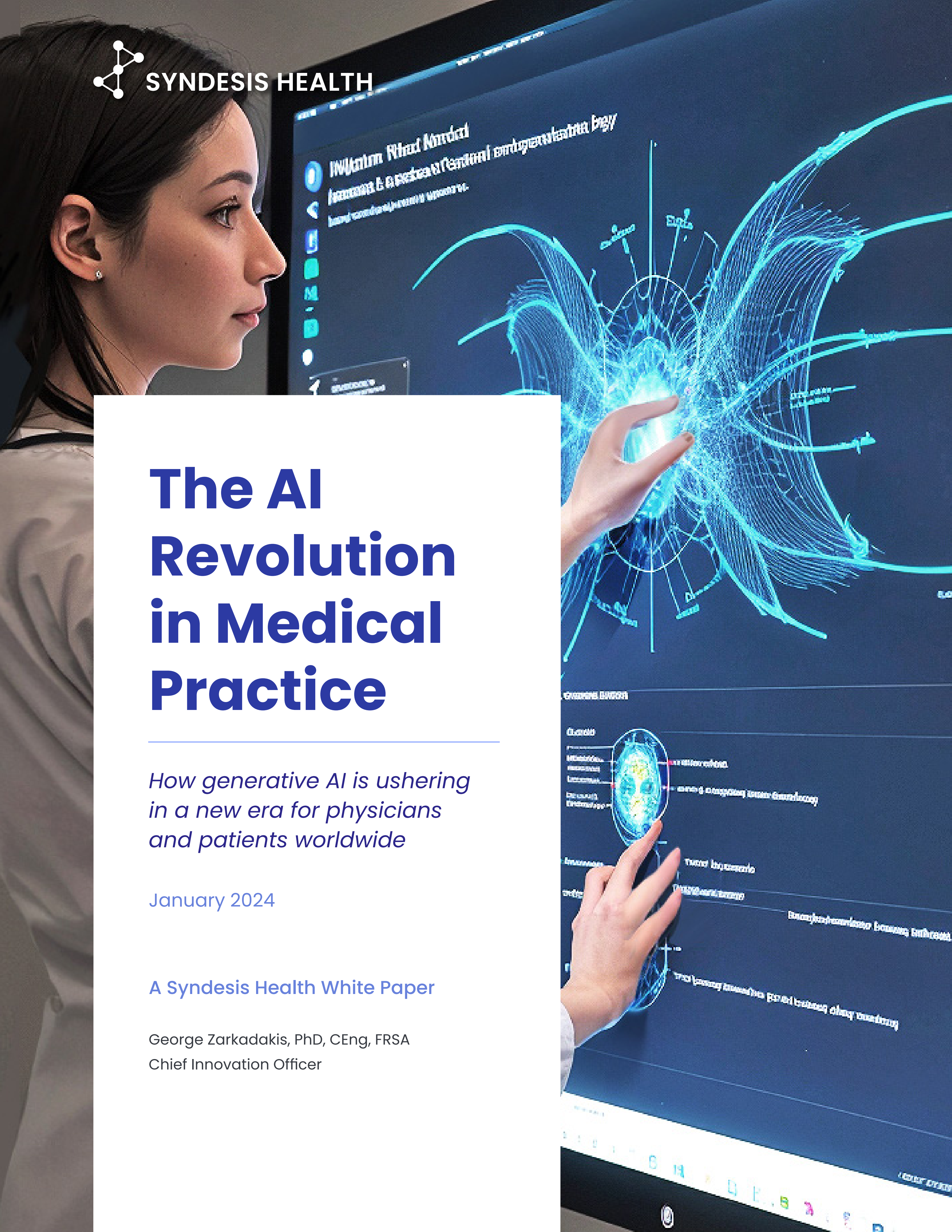


SYNDESIS HEALTH

The AI Revolution in Medical Practice

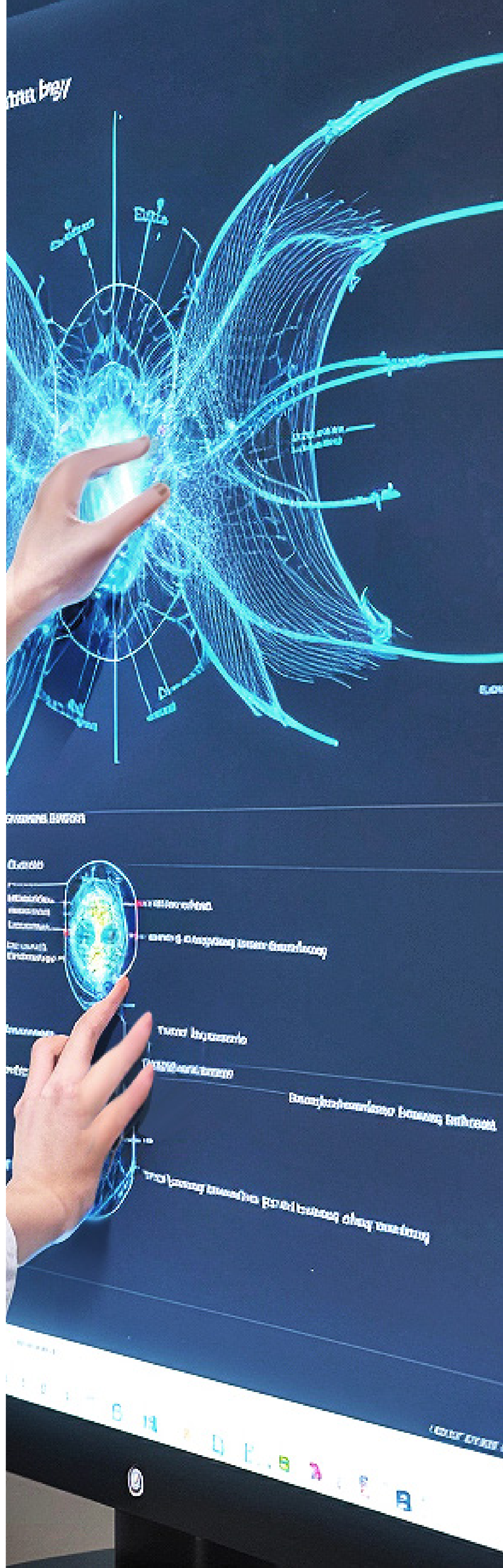
How generative AI is ushering in a new era for physicians and patients worldwide

January 2024

A Syndesis Health White Paper

George Zarkadakis, PhD, CEng, FRSA
Chief Innovation Officer

Table of contents



1. Why We Need AI in Medicine

Artificial Intelligence (AI) was born in the summer of 1956 during the now famous Dartmouth College conference with the ambition of simulating human-level intelligent behavior and critical thinking in computer systems. There have been many ups and downs since then, including several "winters" during which AI fell sorely short of its lofty ambition and, consequently, lost funding and talent to other scientific and engineering disciplines.

However, since the early 2000s, several major breakthroughs by pioneers, such as Geoffrey Hinton, Yann LeCun, and Yoshua Bengio, have demonstrated how a combination of highly sophisticated, statistics-based algorithms, high-speed parallel computing, and large volumes of data, could deliver the promise and dream of AI.

This new approach in AI mimicked the human brain and proposed "artificial neural networks" that learned how to perform pattern recognition on large data sets. Investment in AI soared after that historical inflection point, with the Big Tech behemoths of Apple, Meta, Google, and Microsoft spending tens of billions per year on AI system development.

Interestingly, by 2016 most of that private and public investment in AI was already flowing into healthcare applications, reflecting the huge potential, and expectation, of AI to transform medical practice and research[1]. In 2023 the AI in healthcare market was valued at USD 14.6 billion and is expected to grow by leaps and bounds to USD 102.7 billion by 2028[2].

Starting on the next page are the four main reasons - or "drivers"- that explain the phenomenal interest in AI applications for medicine:

In this white paper, we will:

1. Review how AI is addressing four big challenges in medicine
2. Explore some of the most prominent use cases and trends of medical AI systems
3. Dive into the ways the new generation of "generative AI" (such as LLMs) is ushering in new opportunities, now and in the future
4. Discuss the risks and challenges of integrating advanced AI systems into everyday medical practice
5. Provide a framework for how healthcare organizations may develop and execute successful strategies for AI adoption

[1] CB Insights Research. *Healthcare Remains the Hottest AI Category for Deals*, 2017. Accessed via: https://www.cbinsights.com/research/artificial-intelligence-healthcare-startups-investors/

