Published in *Artificial Intelligence for Better or Worse* in 2019 by Future Leaders (www.futureleaders.com.au)

Artificial intelligence in healthcare

Khoa Cao and Luke Oakden-Rayner

What is AI?

It may surprise you, but the first dreams of artificial intelligence did not arise in the basements of an MIT engineering lab, or the cosy rooms of an Oxford college. In fact, these dreams did not arise at any modern university or institution. *The Iliad*, Homer's epic about the Trojan War, provides the oldest surviving record of a description of artificial intelligence:

> *He had fitted golden wheels to their feet so that they could run off to a meeting of the gods and return home again, all self-propelled — an amazing sight ... They were made of gold, but looked like real girls and could not only speak and use their limbs, but were also endowed with intelligence and had learned their skills from immortal gods.*[1]

Since then, artificial intelligence has been found in myths and tales across the world, from Yan Shi's life-sized automaton in ancient China, to mechanical robots that protected Buddha's relics, to al-Jazari's drink-serving waitress in the Islamic Golden Age, and to Da Vinci's mechanical knight.

Modern artificial intelligence (AI for short) found its roots with Alan Turing's seminal paper in 1950, which proposed the "Turing Test", a test to examine whether a machine can act indistinguishably from a human. In 1956, Dartmouth held a

workshop widely considered to be the birthplace of AI. Despite having a turnout of 20 attendees, the workshop was the first to establish the name "artificial intelligence", and unveil the very first AI computer program, which could prove 38 mathematical theorems, some more elegantly than known proofs.

The history of AI in the previous decades have been characterised as seasons, an apt metaphor to describe the flourishing and decline of the field. Following the Dartmouth Conference, the floodgates opened for what is now called the first "AI spring". Generous funding and unbridled optimism permeated the field. An interview with the founder of MIT's AI Lab in 1970 summed up the confidence — "from three to eight years, we will have a machine with the general intelligence of an average human being".[2] As the fog of optimism cleared, researchers realised the difficulty of tackling AI. A lack of progress in the field led to the first "AI winter", a period when it was near impossible to obtain funding for AI research.

Winter transitioned to the second AI spring in the 1980s with the rise of "expert systems", which could reason through hundreds (or thousands) of human-written rules. Many of the first expert systems were built for doctors. Stanford's MYCIN program ranked bacteria based on a patient's infectious symptoms with 65% accuracy. Pittsburgh's INTERNIST had rules for 80% of all diagnoses in internal medicine. This sounds impressive, and programs performed admirably compared to human experts, but researchers quickly ran into two problems. First, the best doctors were those who had gained intuition through a lifetime of experience. Imagine writing all the rules for every disease — there might be millions of them! Second, these systems lacked common sense. Putting in unusual or

unexpected data could lead to grotesque mistakes. These shortcomings led to the second AI winter several years later.

As access to large amounts of data ("big data") and computing power increased, a small research field known as deep learning began to excel, leading to today's AI spring. Deep learning relies on "neural networks", a collection of connected nodes that can receive and transmit a signal from one node to another. In reality, these nodes are simply a set of numbers that influence and are modified by other connected numbers. Despite its name, neural networks are only loosely based on the brain's neurons, which exhibit far more complicated behaviour, and the word "deep" refers to the many "layers" in neural networks.

Neural networks are typically "trained" by providing a set of data with the "correct" answer, before using neat mathematical methods to improve the system's prediction ability. If you were from Mars and did not know what a cat or dog was, one way to train you would be to provide an album full of pictures which are labelled either "cat" or "dog"! Instead of writing thousands of different rules like expert systems, neural networks figure out the rules for themselves.

The neural network has provided the foundation for many new and active fields of research. One extremely successful application is computer vision, where neural networks have been able to understand the content of images, such as recognising objects in photographs or identifying diseases in medical scans. The networks in computer vision can be extended to videos (which are just a series of images) and audio (because sound waves can be laid out like an image). Computer vision techniques are being used by Google's

Waymo and Elon Musk's Tesla to develop self-driving cars — a challenging task given how many obstacles a driver may run into!

