AI@KI: The First Year[1]

Magnus Boman

25 January 2021

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

A year of meetings with people interested in artificial intelligence (AI) is summarised in the form of a requirements specification. All interviewees and discussants were from Karolinska Institutet (KI) or its surrounding ecosystem of research. When met, the requirements paint a picture of a future in which the resources necessary for employing AI in clinical research are in place, with smooth processes in place for the sharing of data, models, and results. KI is in this future an international top player in ethical and efficient AI deployment. That said, the project focus is 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 that is feasible to implement and support be devised by the president of KI, the main stakeholder.

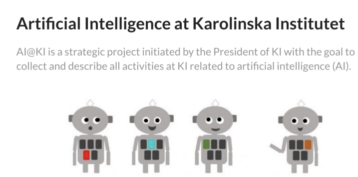


Figure 1: From the project Web page, graphics by Marie Lind.

Executive Summary

The main findings from the first year of this two-year project can be split into positive and negative one-liners where the latter prompt actions to achieve organizational change.[2]

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

+ There are KI people conducting health-related research that use state-of-the-art AI methods with good results
+ Results from AI employment at KI are in some cases of such high quality that they could be published as computer science research and not only as research in the life sciences
+ Many interdiscplinary collaborations are ongoing and some such constellations have sustainable long-term AI employment as a goal
+ The sentiment towards AI is very positive at KI
+ There is full management support for AI use

The requirements discussed in the bread text below, most explicitly in a section named Requirements towards the end, are describing 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

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

— There is no single unit, centre, or point of contact at KI for AI-related questions and support

— While individual research leaders 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, ranging from merely having heard of methods and models to refining and further developing them through applications in labs or at a clinic, something which renders streamlined efforts on further education impossible

— There are research leaders that consider AI methods to be something they have tested and failed to achieve satisfying results, prompting skepticism towards AI in general, even when only a few machine learning methods were used and only to a limited extent

There are also 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

o New life science technology is 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

THE PROJECT HAD A PREDECESSOR that went by the same AI@KI moniker, a series of meetings of people at KI interested in AI. This had consequences. Firstly, some people I interviewed expected me to follow up on promises given in or around 2018 about support for their AI efforts, in most particular ways. Second, I had a list of people that were interested in AI, at least at that time. This helped to quickly identify candidates. Third, some thinking had gone into how to best structure discussions around AI at KI, documented on an internal repository of meeting notes. Most of the people engaged with the old project were also still around to assist me, for example with advice on who to meet first. With this legacy, I mapped out a plan for how to achieve maximum coverage of relevant activities in minimal time. With only one day per week to devote to the project,[3] I settled on semi-structured interviews with principle investigators, and seminars and mentorship with young researchers.[4] A dynamic shortlist of people to interview and advise has been in hand since early February, and it is longer now than ever before, a sign of the project's wide scope but also of its timeliness: people are continuously added to the list whom only recently picked up on AI methods and techniques.

[3] I have applied for external funding to double my 2021 project efforts but the verdict has been delayed.

[4] I have consciously tried to avoid the Senior/Junior dichotomy, but I have found myself still using it at times when speaking about KI, it is as if I have been bitten by a bug.

At the end of 2021, a final report will be delivered with a full maturity analysis of AI at KI. This report is covering about 35 people and their respective work and this will be an important part of that full analysis, which is expected to cover more than 100 researchers. The respective AI interests of the 35 already interviewed is wide, but comes 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 with unsupervised methods. While it is not meaningful to define AI any stricter for the purpose of this report than something having to do with learning structures, I note that ethical considerations of AI often go beyond what was just described. This is fine, since ethical AI needs future-proofing, and I have met with experts on ethics at KI in 2020 for the purpose of pinpointing such aspects.[5] That said, the longest perspectives on AI development—such as singularity research or sentient 'strong AI' systems—have been left out, since they are on the horizon of neither pre-clinical researchers, nor clinical users of AI technology at Karolinska.

[5] A first meeting with Gert Helgesson and Niklas Juth from the Stockholm Centre for Healthcare Ethics at KI/LIME already covered a multitude of topics, including bias in machine learning training corpora, machines for enhancement of humans, screening ethics, and the standardisation of self-learning algorithms.

technology is there

employs the technology

enlist the help of technology experts

secures technology expert help in the long-term

knows, thanks to the expertise, what the technology can do

knows, resting on the expertise, what the technology can do for them

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.