[2] *Artificial Intelligence in Healthcare Market Report*, published by Markets and Markets, January 2023, Accessed via: https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-healthcare-market/

Medical Data Deluge

An enormous, complex, and exponentially growing volume of diverse data is demanding attention from medical researchers and practitioners. Readings from medical devices, such as sensors and ECGs, X-rays and other high resolution radiology images, Electronic Medical Records (EMRs), prescriptions and biometric data from wearable devices, lab tests, genomic and other multiomics information, are modern medicine's "big data". Despite the variability of these (usually) non-standardized data sources, physicians are required to somehow access, integrate and utilize this big data to improve patient outcomes (while complying with privacy regulations). These tasks are no longer humanly possible; however, the proper and ethical use of AI acting as a cognitive multiplier could "augment" human abilities. By bringing together and making sense of big medical data, AI can enable physicians to boost their productivity and expand their capabilities in diagnosis and disease management.

Skyrocketing Costs

High inflation, rising wages, clinical workforce shortages, and an aging population (particularly in developed countries) are putting increased pressure on national budgets for healthcare. According to PwC, medical costs are projected to rise 7% annually in the US, and a similar trend is apparent across a host of other countries too[3]. Medication cost trends are not slowing down either. Accelerated approvals of new cell and

[3] *Medical Cost Trend: Behind the Numbers 2024*, Report by PwC, Accessed via: _https://www.pwc.com/us/en/industries/health-industries/library/behind-the-numbers.html

gene therapies add further inflationary impact to pharmaceutical pricing. All those trends combine into a perfect storm of cost for healthcare systems, physicians, and patients. AI has the potential to slow down, and in many cases reverse, those rising costs by delivering cost efficiencies and higher productivity. For example, AI can be used to optimize the distribution of scarce medical resources by providing new analytical insights into cost management or by reducing the time required for new drug development and regulatory approval. A report by Harvard researchers has found that AI could save the US between 5% and 10% in healthcare spending, which translates to USD 360 billion annually if adopted more widely[4]. This number does not consider the personal and social benefit of better health outcomes for patients, or the impact on the broader economy of productivity lost due to illness.

Exponential Growth of Medical Knowledge

According to the US National Institutes of Health, the number of active, peer-to-peer reviewed clinical medical journals has expanded to approximately 28,000, collectively publishing more than 1.8 million articles per year only a few years ago, and constantly expanding further[5]. It is simply impossible for practicing doctors to be able to access and absorb all the knowledge that is produced annually at such a dazzling rate. Hence, many of the breakthrough studies published in these journals take years to be translated into everyday practice at the level of a hospital ward or a clinic. This lag between research results and clinical practice unfortunately translates into multiple lost opportunities for patients that may sometimes make the difference between life and death. As we will examine in more detail later in this paper, advanced forms of AI, such as Large Language Models (LLMs), can scour and process burgeoning medical literature and provide a highly efficient interface to medical knowledge.

Scarcity of Deep Medical Expertise

One of the main reasons for health inequity at national, regional, and global levels is the unequal distribution of deep medical expertise. AI can mitigate this obstacle by democratizing medical expertise. For example, AI can make available medical expert systems that compare patient cases with hundreds or thousands of other similar cases and advise doctors on possible diagnosis, as well as optimal disease management and therapy strategies.

[4] Sahni N et al, The Potential Impact of Artificial Intelligence on Healthcare Spending, (January 2023 – working paper), In: NBER, accessed via: https://www.nber.org/books-and-chapters/economics-artificial-intelligence-health-care-challenges/potential-impact-artificial-intelligence-healthcare-spending).

[5] Dai N, Xu D, Zhong X, Li L, Ling Q, Bu Z. Build infrastructure in publishing scientific journals to benefit medical scientists. Chin J Cancer Res. 2014 Feb;26(1):119-23. doi: 10.3978/j.issn.1000-9604.2014.02.10. PMID: 24653634; PMCID: PMC3937756.