A second field is natural language processing, which tackles tasks such as speech, writing and translation. Automated speech can be found in virtual assistants, such as Amazon's Alexa and Apple's Siri. One of the most advanced systems today is Google's Assistant, which adds in natural "umms" and "ahhs" and can call restaurants to reserve dinner on your behalf. OpenAI's GPT-2, which has not yet been released due to fears of malicious applications, only requires a short "news prompt" before generating an entire article from scratch.

The final application is reinforcement learning. In this technique, programs are termed "agents", who can complete "actions" based on their "observations" of the world to maximise a certain "reward". This is another way of describing how humans work. Imagine a new office employee. If she works hard and is rewarded with a fancy new title or an envelope full of cash, she will be motivated to continue her hard work. If instead she is rewarded with sleep deprivation and isolation, her persistence will rapidly deteriorate.

Reinforcement learning agents take the "trial and error" approach, experiencing millions of different actions in millions of different scenarios before learning the best action for every situation. Many of its successes have been in playing games, which provide a clearly structured environment and reward system. Recent landmark successes have included DeepMind's AlphaGo, which beat the world champion in Go, a Chinese strategy game with more combinations than atoms in the universe. This achievement was recently superseded by

OpenAI's Five, which beat the reigning world champions in Dota 2, a cooperative strategy game requiring teams of five human players, with each player having 170,000 possible actions at any time. OpenAI achieved this through pitting the program against itself and playing for 45,000 years of gameplay over a period of 10 months in real life.

Despite these exciting advances, neural networks are not without their shortcomings, which are crucial to understand in developing or evaluating any medical AI system.

First, neural networks are extremely data-dependent. "Garbage in, garbage out" is a common adage within AI circles. If the data is not labelled correctly, if there isn't enough of it, or if the data is biased (e.g. only Caucasian skin images in a skin lesion database), then the system will not be reliable enough. In the medical field, specific challenges include the accumulation of enough data, privacy and consent concerns and the labelling effort required from busy doctors. Neural networks can require up to millions of data points before it figures out the rules, and getting doctors to produce these data points is time-consuming and expensive.

Second, many neural network architectures are "black boxes". The majority of neural networks cannot explain why a decision was reached, an important requirement in medicine, and this remains an active area of research. Neural networks are also domain-specific. A program trained to excel in playing chess cannot play checkers or do anything else.

Lastly, neural networks still lack common sense and can be tricked. This is best demonstrated by "one-pixel attacks", where a carefully placed black dot on an image of a horse can

completely break the neural network, leading it conclude it is a frog. It's hard to imagine this trick fooling a human!

Setting the scene for medical AI

The widespread adoption of electronic medical records in many high-income countries has led to a data "explosion", with the industry anticipated to have 2,314 exabytes (1 exabyte = 1 billion gigabytes) of data in 2020. Although this may be cause for optimism, use of healthcare data is often complicated. Estimates indicate that 80% of healthcare data is unstructured (such as free text), which adds an extra layer of complexity in AI training. There is significant variation in how doctors write notes and dealing with this variation can be prohibitive in a broad range of medical AI problems. Data is also most powerful when hospitals share their data, but "data silos" are common, often as a response to privacy and commercial concerns.

Most countries are labouring to contain healthcare costs, with rising chronic diseases and an aging population. Automation of labour, ranging from medical paperwork to diagnosis support, presents an attractive solution that may improve efficiency and accessibility. Although AI programs may be expensive to develop and validate, they are an intangible asset and can be replicated and scaled at near-zero cost. A present challenge is determining legal responsibility. While the first wave of AI programs may place legal responsibility on the physician (by marketing themselves as decision aids), more and more programs may claim full diagnostic or therapeutic autonomy. In such scenarios, companies who develop the programs may be held liable (as is the current case with

medical devices), and licenses may be provided (and removed if too many mistakes are made).