Current users of machine learning have been placed on a ladder I have 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. Only half a dozen are at the bottom rung, with at least five steps fully explored. The AI@KI project is very much studying a moving target, however, and in my first-year reporting to the stakeholders, I have sometimes checked back on individuals or groups, noting that they have recently climbed down another step.

Because the project goals have long-term strategic implications, my work has not analysed current or short-term risks with AI employment. I will incorporate long-term risks into my final report at the end of 2021. What I can observe and assess will here only constitute some pieces of a large puzzle. A thorough risk analysis requires co-creation and full stakeholder involvement. What I will do is engage with experts on ethics and also further engage with the communities of practice in the KI ecosystem. Perceived and foreseen risks of failure can then be mixed with more techno-positive outlooks and opportunities (see Figure 3 for a concrete example of what that could result in).

Figure 3: The risk analysis report resulting from a co-created study of what AI can do for our future sustainable cities, which I led. This SWOT analysis (in Swedish) was reported on in 2018 for the strategic innovation programme Viable Cities. Something similar could be done for KI after top-down strategy decisions on AI have been taken.

1.2 *Data*

Folklore has it that 30-45 per cent of the work in any AI project is spent on pre-processing digital data, getting it ready for machine learning. Different machine learning methods have different tolerances for missing values and noise, for instance, which also affects the time spent in pre-processing. Imputation is sometimes necessary and at other times forbidden. Labels and a 'gold standard' for supervised learning are often used, albeit rarely complete from the outset. If labels are not used, human annotations are disregarded, for example when clustering data points in unsupervised machine learning. But choices of clustering methods and visualization techniques still require human input. Regardless of the situation, researchers must get their hands dirty with the data and spend time on plotting distributions and on various statistical tests, like Pearson correlations and cross-entropy measures. This all means that 30-60 per cent is a better estimate for how much time is spent on pre-processing, with most projects lying in the upper half of that interval. To be realistic about the time required is important, for many reasons, including the following.[6]

- Most researchers consider pre-processing a tedious task and would prefer to 'get on with the work'
- AI is sold 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 enough
- 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 (this is when the first email reply from the judicial department arrives)

1.3 *Explainability*

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.[7] Arguably, the hype around such systems and what they could support the clinic with, helped create the 'AI winter' that followed. 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 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 algorithms

[6] 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 and other more esoteric means to processing, it is often fruitful to look at digital 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.

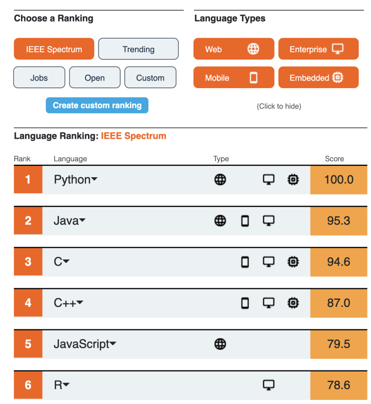


Figure 4: General ranking of programming language popularity, and what type of language they represent. Results according to an app built by respected engineering journal IEEE Spectrum by N Diakopoulous, rebuilt by M Bagavandas and G Singh, and updated by P Kulkarni. Even across all application areas, Python and R are top-6. My own favourite language right now, Racket, is ranked last on spot 55, with a zero per cent popularity. So, we should perhaps take this ranking as an illustrative example, even if eleven different data sources were combined in the app.

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

became a judicial matter. Fears around what role GDPR would play further increased the need for judicial support. This, in turn, made researchers sometimes abstain from computing optima, and 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.[8] When it was recently demonstrated how easy it was to corrupt such explanations,[9] 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.

[8] 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.

[9] Slack, D., Hilgard, S., Jia, E., Singh, S., & Lakkaraju, H. (2019). How can we fool LIME and SHAP? Adversarial Attacks on Post hoc Explanation Methods. arXiv preprint arXiv:1911.02508.

1.4 Programming

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 F-score, balanced accuracy, AUC/ROC, p values, the list is endless.[10] 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 regularise, or how to mix different penalties in regression. The language of choice usually sports libraries of code for such

[10] 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.