2. Flowcharts vs Data: a Short History of Medical AI

The 1980s and 1990s saw a proliferation of "expert systems" that paved the way for AI's role in medicine. One such system, MYCIN[6], was able to identify bacteria causing severe infections and recommend antibiotics with the appropriate dosage based on a patient's body weight. Another noteworthy expert system was CADUCEUS (Figure 1) that could diagnose up to 1,000 different diseases. Systems such as these took a patient's history and symptoms as inputs, alongside clinical data, and used abductive reasoning to perform differential diagnosis in a way that a human doctor would. In other words, they followed a "flowchart-based" approach that translated the process of history-taking through a series of questions and then arrived at a probable diagnosis by combining the symptom complex presented[7].

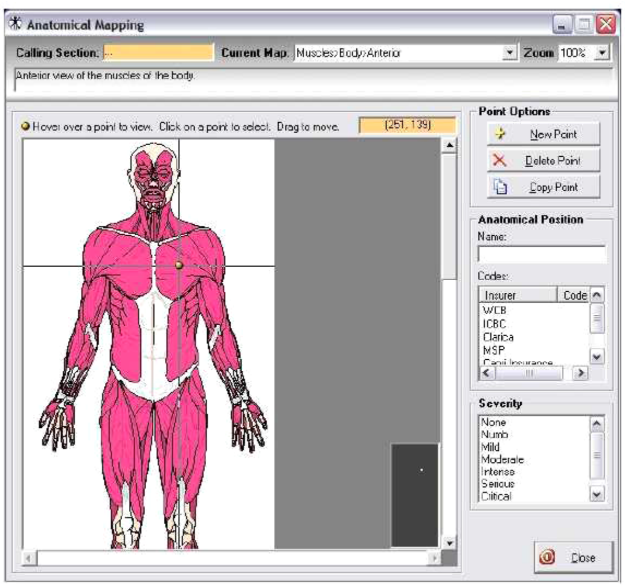


Figure 1. Caduceus Mapping Module, 2012

The outcomes of these early AI approaches to medical diagnosis were shown to be limited. This was because of the challenges of coding machines to recognize so-called "tacit knowledge", i.e., non-explicitly stated cues observed by human doctors during a patient encounter that guide their thinking processes. The limitations of those early "flowchart" approaches explains why none of these medical expert systems were adopted in regular medical practice.

The advent of deep learning techniques in the 2000s provided a different approach to automating medical diagnosis. Instead of trying to explicitly code the inference steps followed by human doctors, this new generation of AI systems utilized pattern recognition in large volumes of medical data. In effect, deep learning systems generated their own, internal inferences, which were then checked for correctness by human experts in a process called "supervised learning". This data-based approach to AI has been a game changer. Intelligent systems in medicine are performing much better than human doctors