Finally, more patients are actively participating in their own healthcare as consumers. AI programs that safely provide medical knowledge will inevitably empower patients. However, a more informed public may choose to opt-out of personal data usage in medical AI research, a decision which is influenced by culture and societal trust. In Sweden, for example, research participation is perceived very positively, which has led to a goldmine for epidemiological and medical AI research and a rise in patient engagement.

Frontiers of medical AI

In this section, we examine some interesting examples of medical AI, from elementary to complex applications. As it would not be possible to cover all current use cases, we hope the selection is representative of the frontiers of medical AI research.

Digital symptom checkers

Digital symptom checkers are applications available to the public, which have become extremely popular in the UK. They suggest a ranking of likely diagnoses based on an individual's self-reported symptoms. Examples include the NHS Triage, Babylon Health, ADA and Your.MD. Some apps use expert-system style branching logic, while others have trained neural networks on real medical records.

While these applications may ease the workload of general practitioners in the UK, one study from the *British Medical Journal* found an accuracy of 34% for the first diagnosis (and

57% for the top three diagnosis), with a higher accuracy for emergency cases (80%) compared to less serious diseases (33%).[3] It remains unclear whether these applications provide a net benefit to society. Although an accuracy of 80% sounds impressive, this translates to a misdiagnosis of one in five emergency cases. Conversely, such applications may combat the shortfall of care from the difficulty of booking same-day GP appointments, overworked emergency departments and patients googling their symptoms.

Similar approaches to digital symptom checkers have been applied to develop one of the most extensive databases of medical information through physician contributions (the Human Diagnosis Project) and to build medical virtual assistants that can answer healthcare questions (MedWhat, Nuance's Dragon MVA, HelloRache). These assistants, or "chatbots", are coded with natural language processing, to understand the question being asked, and to structure the response based on a large volume of medical information.

Event prediction

Event prediction with neural networks is also popular. These applications provide the likelihood of a defined event given related variables. An example is an AI model developed by Partners Health Network, a large hospital network in Massachusetts, which was able to predict the risk of hospital readmission in 30 days with an accuracy of 76.4% based on over 3,500 variables from electronic medical records.[4] This enabled care teams to target interventions at the highest-risk patients to improve clinical outcomes and prevent further readmissions. Investing an extra day, or giving more information to the right

patient, can be the difference between a re-admission and safe care in the community. The Assistance Publique-Hôpitaux de Paris, one of the current leaders in event prediction, have used medical record data to predict hospital admission rates, recommend staffing levels, and suggest the best hospital for patients to attend to receive the most efficient care.

Programs have also been developed to accelerate clinical trial recruitment (such as HealthMatch) and predict the probability of survival over a certain timeframe for palliative patients, a challenging aspect of end of life care (Stanford ML Group). The challenge of building these systems lies in the implementation of an electronic medical record that can capture clinical data with sufficient accuracy and detail. The majority of free text data has too much variation between doctors to allow for effective event prediction. It also remains to be seen whether prediction of medical events from wearable data (such as the Apple Watch or FitBit) will enter widespread use. Wearables are limited to relatively basic biosignals (such as a heart signal, oxygenation or heart rate), but more creative uses have been developed for invasive implants, such as Medtronic's Artificial Pancreas, which has recently been approved for use in the automated management of blood sugar levels.

Computer vision

As one of the more advanced AI fields, some of the world's largest companies have started to tackle computer vision problems in medicine. The majority of these problems are found in radiology, dermatology, ophthalmology and pathology, which are specialties heavily reliant on images. Both Stanford University and IBM have published research on the

classification of skin cancer with deep learning, demonstrating dermatologist-level accuracy. In ophthalmology, Google DeepMind has trained a medical AI software to detect over 50 eye diseases with 94.5% accuracy on specialised eye scans. Pathology and radiology have seen extensive research efforts, ranging from detection of pneumonia in chest x-rays and lung cancer in CT scans, to interpreting biopsies to look for cancer cells.