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' 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.[11]

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

For AI programming, a dedicated AI software platform is employed, such as TensorFlow, Keras and 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 created at KI around the so-called Falafel seminar series, detailed in the next section.

2 *The Falafel Seminars*

Given the restrictions to larger congregations of people in 2020 and ongoing, it could have been hard to bootstrap any such 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[12] 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 70 people and new names are continuously 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:

[12] 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 beginners, Friday October 9, 2020

2. Magnus Boman: The AI@KI project, some preliminary findings, Friday October 23, 2020

3. Fehmi Ben Abdesslem: Machine Learning for medical applications in Python, Friday November 6, 2020

4. Joanna Hård: Phylogenetic Fatemapping, Friday November 20, 2020

5. Peter Sjögårde: Community detection for subject mapping of AI publications at KI, Friday December 4, 2020

6. Evangelia Gogoulou: Natural Language Processing for medical applications, Friday January 15, 2021

7. Helga Westerlind: Machine learning for prediction of treatment outcomes in rheumatoid arthritis, Friday January 29, 2021

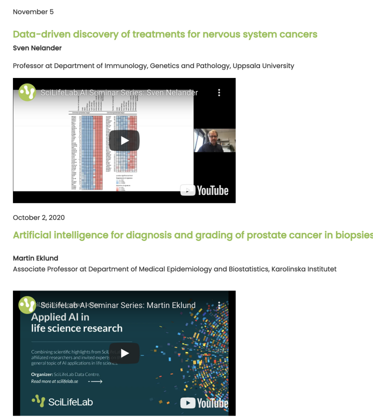


Figure 5: The Web page for the SciLifeLab AI seminar series, which hosted two talks in the autumn of 2020.

For February and March, we have three people with a longer history of AI use lined up, as the plan is 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 is slowly building at KI, and new friends are being made.

The falafel series is not the only AI seminar series in the KI ecosystem. The SciLifeLab Data Centre has recently started a series branded as *Applied AI in life science research*. AI Sweden sports a *Knowledge library* with recorded seminars, and there are relatively many relevant events planned for 2021.[13]

[13] A good example is the bi-weekly *Swedish NLP webinars* seminar series for practitioners using natural language processing, organised together with RISE.

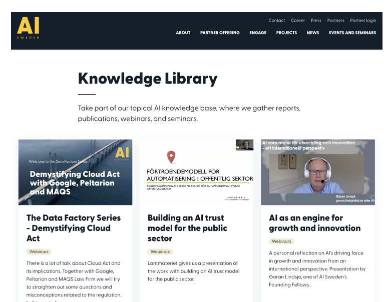


Figure 6: The Web page for the AI Sweden knowledge library.

3 *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. In 2020, that same foundation put SEK3.1B into Data-Driven Life Sciences (DDLS) over the next 12 years.

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

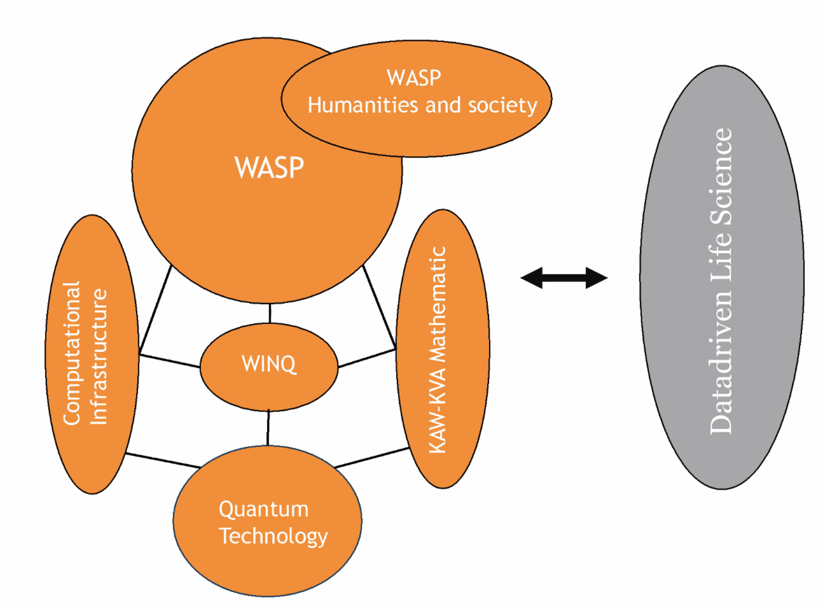


Figure 7: The WASP main programme not only spawned WASP-HS but is also directly related to the Wallenberg Initiative on Networks and Quantum information (WINQ), with many AI connections. A new supercluster for computations—notably for training deep learning models—is about to be made available to researchers.

