OPEN ACCESS

Learning from chess engines: how reinforcement learning could redefine clinical decision-making in rheumatology

Thomas Hügle

Handling editor Josef S Smolen

Department of Rheumatology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

Correspondence to
Professor Thomas Hügle, Department of Rheumatology, University of Lausanne, Lausanne 1011, Switzerland; thomas.hugle@chuv.ch

Received 10 January 2022
Accepted 27 January 2022
Published Online First 8 February 2022

To cite: Hügle T. *Ann Rheum Dis* 2022;**81**:1072–1075.

It is the year 2035. For many years now, the concept of 'shared decision making' has looked nothing like it did in earlier times. Many clinical decisions, such as dose adjustments of methotrexate or certain biologics, are made neither by the rheumatologist nor by the patient, but by computer systems which are more or less autonomous. These consist of digital biomarkers, implanted or skin-integrated sensors and drug delivery systems based on microtechnology and nanotechnology, which have been used for some time in diabetes care. In the meantime, it has been shown that for rheumatoid arthritis and other rheumatological disorders, the disease activity and quality of life can be better controlled with these self-learning systems (formerly called artificial intelligence) than by the rheumatologist alone. Even in the case of non-drug treatments, such as physiotherapy or diet, the patient now receives personalised support through various algorithms. In any desired situation, the options are systematically assessed for their effectiveness and the best ones are suggested. If the treating rheumatologist retires, many years of experience about the individual course of the patient's disease are not lost, but the model continues to improve. It combines existing and new data, enabling it to treat more accurately with every passing day. Non-individual treatment recommendations for diseases no longer exist and treat-to-target strategies are not reviewed every 3–6 months, but daily to hourly. Of course, rheumatologists still exist. But their role has changed, especially when it comes to treating patients with common diseases and uncomplicated disease courses.

How did this development happen? As is often the case, such knowledge was initially developed outside of medicine. Learning systems initially came from the gaming industry, robotics and autonomous driving. In each of these fields, simulators are available that can be used to generate enormous amounts of data in order to test and improve machine-generated decisions. Chess is an excellent example of this.

To understand this better, let us return to the present. In the following, 10 theses are developed to underly the vision described above:

In December 2021, the World Chess Championship took place. Magnus Carlsen won again, retaining his status as World Chess Champion. He made fewer mistakes than his opponent Ian Nepomniachtchi and repeatedly generated surprise with unexpected moves that the chess computer had not predicted. During the live broadcast and in countless YouTube videos, renowned grandmasters commented on every move, every decision of the opponents, and discussed possible better alternatives. The 'gold standard', the best possible move, always comes from a chess computer such as Stockfish.[1] After all, since Deep Blue's victory over Garry Kasparov in 1997, chess computers have been considered invincible due to their computing power.

As another milestone, in 2017, Google's chess computer AlphaZero again defeated Stockfish.[2] AlphaZero, which also consists of neural networks (a subform of machine learning), took a whole 4 hours to learn chess. Unlike Stockfish, AlphaZero was not given any tactical instructions or human chess games from the past. It just knew the basic chess rules, and thus acted completely autonomously through reinforcement learning (RL). In RL, the so-called 'agent' determines which action offers the best decision in a certain situation and at different points in time (figure 1). A reward function is used to determine the best strategy to achieve a medium-to-long-term goal. In this case, to be able to checkmate the opponent at a later point in time. In doing so, the agent may accept sacrifices, even if this generates a worse position in the meantime. The underlying Action-Value formula in RL is the Q-function (Q stands for quality): $Q^{\pi}(s_t, a_t) = E\left[(R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \ldots | s_t, a_t)\right]$. The Q-value for state (s) and a given action (a) is calculated by the expected discounted cumulative reward E, given that state and certain actions. Of note, thanks to the simulator, the chess computer is able to make use of an almost infinite amount of data. To understand this dimension better, there are up to 10^{120} different game courses in chess which can be simulated by the computer.