While these successes are impressive, it is worth remembering that neural networks remain extremely task specific (a software trained to detect pneumonia would not be able to detect other diseases). Although many companies have published computer vision research, the path to implementation in actual medical practice remains obscure. A big part of the challenge, for example, is integrating AI software with current hospital IT systems. This first generation of approved AI software may be implemented with heavy restrictions, and it remains unclear how regulatory bodies will manage the risks and benefits of this technology. One particular issue is how to regulate AI systems that can learn and be updated (e.g. whether companies are allowed to update at all or need to repeat the approval process with updates), which could enable improved accuracy over time, but risk greater unpredictability.

Other notable uses

While these may represent some dominant medical AI examples today, the use of neural networks in healthcare is incredibly diverse. Neural networks have been used by BenevolentAI to read through millions of academic papers and clinical trials to predict new drugs and their anticipated effect.

Google DeepMind have used AI to predict the shape of proteins based on its genetic sequence (an extremely challenging problem defined by Levinthal's paradox). Networks may also be used to identify the most important genes in specific diseases to better understand the genome, although it remains unclear how this may be combined with genetic editing technology provided the multifactorial nature of most diseases. The neural networks used for images and videos can be adapted for audio, such as ResApp's prediction of childhood respiratory diseases based on cough and breathing sounds.

AI is also being developed for technologies that may sound closer to science fiction. Neural networks may present an important solution to interpreting the chaos of brain signals and developing accurate brain-computer interfaces, which can convert thoughts to actions. Although it is currently "feasible", modern interfaces require extensive training by individual users (sitting in a dark room for days completing only a single action) and remains rudimentary (such as being able to translate thoughts to only 10 words). Finally, a few companies are hoping to combine computer vision, reinforcement learning and surgical robotics for truly automated surgery. Although theoretically possible, this is likely to be extremely challenging, with recent automated robots suturing at only one-third the speed of a surgeon, with an 86% success rate.[5] Most surgical procedures are far more complicated than simple suturing.

The future

Predictions for AI in the past have been surprisingly bad. Experts at the 1956 Dartmouth Conference suggested that AI would be solved by 1957. In 2017, a survey of 350 AI

researchers from the University of Oxford's Future of Humanity Institute demonstrated a broad range of predictions of when AI would exhibit "general intelligence" — an ability for one program to perform any intellectual task a human can, from 2025 to never, with Asian researchers more optimistic than their Western counterparts.[6] It is clear that no-one seems to know what the precise future of AI is, whether the field may enter another winter, reach stagnation from research problems that are too hard to solve, or develop super-intelligent agents to reach the technological singularity, a hypothetical event where technological growth becomes uncontrollable and human society changes irreversibly. Regardless of where your opinion lies, research today will influence the future of AI and by consequence, medicine.

Today, most AI research relies on developing more efficient architectures and learning algorithms for neural networks. Networks are becoming more accurate, faster to train, and require less data (and power). Coupled with stronger computers, it would not be a wild prediction that networks will continue to improve in these aspects. Some interest has arisen in combining the strengths of structured expert systems, which can be "taught" rules rapidly, with the strengths of unstructured neural networks, which can learn more extensive rules, to form "hybrid" statistical and symbolic systems. The computational power required to train AI algorithms remains high, and approaches to reduce this burden include developing specially designed electronics (e.g. FPGAs), using quantum algorithms, and developing computers which mimic our brain's neurons (the brain uses 25W of power, while the average AI program requires 2000W to train). Such advances in

AI will lead to faster, more accurate and potentially more complex medical AI programs.

It is unclear whether the current research paradigm can lead to AI that succeeds beyond narrowly defined tasks. Our brains are significantly more complex than modern neural networks, and entirely different structures may be required to emulate general intellect. One example of a promising research direction are "active inference" agents, which aim to minimise surprise rather than maximise reward (reinforcement learning), a seemingly small difference that may be a stepping stone to creating more generalised intellect, because agents have intrinsic motivation to explore the environment.[7] While intellect presents one significant challenge, our current understanding of consciousness is entirely insufficient to predict whether silicon-based agents can ever develop it. We have not been able to consistently define consciousness, where it comes from or how to build it. Future AI might be extremely intelligent across many domains but lack true agency or consciousness unless it is explicitly defined by a human.

