants.
PathFX	Survival prediction of metastatic bone cancer patients to support in treatment decision.
DeepMed	Decision support system to classify fractures according to guidelines to support in treatment decision.
I-AID	Integrated AI Diagnostics - Three pilots, all within image processing.

4 respondents
- Bottom-up perspective
- KI-associated researchers
- Decision support within Speciality care
- Combined positions

STAKEHOLDERS

Stakeholder	# Respondents	Organization(s)	Role(s)
Health Care Provider	2	Karolinska Universtiy Hospital	Chief Medical Information Officer
Health Care Provider outside stockholm	2	Region Halland Hospital Group Sweden & Region Västra Götaland	Chief Strategy Officer & eHealth Unit, Department of Healthcare digitalization
Karolinska Institutet	2	Karolinska Institutet	Management
Academic innovation office	1	KI Innovation AB	Project Manager /Business Coach
The Health and medical care administration	2	The Health and medical care administration, Region Stockholm	Senior Project Manager & Digitalization and Strategic Planning
SciLifeLab	1	Clinical Genomics & Genomic Medicine Sweden	Head of Clinical Genomics Facility

10 respondents
- Top-down perspective
- Select stakeholders represented but not all, missing e.g. hospital innovation offices

Figure 7: The four bottom-up (innovator) activities at left name projects otherwise mentioned in this report. For example, DeepNews Neo was covered in the half-time version of this report and that text is available in the appendix. The ten top-down (decision maker) respondents at right cover many parts of the stakeholder map, and even parts excluded here like the innovation offices have affected (and have been affected by) AI@KI, via various interactions in the last two years.

In the first half of 2021, Sophie Monsén Lerenius conducted 14 semi-structured interviews within AI@KI, covering important projects with AI elements, and many key stakeholders (Figure 7).[47] Among the four AI projects, PathFX—a system for survival prediction of metastatic bone cancer patients to support in treatment decision making—was deemed most mature. It is integrated into the care plan of the Orthopaedic unit at Karolinska and its a CE-marked system at Technological Readiness Level TRL8. Sophie embraced the almost impossible task of mapping out the entities involved with interplay between academia and healthcare in Stockholm (Figure 8), in keeping with the ambition for AI@KI to cover as much as possible of the KI ecosystem. Such an illustration could be made almost arbitrarily complex and we will return to this mapping later in the section that follows.

[47] Images in this subsection are adapted from slides made by Sophie and in some cases occurring in her master thesis in health informatics, supervised by Sabine Koch. The full thesis is part of the digital repository for this report.

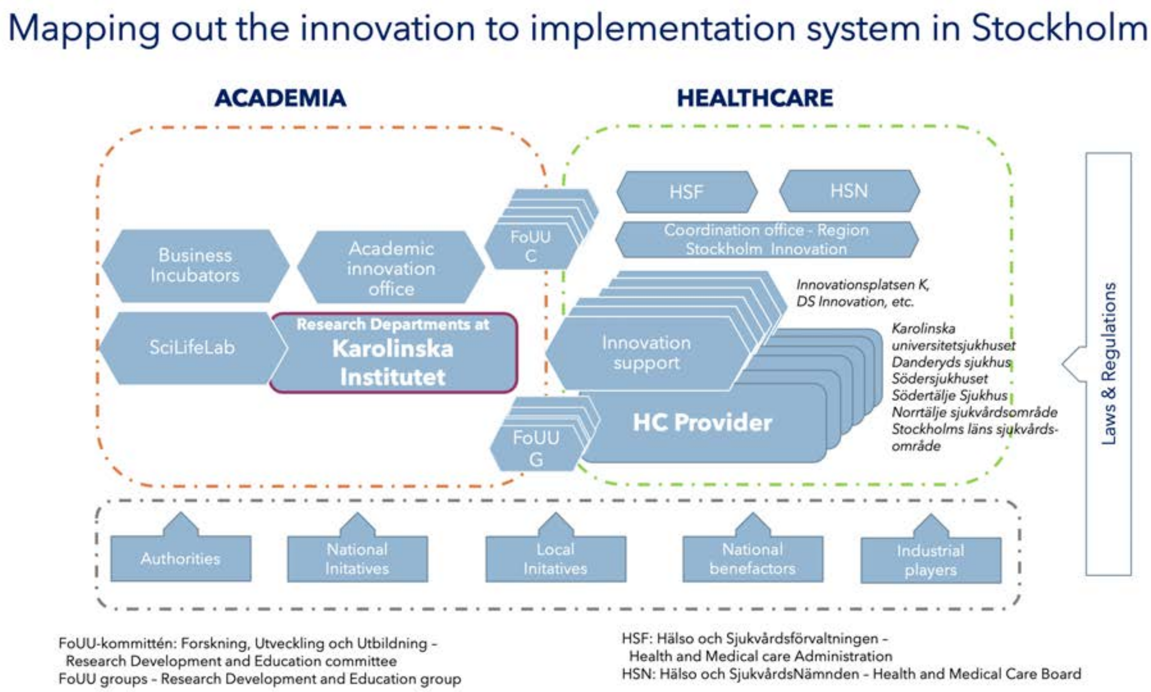


Figure 8: Some of the entities potentially involved when a researcher at KI implements an innovative idea related to AI.

A gap between research projects and the reality of using AI in clinical practice in Sweden was identified, reinforcing challenges associated with the implementation process. The uptake of technical innovations related to AI were explained in terms of barriers and facilitators. For Swedish circumstances, Sophie found that there is insufficient knowledge about the opportunities that AI could bring to Swedish healthcare, and also that the vast amount of promising AI projects underlines the importance of an in depth and consolidated understanding of the challenges and facilitators in these projects. In a directed qualitative content analysis, she identified five themes, with their barriers (Figure 9) and facilitators (Figure 10).

From the interviews, examples were collected that illustrated the importance of the broad set of competences needed both to develop but more importantly also to implement an AI solution. Below is a list of the skills mentioned in interviews as crucial.

- Medicine – an understanding of the need and/or potential
- Technical – basics of AI
- Implementation – experience of change management and leadership
- Legal and regulatory – MDR, Regulation on in vitro diagnostic medical devices (IVDR), CE marking, intellectual property, etc.

- Ethics and data security – GDPR, data retention, etc.
- Tenders – process and tactics of The Public Procurement Act
- IT Infrastructure – strategy, platforms, etc.
- IT maintenance – DevOps, i.e. software development (Dev) and IT operations (Ops)
- Informatics - standards, terminology, data hygiene, etc.
- Intelligence – latest developments and other solutions on the market
- Health Economics – evaluate and propose value propositions

Key barriers within each identified theme

Data & Informatics	• Insufficient IT infrastructures and absence of technical innovation environments • Low data maturity; lack of data governance structures and general data hygiene
Business Model	• Difficult to develop sustainable business models
Culture & Competence	• Broad set of competences required • Reimbursement model counteract incentives to innovate • Decentralization and a general fear of "doing wrong"
Innovation to implementation process	• Complex structures for innovation, clinical implementation and collaboration • Unclear process for clinical implementation after pilot phase • Insufficient financing of the clinical research and implementation phase
Regulatory & Legal	• General ambiguity on how to interpret regulations • Requirements of The Public Procurement Act (LOU) and new Medical Device Regulation (MDR)

Figure 9: Five common themes emerged from interviews. Key barriers for the clinical implementation of AI in healthcare in Stockholm were identified within each theme, from both a case and stakeholder perspective.