- *Reinforcement learning (RL), a subtype of machine learning, specialises in making the best possible decisions in a given environment and can far surpass human abilities through simulators.*

This is a different type of machine learning than that which is currently most used in medicine; classical supervised learning. If you look at the list of current FDA-approved machine learning algorithms, there are already over 100 applications.[3] These are mainly used for automated image recognition in radiology or for the detection of cardiac arrhythmias, for example. In most cases, these are automations to support non-complex medical tasks rather than genuine clinical decision-making aids. Almost always, these models have been trained and validated through supervised learning on labelled data sets such as X-ray images (table 1).

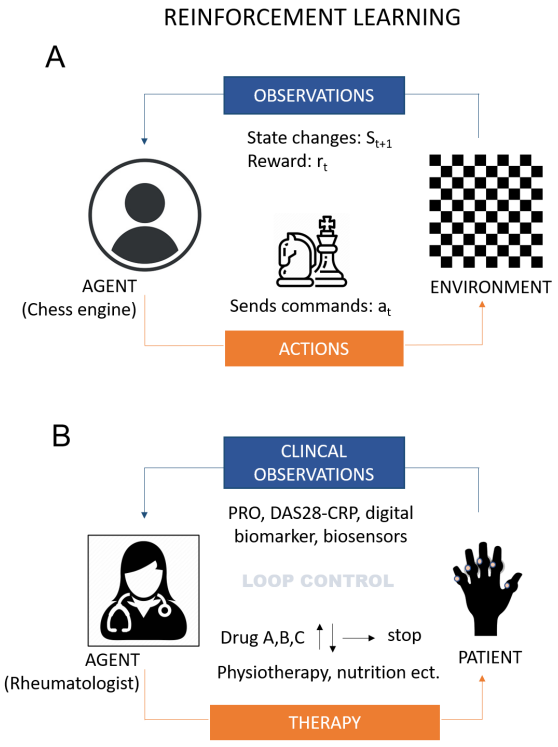

Figure 1 State-action pairs in the reinforcement learning concept using the example of chess (A) with transfer to rheumatology (B). An agent recognises the current situation (state) and independently takes an action. A reward function evaluates the respective decisions with regard to a certain goal, for example, remission. By this loop control, the system constantly improves its decisions. This could be a closed loop in the case of a drug pump and reliable biosensors and digital biomarkers, respectively. PRO, patient-reported outcomes.

Almost always, human labels are the ground truth that cannot be surpassed by the machine. There are exceptions when, for example, subsequent biopsy results are used to train the recognition of tumours on radiographic images. Notwithstanding, these data sets are (and remain) incomparably smaller than data sets from simulators such as in chess or autonomous driving. So, a substantial problem in medicine is that there is no realistic disease simulator in which treatments can be tried out and thus the amount of data cannot be increased while still maintain the quality.

- *In classical supervised learning, models are trained from existing fixed data sets that have been labelled by humans. Such algorithms therefore are supportive and time-saving, but they can never outperform the human performance.*

RL, on the other hand, also recognises and promotes prospectively raised actions through reward functions that lead to a sustainable, good result in the medium and long term. This is what we also expect from medical decisions. However, to a certain extent this means trial and error, which is medically and ethically problematic. On the other hand, we often try new experimental treatments in clinical trials, although under strictly defined conditions and in a fairly controlled manner. RL is therefore exciting because it is a granular decision-making aid for small steps at any desired time. In rheumatology, this could support smaller and less 'invasive' first-line interventions, such as an adjustment of the methotrexate or cortisone dose or non-drug interventions (physiotherapy, dietary changes, etc).