has been met partly by giving the Swedish funding agencies money ear-marked for AI research and innovation. The following partial list includes efforts that has bearing on health and medicine:

- 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[14]
- 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 national Strategic Innovation Programmes (SIPs) got additional funds for AI activities, spawning a range of small AI project in 2019-20
- 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[15]
- 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

[14] 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* led by Francis Lee, who will co-lead with me a roundtable discussion on March 25, 2021 with 50-60 participants.

[15] KI is a partner, represented by myself and Sabine Koch (professor at LIME).

- 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[16]
- 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, with the mission of providing patients with faster diagnosis and better treatment

[16] I had the pleasure of co-organising the 6th and 7th KI/SciLifeLab/RIKEN Symposium, the theme of which was *Biomedical Data for Artificial Intelligence—The role of AI in the future direction of Life Science research*, which produced a White Paper on reference datasets for biomedical research with AI methods, recently submitted for publication. In these symposia—the sixth in Yokohama in 2019 plus a digital mini-symposium in 2020—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.

4 *Applied AI at Karolinska*

THE IDEA THAT AI-TOOLS CAN PROVIDE actionable insight at the clinic is affecting pre-clinical research. Automatic decision support, automation of human tasks, and augmented researchers and practitioners are all on the horizon. In this section, I will make a small selection of cases high-lighted to me as I completed semi-structured interviews in 2020. 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. It is really very far away from a complete picture too: I aim a lot higher with respect to coverage for the end of 2021.

4.1 *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.[17] 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.

[17] 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.

The next gen high-temperature superconducting quantum interference device magnetometer (high-Tc SQUID) is itself extremely interesting from a quantum information perspective, and I have contacted physicists at KTH to discuss possible interdisciplinary development of such devices and their use in the future. 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 Dr Lundqvist, 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.

4.2 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 professor Michael Skoglund, the team has published on a Hidden Markov Model for sequential physiological data analysis, for instance.[18] 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.

[18] 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.

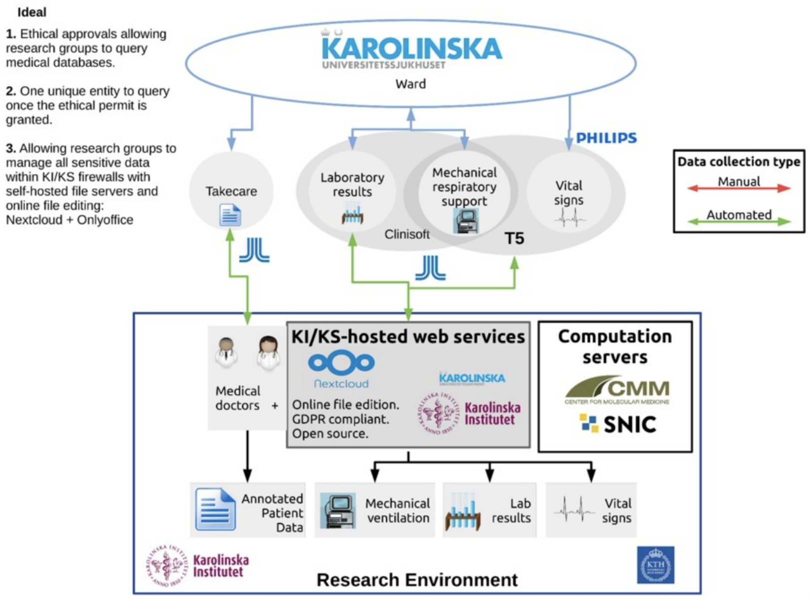


Figure 8: The DeepNEWS infrastructure, as envisioned by Herlenius *et al.*. The system is designed to be scalable to many kinds of monitoring, enabling collaboration between different parts of the hospital. Such collaborations could include a central point for (big) data storage and management.