Key facilitators within each identified theme

Data & Informatics	• Increased collaboration with regards to health data • Focus on IT infrastructure for development and implementation
Business Model	• New co-creation models evolving (Industry – Academia – Health care)
Culture & Competence	• Increased focus on innovation and the development of care • Interdisciplinary skills and collaborations • Cultural shift towards data maturity
Innovation to implementation process	• A more structured and systematic way of working together with innovation • Strengthened support and governance • Strengthened collaborations between academia, health care and industry
Regulatory & Legal	• The future will bring more clarity on regulations • Opportunity to join forces on policy issues

Figure 10: Key facilitators and enablers to accelerate clinical implementation of AI in healthcare.

To unify healthcare and academic research means to investigate data from its primary and secondary use, respectively, and in tandem. In this way, a joint strategic path can be mapped out, given that

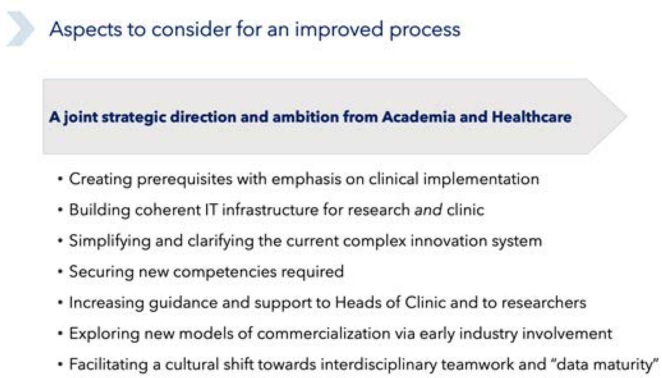


Figure 11: Seven verbs for seven aspects to cover on the strategic path. It is not enough for researchers to (bottom-up) demonstrate the efficiency and usefulness of AI models or to innovate by introducing AI in their information flows. Such organic growth and refinement needs a certain amount of (top-down) leadership, coordination and follow-up.

a number of aspects that Sophie painstakingly identified are taken into consideration (see Figure 11). After Sophie completed her work in AI@KI, one of the project leaders she interviewed, Max Gordon at the orthopaedic unit at Danderyd university hospital, saw a future he, with his colleagues, had envisaged for quite some time start to come true. Max co-founded DeepMed AB in 2016 to help take innovative uses of deep learning for fracture classification to orthopaedic decision making and had made the journey from clinical problem to a deployed clinical application within the hospital environment. At a site visit with Sophie, we also spoke to Max about the desire to continue on the path of rolling out clinical AI applications. It was therefore a pleasure to learn of Clinical Artificial Intelligence Laboratory (CAIR-Lab), a new innovation laboratory in which is used the experience from the orthopaedic AI application to aid various other specialties building their own clinical applications based on AI. The three key elements of CAIR-Lab are:

1. Description of the clinical problem
2. Relevant patient data
3. Technical knowledge

During the course of AI@KI, I have often been asked to suggest problems that AI techniques could solve. This I have declined, stressing the importance of the problem originating from the environment and people populating it in which the solution would be deployed. I have also noted the difficulty with which those same people presented patient data that were useful to AI algorithms. This is most easily explained by noting that humans and machines learn in conceptually different ways, and therefore data projections that well meaning non-experts have curated are not always ideal. For example, structured text is under some circumstances less useful to learning machines than free text is. It is therefore joyful to see CAIR-Lab described as supporting these elements. With the elements identified by Sophie, we begin to approach a recipe for counteracting that great risk of AI development, to get stuck at the level of successful pilot. The other projects scrutinised by Sophie are also very informative in this light, and the details on those are in her thesis, see the appendix.

6.2 *Machine Learning for Clinicians*

Some of the recommendations from Sophie's work can be met by looking at AI as yet another methodology, a form of modern statistics, with some explicit learning representations and models. For AI to be put to use at the clinic, there are then lessons to be learned from how statistics—and biostatistics in particular—and health informatics have historically proved useful. Clinicum is an entity within which such studies could take place, even if it is still under development rather than in place. Sophie chose to map it out in her discussion on these matters, see Figure 12, where she related the local initiatives to national ones, and where KI got related to the region.

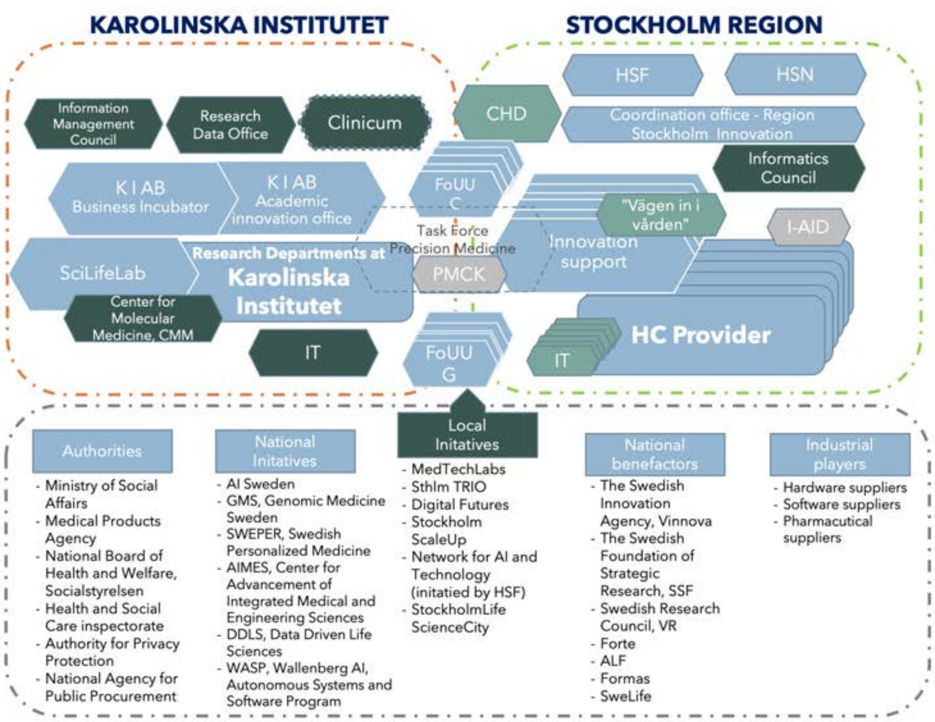


Figure 12: Even more organisational complexity in this illustration, where Clinicum is a possible KI unit at left. Details can be found in Sophie's thesis. We will return to the precision medicine initiatives at centre, shared between KI and the hospital in the section that follows.

With Sandra Eloranta, biostatistician at KI/KEP and also involved with the development of Clinicum, I have been discussing methodology in general and the relation between biostatistics and AI in particular. We decided to write a guide on machine learning for clinicians together, thinking such a guide could be a means to both bottom-up and top-down efforts.[48] We have attempted to explain the differences between statistical inference and machine learning from a methodological standpoint. Even if the guide has a potential general readership, we both draw heavily on our KI experiences: Sandra wrote a deep dive on predicting survival outcomes after aggressive lymphoma and I wrote one on precision psychiatry.

[48] Forthcoming publication as: S. Eloranta and M. Boman (2022) Predictive models for clinical decision making: Deep dives in practical machine learning. Ask either of us if you would like to read a near-finished draft.