- *RL in the clinical setting will initially take over smaller, ethically justifiable interventions, where there is greater leniency for wrong decisions.*

In the future, we might have to allow the machine to make mistakes, at least to a certain extent, when necessary. After all, we make wrong decisions in the clinic every day. Why should not we allow the computer to do that if it learns to do better next time and the decision is made within certain rules? This approach differs substantially from the supervised machine learning that is currently applied in medicine. In supervised learning, clinical decision support consists primarily of predictions of specific events, such as the future disease status (eg, remission or flares).[4] Through regression analyses with deep neural networks, for example, algorithms trained on clinical data are already used to predict numerical values such as the DAS28-BSR at next visit.[5] Predictions are therefore decision-making aids by providing a more or less concrete look into the future. Potentially, this could improve the quality of therapy through a treat-to-predicted-target concept.

- *Clinical predictions of future disease states using supervised and unsupervised learning are already possible for rheumatoid arthritis, but always refer to previous observations.*

RL algorithms can also be trained retrospectively on large, existing data sets. This has been investigated, for example, for mechanical ventilation or fluid management in over 60 000 Intensive Care Unit stays.[6 7] However, RL becomes really exciting when it is no longer a human person who monitors

Table 1 Explanation of terms and concepts

Artificial intelligence (AI)	General term when computer systems take over tasks that are typically assigned to human attributes such as learning, recognising, planning and so on. Can also be robots or cars that move independently in their environment.
Algorithm	Set of steps for a computer program to accomplish a task or to solve a problem.
Machine learning (ML)	Subform of AI. Computer systems that learn and adapt independently from *data* without following explicit instructions. Can be prediction models or image recognition.
Supervised learning (SL)	Subform of ML. Models are trained and validated in existing, labelled data sets. These are typically used for classification tasks, for example, to predict future disease states or to detect pathologies on images.
Unsupervised learning (UL)	Subform of ML. Models are created from unlabelled data, for example, for clustering or outlier detection in electronic medical records.
Reinforcement learning (RL)	Subform of ML. Models that can make prospective decisions on their own and constantly improve them depending on the results. Works through a reward function (trial and error). Only good actions continue.
Q-learning	Subform of RL. A model-free, flexible RL algorithm to learn the value of a certain action. Random actions outside a specific system can be learnt, for example, by imitating and improving expert actions.
Artificial neural networks	A set of algorithms, modelled loosely after the human brain, in the form of different layers similar to neurons. A powerful tool which can be used for supervised, unsupervised or RL.

the reward function, but the environment is checked by the machine itself and actions are carried out independently. This is already possible in diabetes.[8] Through constant blood glucose measurement as a biomarker, the situation is assessed, and the micropump automatically injects an appropriate dose of insulin. Of note, the algorithm was trained beforehand to know how much insulin is approximately necessary in each situation and has strict constraints on the maximum amount of insulin that can be injected to avoid hypoglycaemia. This could potentially be carried out in rheumatology through the analysis of patient-reported outcomes (PROs), digital biomarkers and skin-integrated biosensor patches for the continuous measurement of inflammatory markers such as C reactive protein or cytokines. Altogether, this could assess the 'state' of the patient. A methotrexate pump or another implanted drug-delivery system could then carry out an 'action' and according to the response, this action will either be corrected next time or not (figure 1B). However, in clear contrast to diabetes, there are multiple possible biomarkers in inflammatory arthritis (not just glucose) and drugs have a much longer duration of action (weeks to months) compared with insulin. Furthermore, multiple antirheumatic drugs are often used at the same time and comorbidities such as fibromyalgia or depression or even side effects might confound digital biomarkers or PROs.

- *Disease-specific digital biomarkers and biosensors detecting inflammation are needed to make RL models more applicable to rheumatic diseases.*

It is essential that rules are imposed on such algorithms. In the case of the chess computer, these are the basic chess rules. Within these rules, anything is possible, even a queen sacrifice. In medicine, no patients or joints can be sacrificed and no regulatory rules can be disregarded. Compared with the regulation of blood glucose, inflammation as a 'system' seems more complex and algorithms must underly even more constraints and imposed rules (eg, in infection). At least for the time being, these rules still concern treatment recommendations, labels for reimbursement, contraindications or allergies.[9] Another important point is that an RL model must recognise when it is *not* in a position to make a decision. Quantitatively or qualitatively insufficient data must be recognised before taking action. This corresponds to situations in everyday clinical practice, where a doctor cannot make a decision without further diagnostics, for example.