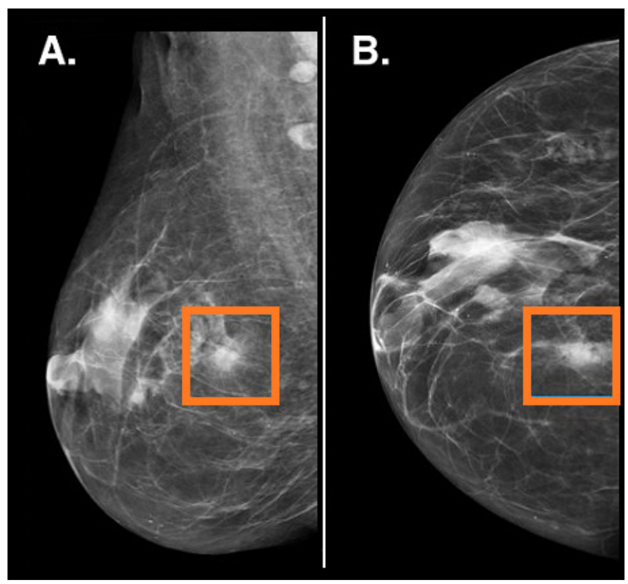


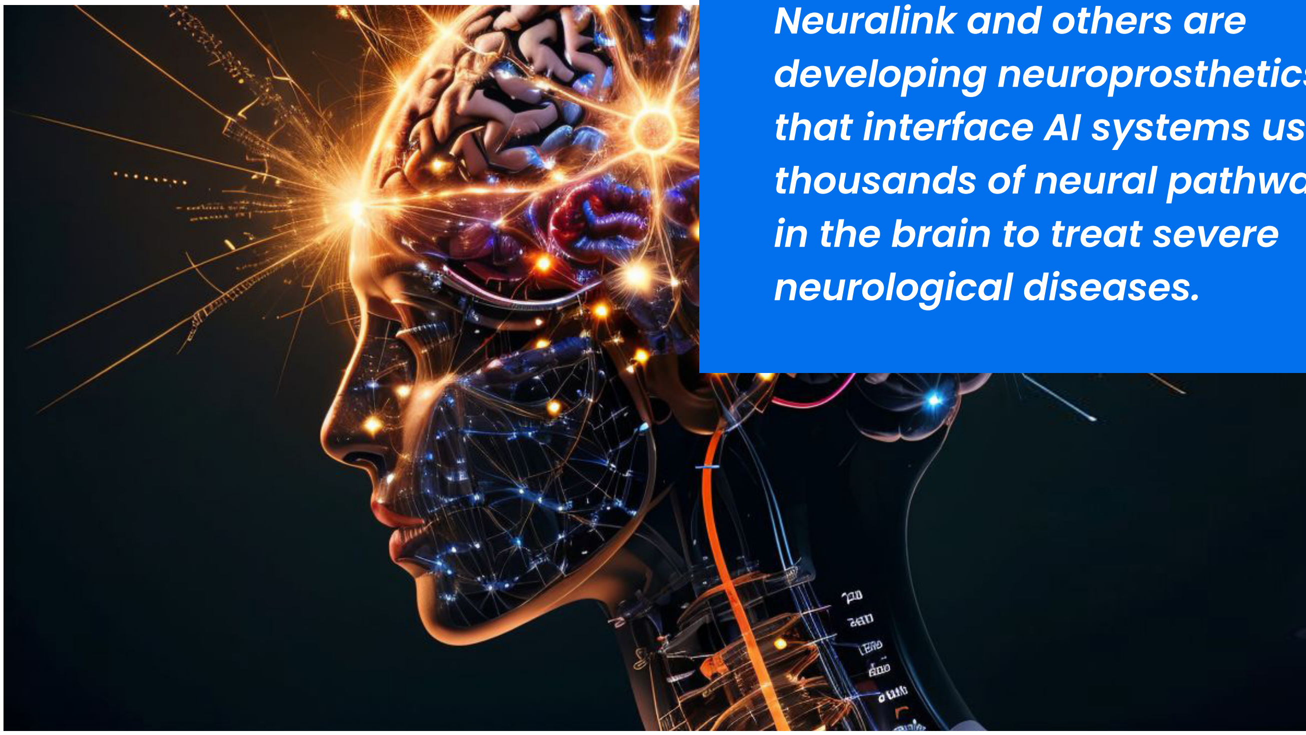
Figure 2. AI identification of tumors in mammograms, Lund University, Lund, Sweden, 2021.

in several cases. For example, in 2020, Google Health announced a deep learning system that was able to outperform human experts in breast cancer screening by cutting the number of people incorrectly referred for further screening with a false positive result by 5.7% - while also detecting 9.4% of potentially missed breast cancer cases[8].

[6] Clancey WJ, Shortliffe EH (1984). Readings in medical artificial intelligence: the first decade. Addison-Wesley Longman Publishing Co., Inc.

[7] Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. J Family Med Prim Care. 2019 Jul;8(7):2328-2331. doi: 10.4103/jfmpc.jfmpc_440_19. PMID: 31463251; PMCID: PMC6691444.

[8] McKinney, S.M., Sieniek, M., Godbole, V. et al. International evaluation of an AI system for breast cancer screening. Nature 577, 89–94 (2020). https://doi.org/10.1038/s41586-019-1799-6.



Neuralink and others are developing neuroprosthetics that interface AI systems using thousands of neural pathways in the brain to treat severe neurological diseases.

More recently, researchers from Lund University in Sweden showed how AI systems can work side-by-side with human doctors to improve their performance on patient outcomes. The researchers performed a randomized, controlled, clinical trial to determine whether an AI system could save radiologists time without endangering patients. This was the first study of AI's ability to diagnose breast cancer from mammograms whose design met the so-called gold standard for medical tests (Figure 2). Their human-plus-machine evaluation procedure enabled radiologists to spend substantially less time per patient while exceeding a baseline for safety[9].