7 *AI and Precision Medicine*

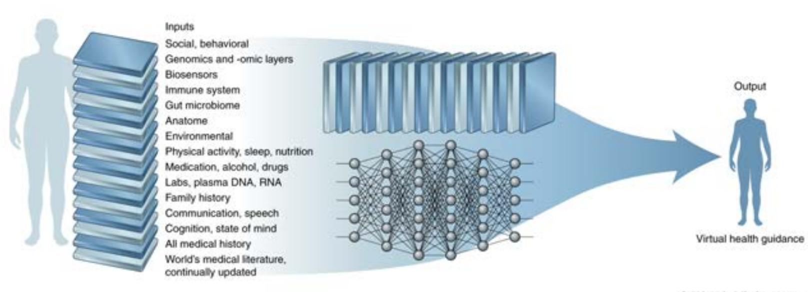
A NUMBER OF FACTORS CONTRIBUTE to the timeliness of rolling out precision medicine (PM) as a key component in healthcare. The number of treatments and therapies offered is increasing and combining them correctly is a complex problem. Increased costs for some treatments are also increasing the risk of unfair care, especially in the light of tight budgets and frequent lack of resources. Because of the cross-disciplinary competence required to adequately address such problems, cross-cultural collaborations are necessary, including tight coupling between research and the employment of its results at the university hospitals. But the crossing of cultures also necessitates combining somatic care with genomics, proteomics and pathology. The increased size and number of health-related databases is an enabler for this to happen. Moreover, in Sweden's National Life Sciences Strategy, an objective is to pioneer the introduction of PM in healthcare.[49]

[49] While there is no mention of PM in the Region Stockholm strategy on IT and digitalisation for 2020-23, AI is mentioned as a technology for increasing accessibility, efficiency and quality.

When Big Data was first introduced, medicine was named a field of application in which huge datasets were ubiquitous. Big Data was the key to turning PM into clinical medicine. Biobanks, image repositories and large collections of video material were among the sources supposed to fuel pipelines for big data analytics. The increasing use of personal monitoring and sensor tracking fuelling the medical Internet of Things, such as mHealth, eHealth and wearable technologies would then seamlessly add data on 'digital biomarkers' over time. This would happen via apps and via collecting other digital traces of individual activity. Because of the volumes, only AI could process the data and so terms like 'intelligent health data analytics' came about. When field-tested, designing AI pipelines for health data turned out to be much harder than first expected, however. Almost all the data is unstructured and requires extensive pre-processing to become useful downstream. Issues concerning privacy also called for attention, leading to lots of computer scientists concentrating on technical matters like pseudonymization, synthetic data, data encoding and decoding, transparency, and last but not least information security. While important, such issues are not at the heart of what Big Data had promised to deliver to medicine, namely new and important correlations and causal relationships in health data, out of human reach due to their complexity. AI tools had similarly promised to automatically deliver results from data-driven methods that would be so-called non-SQL: the innovative next step after relational databases having successfully been applied to data lakes and data warehouses.

What happened instead was an intense focus on the individual, leveraging on health analytics by indexing relevance in huge data sets on individual health profiling. My favourite term for this is $n = 1$ *medicine*. If a vast space of unstructured data is trawled for every data point relative to the social security number—diagnoses, anamneses or observed values of an individual—we can go from population statistics to customised health advice and care to a single person. This works because the relevant dataset is shrunk in volume, making AI methods feasible and even easy to employ. It also sails past the privacy barrier, because we can focus on one individual, and possibly some relatives and some environmental data, and everything we investigate is prompted by current or future health issues in this individual. As per usual, when your own health or that of your loved ones is at risk, privacy goes out the window. This paved the way for digital phenotyping, a key step in achieving so-called *P4 medicine*: personalised, predictive, preventive and participatory modern medicine. I have argued, with colleagues that apply AI methods to psychiatry, for a fifth P for 'Psychological' to be added,[50] making it to P5 medicine.

[50] See Boman, M and S Velupillai (2021) P5 Medicine and Slowfood AI: Data Science and Mental Health, *Medium*, 12 Jan 2021.



A virtual medical coach that uses comprehensive input from an individual that is deep learned to provide recommendations for preserving the person's health.

Figure 13: In an excellent article in *Nature Medicine* from which this illustration was taken, Eric Topol fleshes out the dream of how deep learning and other AI techniques can help realise PM.

At KI, the PM task force, led by Anna Martling, includes the development of new precision diagnostics.[51] Another part deals with the data, while a third area concerns a virtual centre—Precision Medicine Centre Karolinska (PMCK)—established to support care in practice. The task force also has parts dealing with prevention, biobanking and

[51] The blog entry by the KI President (Nov 12, 2020, in Swedish) provides background to this collaborative effort with Region Stockholm. I was very happy to join the task force on February 1, 2021.

industry connections.

Because PM must be about reducing the health burden for our population, people in the largest disease categories must also be helped by it. This can be achieved in two ways. First, small and highly specified cohorts of patients can be considered by clustering individual cases. Second, by considering the 90 per cent or so of the PM process flows that are near identical for all diseases, we can leverage on success stories. Some diagnoses for which PM should be able to deliver important means to reduce human suffering and societal costs, with AI playing a part, include:[52]

- Cancer (precision oncology)
- Hematology
- Rare diseases
- Multiple sclerosis
- Mental health (precision psychiatry)

All involve the customisation of drugs and treatment. If the target is an individual, we could refer to an organ. If we talk about the coating of a pill instead, we could generalise to small subpopulations, keeping the precision qualities intact. Understanding the etiology of a disease through molecular epidemiology could then become reality, mixing macro- with micro-methods (or meso-methods if we consider subpopulations). There are also differences between classes of diagnoses when it comes to the feasibility of AI. For rare diseases, to be able to compare a patient to historical data on similar patients and their subsequent diagnoses could provide important decision support functions.

For cancer, imaging has so far been more fruitful, but the most promising avenue for also involving AI seems to be molecular tumour boards. The BioLymph project at KI, led by Karin Ekström-Smedby, is chairing meetings discussing retrospective lymphoma cases, with an eye on a PM-oriented future where such meetings could be part of clinical reality to the benefit of patients under current investigation. This is directly linked to diagnostic development within the PM task force, because what is considered covers assays designed for comprehensive molecular profiling.[53] A molecular tumour board meeting is an incredibly interdisciplinary affair already, and a clinical decision support system to be considered could use AI methods for multimodal fusion, but the sheer information overload could also benefit from more classical expert system augmentation. With gene aberrations as a tool for patient stratification, AI methods for multimodal fusion could then be employed within interesting

[52] In a recent agreement between the Ministry of Health and Social Affairs (Socialdepartementet) and SKR on fair and efficient treatment of cancer patients, AI is mentioned in two contexts. For prevention and early discovery (Section 4.1), image diagnostics using AI is high-lighted, and as part of a research and competence discussion for regional cancer centres (Section 6.2.2) AI is cited as worthy of support. Such links between PM and AI are noteworthy, and potentially of great future importance.

[53] At KI, Richard Rosenquist is leading such efforts, combining genomic and epigenomic characterisation using high-throughput technologies with state-of-the-art evidence-based documentations of the vast space of variables in such characterisations, for the individual patient.

subpopulations, making this a candidate forerunner for precision health (Figure 14).