4.3 *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.[19] With his group, he aims to allow pathologists to consider genetic and clinical data for risk prediction.[20] 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 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.

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

[20] 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.

4.4 *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.[21] 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 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.

[21] 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.

4.5 *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.[22] 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

[22] 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.

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 Hospital since 2017 for history-taking from patients with chest pain in the emergency department.[23] 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 re-solving 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, images, treatments, and hospital course) to interpret the significance of findings.

[23] 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.

5 *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 is 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.

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 its 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.

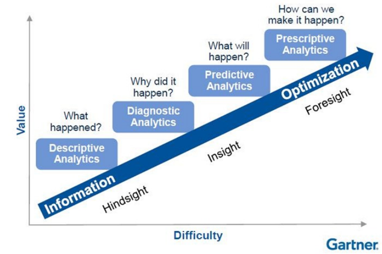


Figure 9: The business intelligence analytics paradigm just before AI and Big Data took over, as shown on Timo Elliot's blog, 14 Feb 2013, courtesy of Gartner. The alleged intelligent analytics paradigm shift pushed the majority of efforts in the direction of the arrow, leaving descriptive analytics almost entirely. In the new paradigm, diagnostic analyses are with AI made faster, with better precision, leveraging on expert systems and other human-machine mergers. Most of the weight from diagnostics is moved over to predictive analytics, avoiding suffering and costs as a result. These savings (sic) are often quantitatively estimated by consultancies, but I have yet to see a calculation that does not ultimately rely on handwaving.

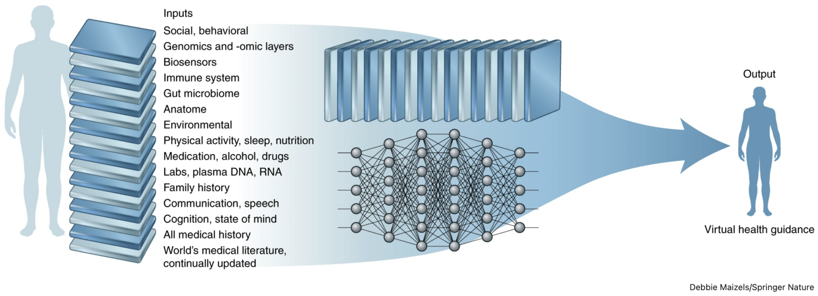
What happened instead was an intense focus on the individual, leveraging on health analytics by indexing relevance in huge data sets on individual health profiling. My favourite term for this is $n = 1$ *medicine*. If a vast space of unstructured data is trawled for every data point relative to the social 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,[24] making it to P5 medicine, but I will here refer to it as *precision medicine*.

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

Arguably the three most eligible kinds of diagnosis for which PM should be able to deliver important means to reduce human suffering, with AI playing a part:[25]

- Cancer (precision oncology)
- Rare diseases
- Mental health (precision psychiatry)

[25] 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.



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 10: In an excellent article in *Nature Medicine*, Eric Topol fleshes out the dream of how deep learning and other AI techniques can help realise PM. The original figure caption reads: The virtual medical coach model with multi-modal data inputs and algorithms to provide individualized guidance.

All three 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. To be able to have long-term positive effects of AI use, learning structures should be saved and re-used. For data to be re-used is an old truth, but that contextualised deeper models could be generalised to new patients is still in its infancy. If such transfer learning becomes easier to implement thanks to PM efforts, it will not only makes diagnostics more efficient, but it will also help predictive models. As a bonus, this mix of PM and AI could help us understand why healthy people stay healthy.

At Karolinska, the PM task force led by Anna Martling has three focus areas, with the development of new diagnostics being one.[26] Another deals with the data infrastructure, while the third area concerns a virtual centre established to support care in practice. Besides the AI-relevant entries listed above (Section 3) as belonging to the KI ecosystem, the following entities play a part for PM+AI.

[26] The blog entry by the KI president (Nov 12, 2020, in Swedish) provides background to this collaborative effort with Region Stockholm.