Can AI replace doctors?

Whether AI can ever truly replace doctors is a controversial topic in the medical field, but a very fun thought experiment nonetheless. If I provided you a billion dollars and asked you to build software strong enough to replace doctors, where would you start? The daily life of doctors is an assortment of tasks, from talking with patients, acquiring clinical information from a mixture of sources, writing notes, processing medical information and miscellaneous procedures. AI programs may provide significant support, granting doctors more time to

spend with patients. However, the cost of automating a wide range of tasks would be high, and it may be simply cheaper to train doctors. The economics of healthcare would therefore prevent a replacement of doctors.

However, a government somewhere might decide this is the wrong conclusion! What would then be next steps? It makes sense to "task-shift", a public health strategy that ensures doctors complete only tasks at their level of training. A highly trained surgeon, for example, should be spending her work day on the most complicated surgical tasks, rather than filling in paperwork (which a specially trained clerk may complete) or opening and closing wounds (which a junior doctor may complete). Task-shifting has been utilised extensively in India's phenomenal Narayana Health, which has reduced the cost of open-heart surgery to $2,000, a procedure that costs $100,000 in the United States, with complication rates comparable to America's best hospitals.[8]

If we task-shift extensively, the primary role of doctors becomes collecting and processing the right information, producing a diagnosis, performing procedural work and providing emotional support (although many would argue that nurses do a far better job). There is nothing to suggest that it would be impossible to build an AI agent that collects pertinent information or provides diagnosis at a superhuman level. The difficulty lies in building software for every single diagnosis, a task that will be extremely challenging if AI remains narrow, but easier if AI gains more "general" intellect and an ability to reason from abstract knowledge. Medical and surgical procedures are more difficult to tackle due to the added

complexity of teaching robots truly autonomous movements, although hybrid networks may present a promising solution. The task is not theoretically impossible, and new surgical paradigms, such as intraluminal robots or nanorobots (famously popularised by Richard Feynman), may completely change medicine.

A final argument against whether medical AI can replace doctors revolves around the emotional aspects of medical care. Disease is an innately emotional experience and the best doctors are those who listen and support us through tough times. The counter to this, however, is that emotional support is not provided only by doctors, but all staff in healthcare facilities, from our nurse to the volunteer who brings us a cup of tea and listens to our stories. Surprisingly, emotional support can also be provided by AI, which can be trained to sound and act more human (such as Siri's off-hand jokes or Google Assistant's "umms" and "ahhs"). Whether society as a whole is comfortable with emotional robots is dependent on culture. For example, Japan is far more relaxed with robots and they are already popular as companions who can care for and talk to the elderly.

All up, medical AI is an exciting field that is filled with significant potential and challenges. Understanding the nuances of artificial intelligence and its medical applications is crucial in delineating hype from reality and guiding the field towards the improvement of healthcare for everyone.

Endnotes

1 Homer FR (1990). *The Odyssey*. Vintage Books.

2 Darrach B (1970). Meet Shaky, the first electronic person. *Life Magazine*, 58B–58D.

3 Semigran HL et al. (2015). Evaluation of symptom checkers for self diagnosis and triage: Audit study. *BMJ*, *351*, h3480.
4 Golas SB et al. (2018). A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: A retrospective analysis of electronic medical records data. *BBMC Medical Informatics and Decision Making, 18*, 44.
5 Sen S et al. (2016. Automating multi-throw multilateral surgical suturing with a mechanical needle guide and sequential convex optimization. In *2016 IEEE International Conference on Robotics and Automation,* 4178–4185.
6 Grace K et al. (2018). When will AI exceed human performance? Evidence from AI experts. *Arxiv*.
7 Friston K (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*, 127.
8 Bhatti YA et al. (2017). The search for the holy grail: Frugal innovation in healthcare from low-income or middle-income countries for reverse innovation to developed countries. *BMJ Innovations*, *3*, 212–220.