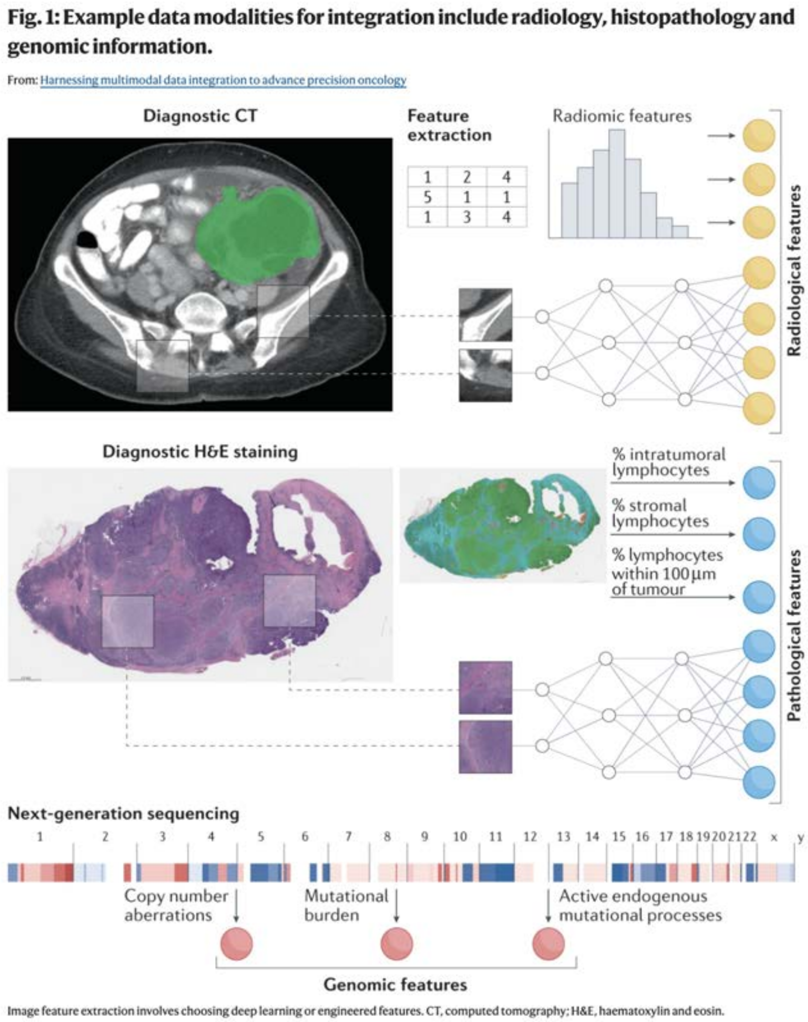


Figure 14: With the help of the scientometrics detailed in Section 5, I could easily find key publications linking AI to PM, in precision pathology, histopathology, gene amplification through tumour morphological features, and many more multimodal joint studies of tissue and disease. Illustration from Boehm, K.M., Khosravi, P., Vanguri, R., Gao, J. and Shah, S.P. (2021) Harnessing multimodal data integration to advance precision oncology. *Nature Reviews Cancer*, pp.1-13. With the help of Karin, Tove Wästerlid and Sandra Eloranta, I have initiated a KI master student investigation into the health informatics aspects required to realise such an approach, which just started. Besides multimodal fusion, deep learning is an option for the modalities illustrated, as indicated at bottom of the figure.

Thomas Frisell (KI/KEP) is leading a new VR project combining AI with PM for studying treatment outcomes in clinical practice targeting multiple sclerosis. The project strategy is described as implementing[54]

> 1) involvement of patients and neurologists in identifying clinically pressing key questions where data from RCTs are not available and is unlikely to ever be sufficient. 2) a nationwide prospective collection of clinical outcomes, and of blood sera, from a large population-based patient cohort, representative of current clinical practice. 3) extensive data added from Sweden's national registries, building the largest and most comprehensive MS DMT cohort in the world. 4) analyzing this rich database with state-of-art clinical epidemiological methods,

54 Quote from the approved application, which also details the important data source of the Combat-MS study (2016-2021).

integrating machine-learning algorithms for clustering and prediction with the causal reasoning and focus on interpretative value in classical epidemiology.

The first point is the general observation that RCTs can not be the only means to clinical trials that pave the way for deployment, simply because of the vast space of therapeutic interventions and underlying environmental factors. Point 2 indicates multimodality and point 3 explains the role of disease modifying therapies (DMTs) for capturing real world evidence. Finally, AI is described as capable of a perspective complementary to causal relations, in effect by analysing associations and correlations. This is a very good example of data-driven and mixed-model modern epidemiology and so serves as yet another illustration of the feasibility of AI for PM.

To achieve long-term and stable positive effects of AI use, learning structures should be saved and re-used. That data is ideally re-used is an old truth, but that contextualised deeper models (like the transformers discussed above) could be generalised to new patients is still in its infancy.[55] If such transfer learning becomes easier to implement thanks to PM efforts, it will not only make diagnostics more efficient, but it will also help predictive models. As a bonus, this mix of PM and AI could help us understand prevention better, and why healthy people stay healthy.

[55] The first paper on transformers has, in spite of being published as recently as October 2018, over 32000 citations. This contextualised model—Google's original BERT model—has since been generalised and seen its usefulness improved for many applications, including biomedical text mining.

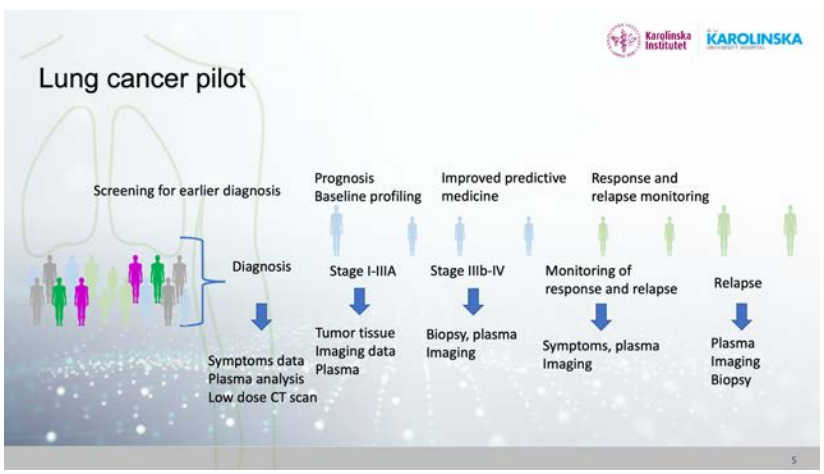


Figure 15: Slight remix of a slide by Janne Lehtiö on the lung cancer pilot, part of diagnostics development within the PM task force. To fuse signals from data that normally is not considered together is a strong point of AI methods and an enabler for precision diagnostics.

Lars E Eriksson is the PI for a three-year project funded by the Sjöberg Foundation that is just starting, *Facilitating Early Diagnosis of Lung Cancer: Transdisciplinary Efforts Combining Data from Patient-Reports, Biomarkers and Imaging*. My own role, in keeping with lung cancer having been chosen as a pilot in the ongoing PM task force, will be to investigate AI methods for multimodal fusion.[56] This builds on already published KI collaborations on symptomatics as a modality for prediction of lung cancer, and I am supervising a KTH

[56] With my former master student Marcos Carbonell, I have developed an open-source solution for fusing and validating multimodal signals. This method is used in several KI projects already, all with publications submitted.

master student in medical technology who is investigating the fusion of symptomatics text, imaging and biomarker modalities.[57]

Besides the AI-relevant entries listed earlier in this report (Section 4) as belonging to the KI ecosystem, the following entities play a part for the unification of PM and AI.