- *In medicine, RL models must underly constraints based on expert knowledge and regulatory issues. An algorithm must be able to reject decisions, for example, due to low data quality or the lack of diagnostic information.*

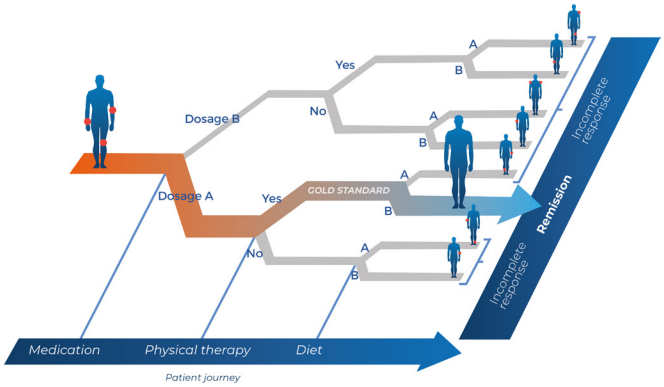

Figure 2 Multimodal decision making by reinforcement learning algorithms at different time points. Adapted from Ref. 15.

The attractive thing about RL as a decision-making professional is without doubt that every small therapy step and every situation can be re-evaluated by the algorithm (figure 2). Treatment recommendations in the form of hierarchical lines of therapy (first line, second line, etc) will no longer exist. Rather, the machine will create situation-dependent, highly granular standards, which include not only drug interventions but also lifestyle interventions. Thus, there will no longer be a rigid treat-to-target concept that is reviewed after 3–6 months. Through RL, it will be possible to achieve long-term targets through smaller and more regular treatment decisions.

- *RL makes treatment recommendations more flexible and granular. On the other hand, treatment targets become more long term.*

Due to the increasing availability of real-world data, such as PROs via apps, information on subjective symptoms, physical activity, nutrition and more can be incorporated into the algorithm more easily. Mental state and work ability can also be recorded regularly. In addition to the classic disease activity measures, other disease outcomes selected by the patient can be optimised. This is also necessary, because a minority of patients with rheumatoid arthritis do not achieve full remission despite targeted therapies. With RL algorithms, the 'point of care' of treatment may shift towards the empowered patient, who can better monitor and control his or her own treatment. Of course, this is restricted to patients appreciating such a computer support.

- *Therapy recommendations by RL will not only refer to medication, but also include physical activity, lifestyle modifications and diet, if this has a positive impact on quality of life.*

Back to chess, Magnus Carlsen stood out with unconventional moves and won. In fact, in chess one can distinguish computer-assisted decisions from human moves, or at least express a suspicion.[10] AlphaZero was a gamechanger. Through pure RL, AlphaZero did not win against Stockfish because it calculated faster. It only examined 60 thousand items per second compared with Stockfish's 60 million. AlphaZero played more creatively, closer to reality. It knew what it was thinking about and what it was ignoring. AlphaZero understood chess better than the calculating machine Stockfish.

Accordingly, RL rather than pure computational power may support human clinical decision-making in future. Computational simulations for rheumatic diseases both on a molecular and clinical level would leverage the performance of such algorithms immensely, although this still seems unthinkable today due to the complexity.[11]

A new approach to RL comes even closer to human reasoning. Originating from robotics, inverse Q-learning with constraints was developed. Here, the Q-function described above follows an expert policy. The machine trains itself to copy the action of an expert as best as possible and improve further while adhering to certain rules.[12] While the perfect rheumatologist to be copied probably does not exist, it could be a safer, smoother way to use artificial intelligence as a clinical decision support tool.