The AI-assisted diagnosis achieved a cancer detection rate of 6.1 per 1,000 patients screened, comparable to the control method and above an established lower limit for safety. The radiologists recalled 2.0% of the control group and 2.2% of the experimental group, and both the control and experimental groups showed the same false-positive rate of 1.5%. The difference in recall rates coupled with the matching false-positive rate suggests that the AI method detected 20 percent more cancer cases than the manual method. Moreover, since approximately 37,000 patients were only examined by one radiologist, the results indicated that AI saved 44.3% of the examination workload without increasing the number of misdiagnosed patients.

AI systems are nowadays routinely used from online scheduling of appointments and check-ins in medical centers, digitization of medical records and reminder calls for follow up appointments, to drug dosage algorithms and adverse effect warnings while prescribing multi-drug combinations. For example, Google's Deep Mind is used by the UK's National Health Service to detect health risks through data collected by a mobile app. Microsoft's Hanover project analyses medical research to predict the most effective cancer drug treatment options for patients. Neuralink and others are developing neuroprosthetics that interface AI systems with thousands of neural pathways in the brain to treat severe neurological diseases, such as "locked-in syndrome" and stroke patients.

[9] Lang K et al, (2023), Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of randomized, controlled non-inferiority, single-blinded screening accuracy study. In: The Lancet Oncology, August 2023, DOI: https://doi.org/10.1016/S1470-2045(23)00298-X.

A plethora of other companies, large and small, are actively developing innovative AI systems across a range of medical applications, such as:

- **Disease diagnostics**
- **Medical treatment**
- **Surgical treatment, often in combination with medical robotics and prosthetics**
- **Health monitoring**
- **Digital consultation**
- **Managing medical data**
- **Personalized treatment**
- **Drug development**
- **Health plan analysis**

In all these applications, high quality medical data are key to training AI algorithms. Indeed, the rise of data-based AI has highlighted the increasing importance and scarcity of high-quality “Real Word Data” (RWD) and “Real World Evidence” (RWE) in shaping the future of AI systems and how to embed them in healthcare practice in an effective and ethical way. As more RWE is collected and made available, machine learning algorithms will adapt and allow for more robust responses and solutions in clinical decision support systems.

3. Generative AI: the New Paradigm

On November 30th, 2022, OpenAI released GPT-3, their biggest Large Language Model (LLM), immediately attracting global media attention and making "generative AI" a household word. In a nutshell, this new breed of AI system can "generate" some type of plausible content, such as text, images, audio, video, or computer code. These systems are also called "foundation models" because they are not confined to specific tasks but can be applied to numerous downstream tasks. They are therefore radically different from previous AI systems that were task-specific and fell under the rubric of "narrow AI". What makes generative, or foundation, AI systems different is a new approach in machine learning called "transformer" that was introduced by Google engineers in 2017[10].

A transformer model is a neural network that learns context – and thus meaning – by tracking relationships in sequential data, such as a series of words in a sentence. They differ from previous attempts to provide natural language responses to human questions, such as Recurrent Neural Networks (RNNs) that processed words sequentially only. By contrast, transformers process whole texts concurrently, and thus build a much better contextual understanding of words. This way transformers can distinguish different meanings of the same word; for example, "interest", which can mean either curiosity or money.

LLMs, such as GPT-4, Claude, or BERT, are trained on enormous data sets that are widely available on the Internet. They break down words into "tokens" that those systems can then process internally to build a list of statistical values (called "vector") that dynamically understand the probability of a token being present, or absent, in the same sentence or other sentences nearby. These values are then used to tune the learning parameters of the model that define the behavior of the AI system and influence how it processes input and formulates output. GPT-4 by OpenAI has 1.5 trillion parameters, BARD by Google 1.6 trillion parameters, while LLaMA by MetaAI (which is

[10] Vaswani, A. et al. Attention is all you need, (2017). In Advances in Neural Information Processing Systems (eds Guyon, I. et al.) 30, 5998–6008.

Quick Guide to LLMs in Medicine

OpenAI ChatGPT - General LLM

Anthropic Claude - General LLM

PubMedBERT - A version of BERT (Bidirectional Encoder Representations from Transformers) fine-tuned on biomedical literature, specifically on articles from PubMed.