- Centrum för hälsodata (Stockholm Center for Health Data) was founded in 2019 to become a one-stop shop for supplying data for research purposes and is part of Region Stockholm. The results of data use should lead to better prevention, diagnostics and treatment, and fair care should be strived for; all in keeping with the goals of PM. The centre also has a mission to work with methodological development for healthcare data and to partner with industry. KI is the organisation with the most requests for Take Care data, numbering 30, but the centre has struggled with legal issues surrounding the EHRs and the external partners in statistics and data delivery.[58]
- Genomiskt medicincentrum Karolinska (GMCK) is a part of GMC Sweden[59] in which universities and hospitals collaborate with Clinical Genomics Stockholm at SciLifeLab and Karolinska Universitetslaboratoriet (KUL) on the hospital side. At the hospital, three pilots are running in cooperation with Tema Cancer and GMCK: lymphoma, breast cancer and lung cancer. In the ongoing clinical prospective Biolymph study, the cancer epidemiology group (KI/KEP) works closely with the medical units of hematology, genetics and pathology at Karolinska where truly inspiring research advances are being made in molecular tumour boards and other forms of multimodal processing of data in which AI could play a part.
- Stockholm Medical Image Laboratory and Education (SMILE) is a core facility at KI and the university hospital. It acts as a meeting platform and a translational hub, with research and development strongly tied to medtech companies. SMILE is connected to the Clinicum proposal and to the Flemingsberg Medical Imaging Facilities.[60]
- The Center for Bioinformatics and Biostatistics (CBB) opened at campus Flemingsberg in 2021, with the aim of sharing knowledge about the use and understanding of bioinformatics and biostatistics research methods. There is much overlap between such research methods and those of AI, as explicated in the guide to machine learning for clinicians mentioned in Section 6.2.

[57] Levitsky, A., Pernemalm, M., Bernhardson, B.M., Forshed, J., Kölbeck, K., Olin, M., Henriksson, R., Lehtiö, J., Tishelman, C. and Eriksson, L.E., 2019. Early symptoms and sensations as predictors of lung cancer: a machine learning multivariate model. Scientific reports, 9(1), pp.1-11.

[58] As of Dec 2021, the average time for data access was 329 days for Take Care data, an unrealistic waiting time for some research studies. Clara Hellner, Director of Research and Innovation at Region Stockholm, interviewed in Läkartidningen in Jan 2022, stated that 36 out of about 90 applications are completed, and that AI-related requests are extra complicated due to the large-size data sets asked for.

[59] The strategic plan for Genomic Medicine Sweden states the following (p.2). *A common platform for genomics data is being built as a national, scalable system that will be transferable to other areas and for the benefit of the entire life science ecosystem, thereby strengthening research, development and innovation. This data platform will be designed to be able to utilise healthcare data using powerful AI-based analytical tools.*

[60] Birgitta Janerot Sjöberg who is responsible for SMILE was also the project leader for *I-AID: Integrated AI Diagnostics*, a Vinnova project that ran from 2017-20, with the purpose of accelerating the use of AI in Healthcare.

8 *The Way Forward*

In order to correctly analyse AI activities strategically, it is necessary to engage and attempt to synchronise with all relevant strategic efforts within the ecosystem at KI. The most relevant ongoing strategic effort is the precision medicine initiative joint between KI and Karolinska. For AI systems to assist with in precision diagnostics and care, it is crucial to understand and respect how systems end up in care and treatment platforms, with or without AI. Going backwards from CE marking, we find successful randomised controlled trials, in turn based on successful research pilots and published peer-reviewed research. There may also be pre-clinical research and results, animal studies, and numerous ways to connect new research findings with established and evidence-based scientific knowledge. No new technology can ignore any of these process flows and since AI has a black box reputation in many circles, such sensemaking is pivotal. Sensemaking efforts on the AI side in turn allows for more visibility to KI, nationally as well as internationally, by demonstrating through validated approaches which success stories are out there and how they may be generalised to other parts of the world. In practice, all therapeutic options can not realistically be evaluated with respect to efficacy in randomised controlled trials. The data-driven, hypothesis-less and exploratory nature of AI methods can here contribute to precision and prevention by looking for associations in real world evidence databases.

In a stack of AI software development layers, most KI researchers are at the top of the stack, looking to see the fruits of their research be realised in a service layer. There is a front-end, usually a service running on a computer, possibly with a clinician as its end-user, accessing the service via a graphical user interface. To make things work front-end, there is a back-end underneath, running some kind of business logic. In the case of AI, this logic layer can be very complex. If different developers are to refine and maintain the logic, an API (Application Programming Interface) needs to hold the logic layer. Under the API is the back-end, with everything related to the communication infrastructure. The back-end connects everything above it to the data, sitting in various databases. At any hospital, it would be surprising if less than four separate databases hold the information needed to reason about a single patient, at any time. It would be equally surprising if in one of the Stockholm hospitals, someone had managed to fuse data from these databases so that it could be smoothly accessed for primary use. Even for secondary use in research, I have yet to see this, but for the goals of precision medicine, it is nevertheless a requirement. In a future perspective, it is therefore necessary to imagine a scenario where this is done, with safety and effectiveness constantly monitored as part of maintenance post-deployment. Some methods, like backcasting, will then allow us to reason backwards to see where the bottlenecks are. The clinical trial before deployment, and the early phases before that. Are the bottlenecks in fact norms and interpretations of regulatory frameworks or are they more crisp and inviolable, for example?

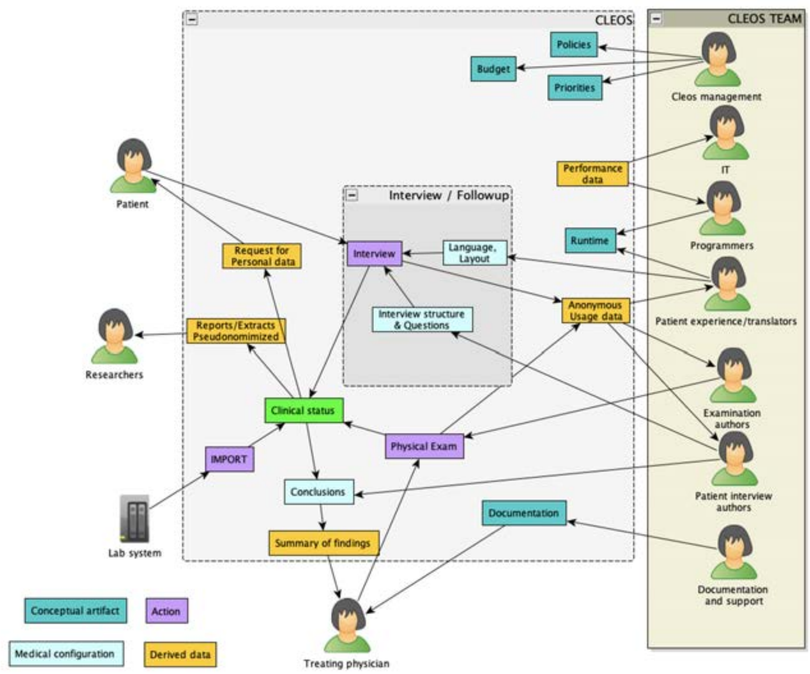


Figure 16: Lars made this slide for a CLEOS team meeting and I include it because it is a fine stakeholder map of a typical project for software developed for health applications. As detailed in the appendix, CLEOS is an expert system, but there is nothing in the map indicating any special AI properties. These come into play only when researchers obtain extracts from CLEOS and use these to suggest refinements or new analytics to the CLEOS team (rightmost panel). Versioning of the software is then used for prototyping or trials, and researchers must take into account the long-term learning of an expert system over time; keeping track of the exact changes to its performance and behaviour is not humanly possible.