- *New algorithms are consciously oriented towards improving existing human actions while respecting certain limitations.*

So, computer algorithms are becoming more human and, to some extent, will be integrated into the shared clinical decision process within the next few years.[13] Of course, there are also several dangers and challenges involved, such as equity and the access to technology, especially for vulnerable groups. And yet, human decisions will always be necessary in medicine due to the complex interrelationships and the fact that

decisions are not always logical. The role of the doctor is not just to make decisions, but to listen to, inform and deal with emotions. Especially with multimorbid or elderly patients, it is not always about bringing the patient in remission, but also about evaluating factors such as polypharmacy or certain side effects that may influence the quality of life. Clinicians fortunately continue to act not only in a data-driven way, but also through experience, empathy and intuition. Those are features that are unlikely to be taken into account by RL-systems in near future. Notwithstanding, future rheumatologists will have to acquire a certain technical understanding of the quality and function of such algorithms, their data sources and medical devices such as sensors or autoinjectors that already exist for methotrexate.[14] In any case, new advances in medicine such as new drugs or biomarkers are first applied by humans and, if necessary and appropriate, can later be made accessible to automated or semiautomated systems as described here. Therefore, it can be concluded that:

- *Neither RL nor other types of artificial intelligence will ever replace a rheumatologist*.

Contributors I am the only author of this article.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study does not involve human participants.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iD

Thomas Hügle http://orcid.org/0000-0002-3276-9581

REFERENCES

1 Stockfish 14 - Open Source Chess Engine (stockfishchess.org).
2 Silver D, Hubert T, Schrittwieser J, *et al*. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science* 2018;362:1140–4.
3 Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices | FDA. Available: www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device
4 Norgeot B, Glicksberg BS, Trupin L, *et al*. Assessment of a deep learning model based on electronic health record data to forecast clinical outcomes in patients with rheumatoid arthritis. *JAMA Netw Open* 2019;2:e190606.
5 Kalweit M, Walker UA, Finckh A, *et al*. Personalized prediction of disease activity in patients with rheumatoid arthritis using an adaptive deep neural network. *PLoS One* 2021;16:e0252289.
6 Peine A, Hallawa A, Bickenbach J, *et al*. Development and validation of a reinforcement learning algorithm to dynamically optimize mechanical ventilation in critical care. *NPJ Digit Med* 2021;4:32.
7 Komorowski M, Celi LA, Badawi O, *et al*. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med* 2018;24:1716–20.
8 Tejedor M, Woldaregay AZ, Godtliebsen F. Reinforcement learning application in diabetes blood glucose control: a systematic review. *Artif Intell Med* 2020;104:101836.
9 Smolen JS, Landewé R, Bijlsma J, *et al*. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann Rheum Dis* 2017;76:960–77.
10 Cheating in chess - Wikipedia.
11 Dent JE, Nardini C. From desk to bed: computational simulations provide indication for rheumatoid arthritis clinical trials. *BMC Syst Biol* 2013;7:10.
12 Kalweit G, Huegle M, Werling M. Deep inverse Q-learning with constraints. *arXiv* 2020.
13 Hügle M, Omoumi P, van Laar JM, *et al*. Applied machine learning and artificial intelligence in rheumatology. *Rheumatol Adv Pract* 2020;4:rkaa005.
14 Saraux A, Hudry C, Zinovieva E, *et al*. Use of Auto-Injector for methotrexate subcutaneous Self-Injections: high satisfaction level and good compliance in SELF-I study, a randomized, open-label, parallel group study. *Rheumatol Ther* 2019;6:47–60.
15 Gottesman O, Johansson F, Komorowski M, *et al*. Guidelines for reinforcement learning in healthcare. *Nat Med* 2019;25:16–18.