BioBERT - Similar to PubMedBERT, BioBERT is a version of BERT fine-tuned on biomedical text, including scientific literature and clinical notes.

ClinicalBERT- A version of BERT tailored for clinical text, developed by researchers to understand clinical narratives and extract medical information.

BlueBERT - A pre-trained BERT model trained on clinical text, designed to handle medical and clinical language

BERT-MIMIC - A variant of BERT trained on the MIMIC-III database, a large electronic health records dataset, to improve performance on clinical natural language processing tasks.

BERT-Medical - Another BERT variant fine-tuned on clinical notes to capture medical context and terminology.

UMLS-Transformer - A transformer-based model designed to map clinical text to concepts in the Unified Medical Language System (UMLS).

MedAlpaca 13b - A large language model specifically fine-tuned to improve question-answering and medical dialogue tasks.

GatorTron - Largest clinical language model.

Med-PaLM - Clinical Knowledge model.

BioGPT - Research Literature model.

AlphaFold - Research for Drug Discovery focusing on protein folding.

MedBERT - A model trained on large amounts of medical and clinical text data. Very useful for medical Naural Language Processing (NLP) tasks like medical information extraction, medical question answering, etc.

BioMegatron - Large biomedical model trained on PubMed abstracts and full text articles. Good for biomedicine text mining.

Bio_Discharge_Summary_BERT - Trained on hospital discharge summaries. Helpful for processing clinical narrative texts.

SciFive - Trained on scientific papers across multiple disciplines including biomedicine. Useful general scientific NLP model.

ClinicalXLNet - Clinical version of XLNet pretrained on medical notes, clinical trials data, etc.

exhibiting multimodal capabilities) has 1.2 trillion parameters. This massive parameter size allows the respective model to capture intricate patterns and dependencies within language, resulting in improved text generation, understanding, and contextual coherence. Presently, LLMs use transformers to process texts up to 25,000 words long, which means they can handle whole research papers, or chapters of books, with very impressive results. For example, by drawing on medical texts present in their training data sets, LLMs, such as ChatGPT or Flan-PaLM, can accurately answer medical questions and achieve passing scores on the US Medical Licensing Exam[11].

Several technology providers are developing LLMs that focus on medicine (see side bar on previous page: "Quick Guide of LLMs in Medicine") and many of them are already affecting how doctors improve patient lives. Triage is one such example. Given that 89% of patients search online for their symptoms before seeking healthcare guidance or connecting with their provider, LLMs may also leverage those searches in performing virtual triages, and thus support and improve triage decisions in a clinical setting. Another area that is already benefiting from LLM solutions is the provision of "expertise on tap", i.e., a way for clinicians to get a "second expert opinion" quickly when faced with a complex patient case. Making specialized medical expertise readily available to clinicians everywhere at the tap of a button can help solve heath inequity across the world.

The results from using LLMs in medical diagnosis are very encouraging. ChatGPT, which is the user-friendly version of GPT-3, is shown to provide the correct diagnosis when asked using plain language 87% of the time, beating both human physicians as well as previous, non-AI, symptom checkers[12]. Moreover, a recent paper tested an evolved LLM system called MedPALM-2 that specializes in medical diagnosis and found that it surpassed ChatGPT and Flan-PaLM, and exceeded the passing score of US Medical Licensing Exam style questions with a score of 86.5%[13]. These are remarkable findings that demonstrate very significant improvements in AI systems over a very short period.

LLMs have the potential to be real game changers in medical diagnosis. Given that language is at the heart of health and medicine, underpinning interactions between patients and healthcare practitioners, we are clearly seeing a new era emerge whereby patients will be routinely health-checked by AI systems using a wide range of data, including real-time data from wearables or medical devices, clinical data from EHRs, unstructured text and image data, as well as genomic and multiomic data, linked to advancing medical knowledge as published in research papers. In the future, an AI "personal physician" will be able to monitor a person's health and well-being, and predict potential illnesses before they manifest, or help a patient manage an existing illness better, with significant positive outcomes for patients and healthcare systems. We are at the beginning of a very exciting new era in medicine where AI systems are evolving from "narrow" problem-solving into "generalist" AI.

[11] Singhal, K. et al, (2022), Large language models encode clinical knowledge. Preprint at https://arxiv.org/abs/2212.13138.