There are currently many people at KI in need of support for specific machine learning models and their efficient implementations: the part sitting mostly in the logic layer of the stack. Here, I have been able to help many to some extent, but far from all, and not to a full extent. In some cases, I have been added to applications for research funding, but as my own time is limited, I have also brokered connections to data scientists and machine learning programmers. Two persons working with me on projects at KI that I have become directly involved with are Fehmi Ben Abdesslem (employed at RISE and affiliated with KI) who is currently working with me on lung cancer symptomatics data, and Lars Sjödin (consultant) who is programming the CLEOS system (Figure 16) for research use at Danderyd university hospital, Karolinska, and also planned for use in psychiatry. I am the examiner for two bachelor students from KTH in 2022 who are analysing digital recordings of history for patients with chest pain acquired by the CLEOS application. The main purpose of the CLEOS-CPDS (Clinical Expert Operating System - Chest Pain Danderyd Study) is to investigate the additional value of self-reported computerised history taking as compared to a standard doctor meeting/interview. The goal is to improve triage of patients in the emergency room, and to develop a decision support tool to be tested in a future trial. Predicting the risk of a heart attack will be realised using machine learning methods.

To increase awareness at KI for all things AI-related, new structures and processes for communicating AI results are required. With Peter Sjögårde, I have hopefully shown the potential of bibliometrics for certain aspects, for instance. But in 2022, the continued progress of Clinicum will be of large relevance to AI efforts, as will various continued seminar series, expanding the lively open culture on involving AI in KI work. Further education will also continue to be offered, to raise the levels of knowledge and capacity of KI as an organisation. This includes a basic AI understanding for all, and competence in more advanced AI methods for

- undergraduate students (and the Medical Student's Union)
- Ph D students
- select faculty
- principal investigators

Acknowledgements

I would like to sincerely thank all the regulars at the Falafel seminars, who spoke, asked questions, suggested topics, followed up with me, and who invited me to read project plans, co-write papers, and much more. This, to me, represented the bottom-up perspective on AI at KI and I am pleased to continue the series from March 2022. The top-down perspective was given in almost monthly reports to the President and Vice President of KI, both of whom gave me extremely clear and concise feedback, augmented by a group which also greatly affected the contents of this report for the better: Jan-Olov Höög, Birgitta Janerot Sjöberg, Sabine Koch, Rebecka Skarstam, Sophie Monsén Lerenius and Carl Johan Sundberg. Rebecka with Sophie in many ways helped me think about how to best structure this report and I thank everyone that shared time for interviews by Sophie or me. Anna Martling, Janne Lehtiö and Staffan Holmin has been most supportive of AI for the precision diagnostics efforts in the task force, as has Patrik Georgii-Hemming on the infrastructure side.

Carsten Daub has taken the time to explain the applications of spatial transcriptomics to me, and he too has helped with infrastructure discussions, not least on reference datasets for research purposes. For understanding the clinical side of things I am grateful to my KI project colleagues—Viktor Kaldo and team, Christian Rück and team, Lars E Eriksson, Joseph Hayes, David Zakim and Fehmi Ben Abdesslem—and all the people who met me or hosted me at visits to the Stockholm hospitals, special award to Max Gordon, Stefan Skare, Tove Wšterlid, Fredrik Strand, Johan Lundin and Eric Herlenius for your energy. The by far wildest AI work at KI I found myself involved with was that described as Quantum Life Sciences, for which I thank my academic collaborators Ebba Carbonnier (SWElife and KI), Erik Aurell (KTH), Igor Pikovski (SU), Daniel Lundqvist (KI) and Göran Johansson (Chalmers) and all the company representatives, in particular, Ulf Hertin (SAS Institute). Therese Wahlström and Marie Lind provided important assistance at LIME, where Carl Johan and Sabine were unbelievably generous hosts and collaborators. Thomas Sjöland was as always my flexible and efficient boss at KTH. Petra Dalunde hosted me at AI Sweden, where I was happy to have technical AI discussions with Daniel Gillblad and Magnus Sahlgren. For my own development on the technical parts, I deeply thank my students and mentees. Frantzeska Papadopoulou Skarp gave me extensive feedback on judicial aspects and Adina Feldman gave me advice on autopoiesis: how KI learns, as an organisation. Last but not least, Sandra Eloranta has helped me align this report with Clinicum as well as with methodological developments in biostatistics, and her meticulous reading of several drafts of this report improved it immensely. All remaining errors are my own responsibility and spotting any might get you a cup of espresso.

APPENDIX

The appendix consists of two parts, the first of which follows on the next page.

- The Falafel Seminar Series
- Applied AI examples[61]
- Project Deliverables

[61] This part is an updated section from the half-time report of the AI@KI project.

The second part of the appendix is available on the KI intranet only, or for those without access, upon request.

- Scientometric search terms[62]
- Magnus Boman, Erik Arner, Carsten Daub and Kazuhiro Sakurada
  Biomedical Data for Artificial Intelligence
  Report from the 6th RIKEN/KI SciLifeLab Symposium, 2021
- Sophie Monsén Lerenius:
  From pilot to clinical practice - Barriers and facilitators in the implementation of artificial intelligence in health care: A multiple case study of Swedish AI projects
  Master thesis in health informatics, KI/LIME, 2021

[62] This document provides a freeze frame of the discussion on precisely which terms to include for the AI@KI scientometry.

THE FALAFEL SEMINAR SERIES

Given the restrictions to larger congregations of people in 2020, it could have been hard to bootstrap any seminar activity, but a high level of interest made it viable to start up seminars in AI modelling and programming in the autumn of 2020. With the purpose of supporting a most variable group of AI-interested people at KI with basic training and insights, the Falafel[63] seminars commenced as a dual participation event, enticing a group of maximum eight to physical attendance, and between one and two dozen more joining online. An email list has been created to stay informed, which is now at more than 100 people, with names continuously being added. The seminars run at 5pm every other Friday, to allow for people in education to participate. The list of topics already covered and planned for give a good overview of what is being discussed:

[63] So named because the underlying theme is health. Just as company meet-ups in technical topics like AI programming are expected to offer beer and pizza, I felt vegan food and water was a natural choice for AI@KI.