- Centrum för hälsodata (Stockholm Center for Health Data) supplies 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.
- Genomiskt medicincentrum Karolinska (GMCK) is a part of GMC Sweden in which universities and hospitals collaborate with Clini-

cal Genomics Stockholm at SciLifeLab and Karolinska Universitetslaboratoriet (KUL) on the hospital side.

- 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.

6 *Education opportunities*

Besides internal effort like the Falafel seminar series and cooperation with other universities via e.g. SciLifeLab and MedTechLabs, there are many ways to further educating KI people on AI.[27] Many of the people I have interviewed have taken some kind of online course, with or without certification. How to get further acquainted with machine learning, for example, is naturally best left to the individual. What I have been doing in 2020 is to provide some good options to those who need advice. For 2021, there are two concrete proposals to consider for a larger group of interested individuals. Firstly, professor Saikat Chatterjee at KTH has suggested giving a course in machine learning at KI. This course would be fairly technical and would tentatively have three modules. The first covers Fundamentals of Machine Learning, the second Deep Neural Nets, and the third would bring up KI-specific machine learning applications. The course would run for about ten weeks total. Second, AI Sweden offers a smorgasbord of courses at different levels. They helped develop the Swedish version of *Elements of AI*, a free 15-30 hours online course on basic AI hosted by Vinnova.[28] AI Sweden tailors AI education for groups at organisations and companies in Sweden, chiefly via the AI Competence hub.

[27] The SMILE core facility offers courses on how to use GPUs and languages like Matlab, for instance, most relevant to AI programming.

[28] The course has been taken by more than half a million people worldwide, and the Swedish partners also include Peltarion and AI Competence for Sweden: a group of seven universities.

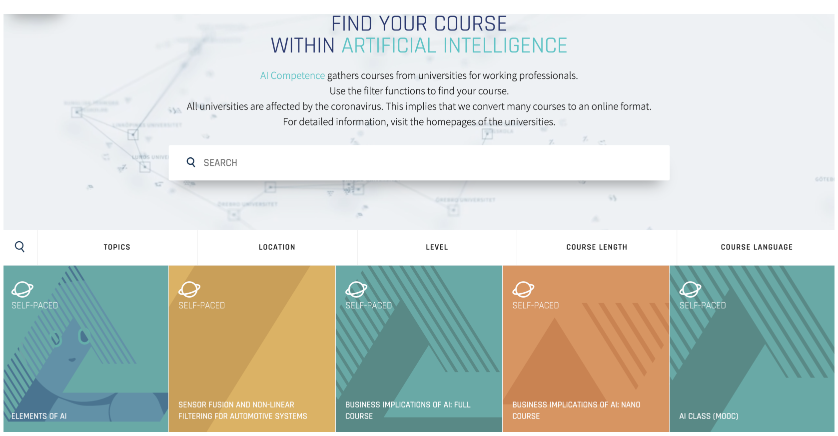


Figure 11: The seven universities offering AI courses to organisations via the AI Competence hub.

7 *Requirements and Corresponding Project Actions*

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 task force. As was explained above, there are many links between the goals of the task force and the capabilities of AI, some of which are already in place at Karolinska. As a first step of linking the efforts, I am happy to be joining the task force in 2021, to act as an AI expert.

There are currently many people at KI in need of support for machine learning models and their efficient implementations. 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 mostly brokered connections to data scientists and machine learning programmers. With Fehmi Ben Abdesslem (employed at RISE and affiliated with KI), who fits this profile, I have myself applied for funding to boost the AI@KI project to be able to ramp up the more practical machine learning assistance. The application was sent to Digital Futures (KTH) and their evaluation of incoming proposals got delayed by three months, but an answer is expected soon.

Noting the gap between successful pilot projects with AI elements and full day-to-day use, following external validation (like an RCT) of the pilot and its refinements, I have initiated an AI@KI activity looking into if this gap could be bridged. This will be done as a master thesis work in health informatics in the spring of 2021 with Sophie Monsen Lerenius, who will work closely with me in continuing the assessment work started in 2020.

The myopia in AI implementations at KI will be fought by continuing and expanding the Falafel seminar series. I have also given a number of introductory lectures at various KI departments in 2020, and these will continue in 2021. This will also assist in the important strategic goal of giving AI at Karolinska a coherent description. Only then can the impressive range of contributions be understood, internally and externally.