1. Rebecka Skarstam: Non-expert programming advice for beginner
2. Magnus Boman: The AI@KI project, some preliminary findings
3. Fehmi Ben Abdesslem: Machine Learning for medical applications in Python
4. Joanna Hård: Phylogenetic Fatemapping
5. Peter Sjögårde: Community detection for subject mapping of AI publications at KI
6. Evangelia Gogoulou: Natural Language Processing for medical applications
7. Helga Westerlind: Machine learning for prediction of treatment outcomes in rheumatoid arthritis
8. Ashley Tate: Multiclass machine learning in R: an exploration into adolescent self-harm and aggression
9. Carsten Daub: Basic machine learning in spatial transcriptomics and what we actually should be doing
10. Saikat Chatterjee: On ability and inability in artificial intelligence
11. Erik Aurell: Direct Coupling Analysis - a simple AI technique with biomedical applications
12. Mikael Huss: Machine learning in biomedical research is taking off
13. Johan Lundin: AI for diagnostics in resource-constrained environments - experiences from field studies in Kenya and Tanzania

14. Sandra Eloranta: Risk prediction in lymphoma - can we do better than the International prognostic Index?

15. Multiple speakers: Special Interest Groups on AI

16. Falafel #16, planned for Friday March 11
    Ashley Tate: Bias in variable importance scores

17. Falafel #17, planned for Friday March 25
    Thomas Frisell: Precision medicine and AI for evaluating approved and off-label treatment strategies for MS patients

The plan was always to mix freely in experience and all other dimensions in this series. Each seminar starts with me giving a quick intro and some relevant news, and then 20-30 minutes of talk, followed by an open discussion. The seminars always end formally at 6pm, but in the physical seminar room, discussions and falafel munching has sometimes lasted much longer. To slightly moderate these discussions has been an immense joy for me, as well as a learning experience. A community has slowly built at KI, and new friends are being made.

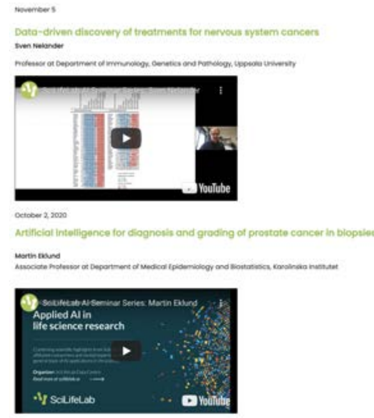


Figure 17: The Web page for the SciLifeLab AI seminar series, which hosted two talks in the autumn of 2020.

APPLIED AI EXAMPLES

The idea that AI-tools can provide actionable insight at the clinic is affecting research at our hospitals. Automatic decision support, automation of human tasks, and augmented researchers and practitioners are all on the horizon. In this section, I made a small selection of cases high-lighted from semi-structured interviews in 2020, updated with 2021 new developments. It is not to be read as a Best In Class, but taken together my selection here does paint a convincing and impressively wide canvas of AI applications.

Interpreting Next Generation MEG-sensor Measurements

At the NatMEG unit (the National facility for magnetoencephalography), next gen MEG-sensors were used to measure weak potentials in the brain of an epilepsy patient.[64] In order to identify complex features in the data registered, a combination of classification algorithms and so-called genetic algorithms were used. Genetic algorithms 'breed' ever-improving solutions to an optimisation problems by evaluating each candidate via a fitness function. The candidates play each other in a tournament or form a converging sequence of values, and here feature vectors were tested by such a fitness function determining the overall similarity between the candidate and the EEG-locked on-scalp interictal epileptiform discharges. This was a world's first MEG measurement on an epilepsy patient and the machine learning algorithm helped identify and classify the discharges. The next gen high-temperature superconducting quantum interference device magnetometer (high-Tc SQUID) is itself extremely interesting from a quantum information perspective, and the new group formed around Quantum Life Sciences in Sweden has shown interest in the NatMEG activities. But sticking here to the machine learning, doctoral student Karin Westin under the lead of Daniel Lundqvist (Neuro division, and head of the unit) suggested a genetic algorithm be employed to create artificial parameter vectors resembling the corresponding real on-scalp data parameters. From this synthetic data, comparisons were made to real discharges and classifications were made based on statistical similarity, through a clever form of anomaly detection that in turn employed a support vector machine. Reading up on the field with the help of Daniel and Karin, I learned that output interpretations from high-Tc SQUID measurements often employ AI methods, but the NatMEG work shows that there is still room for innovation.

[64] S Westin, K., ..., Lundqvist, D. (2020). Detection of interictal epileptiform discharges: A comparison of on-scalp MEG and conventional MEG measurements. *Clinical Neurophysiology*, 131(8), 1711-1720.

Real-Time Decision Support for Sepsis Detection

At CMM/KI and the pediatric departments; including NeoIVA, Pediatric IVA and infectious disease wards, researchers under the lead of professor Eric Herlenius at the department of Women's and Children's Health have developed a deep learning system for early detection of sepsis. Together with experts from KTH like professors Michael Skoglund and Saikat Chatterjee, the team has published on a Hidden Markov Model for sequential physiological data analysis, for instance.[65] Because of the constant monitoring of the preterm babies, any clinical decision support turns into a big data problem: the data must be sieved through and important values harvested. Besides this automated monitoring, there are manual registrations of weight and other relevant parameters. The Deep Machine Learning-based Novel Early Warning System (DeepNEWS) sports an algorithm customised to a Swedish hospital environment and covers the entire population in NeoIVA. An XGBoost model provides for binary (yes/no) classification of sepsis and infection. A clever combination of vital parameters into a predictive model allows for physiomarker indication of important problems in real-time. A risk reduction strategy recommended by the model can then suggest the optimal intervention and do so in time to prevent disastrous consequences. Lessons learned from applying DeepNEWS to data on preterm babies have also allowed for much more extensive data monitoring of patients at the Karolinska university hospital, currently at over 1000 beds. Monitoring data from over 600 CoVID-19 patients has already been collected and analyses are ongoing. In 2022 and onwards, these analyses will be augmented via a collaboration with KTH researchers (myself being one) and with Sabine Koch (KI/LIME) in a VR project called *Scalable Federated Learning*, PI: Dejan Kostic, KTH.

[65] Honoré, A., Liu, D., Forsberg, D., Coste, K., Herlenius, E., Chatterjee, S., Skoglund, M. (2020). Hidden Markov Models for sepsis detection in preterm infants. *IEEE Intl Conf on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1130-1134. It is interesting to note that the first author is a Ph D student at KTH, doing his work situated at Karolinska. Such people migration is sometimes required for a technical topic like AI implementation, but also for securing long-term expertise and engagement on the side of the university hospital. On top of Antoine's stint at KI, his supervisor Saikat will from 2022 be half-time with the group, thanks to an SSF grant.

Improving Spatial Transcriptomics Data to Detect Cancer Signatures

Carsten Daub at Biosciences and Nutrition, KI Syd—also currently a director for the SciLifeLab National Genomics Infrastructure (NGI)—is conducting research into AI for automated image analysis.[66] With his group, he aims to allow pathologists to consider genetic and clinical data for risk prediction.[67] For breast cancer, there are early molecular RNA cancer signatures that can be detected by spatial transcriptomics technology before cancer is apparent in image-based pathology, and such signatures might be recognisable in histology images. The cancer sub-type and severity level of breast cancer can be assessed by histology images once image recognition is trained on expression signatures. In short, given a tissue region classification, a gene-independent machine learning identification of cancer can

[66] He also co-organised the last three KI/SciLifeLab/RIKEN symposia on biomedical data for AI.

[67] Yoosuf, N., Navarro, J.F., Salmén, F., Ståhl, P.L. and Daub, C.O. (2020) Identification and transfer of spatial transcriptomics signatures for cancer diagnosis. *Breast Cancer Research* 22(1), p. 6.

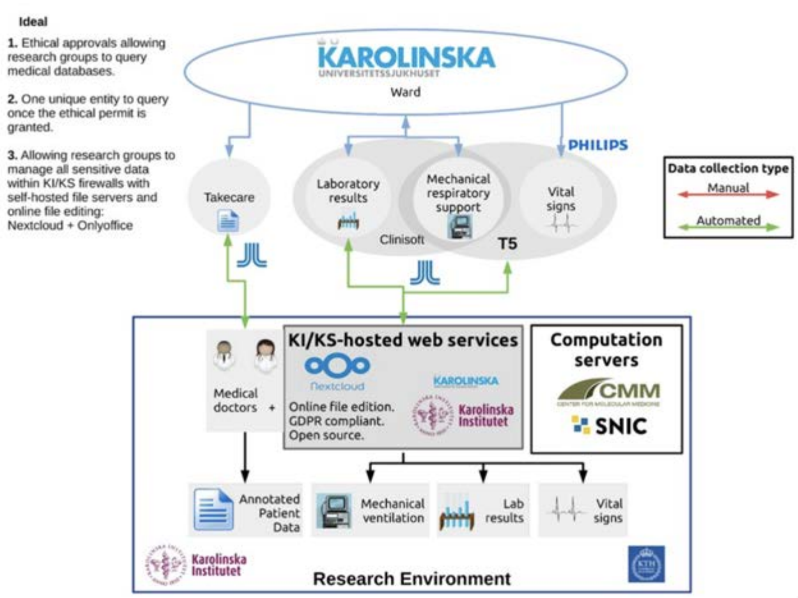


Figure 18: The DeepNEWS infrastructure, as envisioned by Herlenius *et al.*. The system is designed to be scalable to many kinds of monitoring, enabling collaboration between different parts of the hospital. Such collaborations could include a central point for (big) data storage and management.

be made. Deep neural networks can then be trained to learn cancer cell migration patterns. Joint work on this has been carried out with Lund university researchers. A long-term goal would be an optimal segmentation of pathology images of breast cancer samples without using the actual spatial transcriptomics data, developed with KTH researchers.

Adaptive Treatments in Mental Health

Sweden has an impressive track record of Internet-based psychological treatments for among others depression, insomnia, social anxiety, panic disorder, chronic stress and body dysmorphic disorder. For some digital psychological behaviour intervention, clinical researchers have studied treatment engagement, symptom change and other factors, in order to predict successful treatment outcome. Professor Viktor Kaldo is the PI of several projects at the department of Clinical Neuroscience looking to go even further with the help of AI. In a translational collaboration with KTH initiated in 2017, the progress for individuals in treatment at the Internet Psychiatry Clinic is monitored via patients' self-ratings and analysed by a learning machine, presenting its predictions to the therapists via a digital decision support tool.[68] It also meant to generalise to other patient populations via transfer learning, and become more useful over time and over task. While earlier work has shown that identifying individuals at risk of failure can reduce non-responders from 81 to 34 per cent, the learning machine is an attempt to automate

[68] Boman, M., ... Kaldo, V. (2019). Learning machines in Internet-delivered psychological treatment. *Progress in Artificial Intelligence*, 8(4), 475-485. Since I am the first author of this article, I hereby declare bias in any assessment of the significance of this work. Suffice to say that an interdisciplinary group of considerable size has formed and that the work is now under external validation regarding its clinical usefulness via a triple-blind RCT, a very rare bird in AI applied work.

the process, further increasing predictive accuracy. This has vast clinical implications, not least because the patients benefit from this AI-based adaptive strategy while in treatment, and more resources can be directed at the patients most in need. Predicting outcomes for depression, social phobia and panic syndrome has been done with statistical models, indicating that at about one third into the treatment, a patient's responder and remitter status can be predicted from the treatment platform data, with good accuracy. A random forest model has been shown to slightly improve upon this, and a learning machine is under implementation that fuses that model output with the output of two other models, based on natural language processing of patient-generated text. This machine can fuse other modalities, such as genetic data and images, in the future to reach a level of accuracy that further recommends and motivates important psychologist interventions.

Intelligent History-Taking

In computerised history-taking, significant laboratory and imaging findings are incorporated into decision support guidelines for physicians each time a data element is added to a patient's file. The CLEOS system—owned and operated by KI and developed by Professor Emeritus David Zakim and his team—is a software implementation that automates this process.[69] How AI can be used to further develop CLEOS into a full-fledged expert system is under investigation at KI/LIME. The experts represented and emulated are specialists that interview adult patients with problems across any organ system. The system has been deployed in a study with about 2000 patients at Danderyd university hospital since 2017 for history-taking from patients with chest pain in the emergency department.[70] Viewed as a decision tree algorithm, the system is relatively large, with about 13000 questions directed by more than 19000 decision nodes that represent questions and rules for interpreting the clinical significance of the data as it is collected. CLEOS operates by formulating a working differential diagnosis generated automatically as it interviews patients. It selects the most appropriate next question at each decision point to rule in or rule out the differential possibilities. CLEOS can recognise automatically that the working differential diagnosis may not be appropriate and can reformulate it to change the pathway of an interview. This same principle of formulating and resolving a differential diagnosis to account for non-normal findings is used to collect a review of systems organ by organ. CLEOS also collects past history, social and family history data. It can use findings from other scalable data sources (laboratory measurements, ECGs,

[69] Zakim, D. *et al.* (2008). Underutilization of information and knowledge in everyday medical practice: Evaluation of a computer-based solution. *BMC Medical Informatics and Decision Making*, 8(1), 1-12.

[70] Brandberg H, ... D Zakim (2020). A prospective cohort study of self-reported computerised medical history taking for acute chest pain: protocol of the CLEOS Chest Pain Danderyd Study (CLEOS-CPDS). *BMJ Open* 2020;10:e031871.

images, treatments, and hospital course) to interpret the significance of findings. In 2021, a chest pain project with CLEOS was initiated at Danderyd university hospital emergency unit. The aim to assess if patient-reported history-taking with CLEOS can reduce the time to acute coronary syndrome diagnosis, improving the diagnostic accuracy and use of resources. Two KTH bachelor projects investigating the possible use of machine learning in this context have recently started up.

PROJECT DELIVERABLES

The AI@KI deliverables include the following items.

- A series of closed structured presentation of concrete AI activities at KI, with assessments of their maturity and impact, in the form of slide decks
- Numerous applications submitted by KI employees, with named contributions to the AI parts, and in some cases also proposed areas of responsibility, and funding
- Pre-peer review assessments of AI elements in research reports
- Master theses at KI with AI elements, scrutinisation and supervision
- Half-time assessments of Ph D students with AI components in their study plans
- Mentorship of young KI researchers with respect to AI aspects
- The Falafel seminar series, still running
- Interaction with AI Sweden, constituting the KI 'in kind' contribution as a member
- Organisation of, and participation in, AI-related events in or around KI
- Study visits to KI departments and facilities
- Presentations to management and to important stakeholders in the KI ecosystem
- A closed knowledge repository, with interview notes, slides, and relevant papers
- Internal dissemination activities, including content for the project Web page, newsletter contributions, and presentations of the project, internally and externally
- Internationally published and peer-reviewed research papers in collaborations with KI colleagues
- This report