[12] Ruth Hailu, Andrew Beam and Ateev Mehrotra. ChatGPT-assisted diagnosis: Is the future suddenly here? In: StatNews (February 13, 2023). Accessed via: https://www.statnews.com/2023/02/13/chatgpt-assisted-diagnosis/

[13] Singhal K et al, (2023), Towards expert-level medical question answering with Large Language Models, In: https://arxiv.org/pdf/2305.09617.pdf)

4. Towards a Generalist Medical AI (GMAI)

The Holy Grail of AI, since its foundation in 1956, has been to develop "Artificial General Intelligence" (AGI), i.e., systems that can fulfill three key human-like characteristics of higher-order intelligence. Firstly, they need to be generalist systems that can be applied across a wide number of tasks, as opposed to systems built for specific tasks in a narrow knowledge domain. Secondly, they need to be capable of "common sense", meaning they have the ability to contextualize meaning depending on the circumstances. Thirdly, and most importantly, they must be able to "transfer learning"—to apply lessons learned in the past from any knowledge domain to solve for novel problems in other knowledge domains. The advent of transformers and LLMs has been the first great leap towards AGI. For the first time, we now have systems that can satisfy the first two of the three required characteristics. Recently, an LLM system called Gato demonstrated the third capability too, also referred to as "multi-modality". Gato can chat, caption images, play video games and control a robot arm, and thus it can be regarded as a truly "generalist" AI system[14].

It is not difficult to imagine Gato-like generalist AI systems in the medical domain that will interpret a combination of medical modalities, such as imaging, EHR data, genomics, graphs, or medical texts[15]. Such systems would be capable of advanced medical reasoning and provide free text explanations, spoken recommendations and image annotations at scale.

For GMAI systems to exist, we will need four main breakthroughs

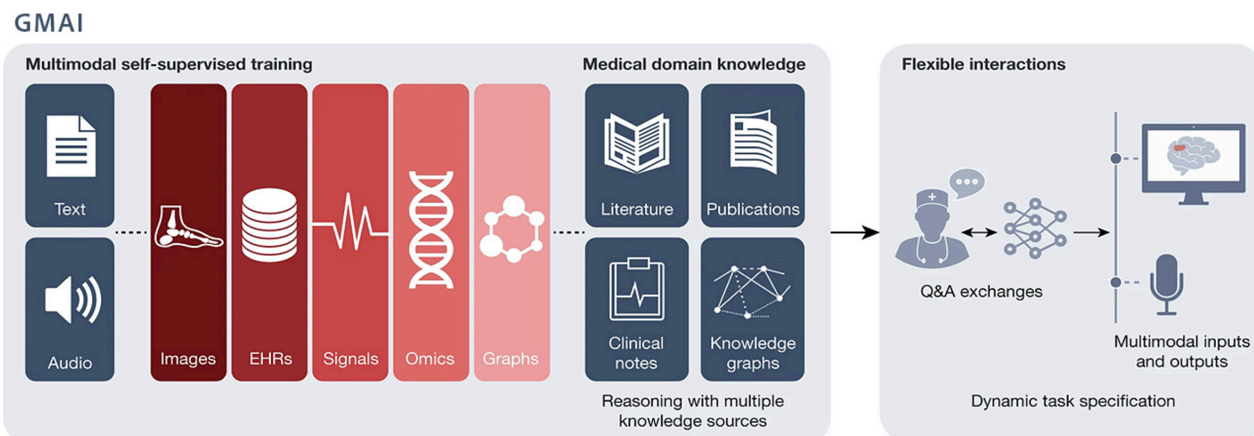
1. **Multi-Modal Architectures** that bring together a multitude of types of medical and clinical data into a single source that can be accessed by GMAIs.
2. **Self-Supervised Learning** whereby GMAI systems will not need additional training when confronted with a novel problem, but will be able to self-learn in context. Such breakthrough is necessary so that GMAI systems can seamlessly adapt to technology change, for instance when a new X-ray machine is commissioned so that they are able to evaluate the new images.
3. **In-Context Learning for Multiple Modalities.** Currently we have LLMs that can process language "tokens" and are capable of in-context learning in text only. GMAIs will have to support flexible combinations of data modalities by utilizing blended stream tokens, such as ones combining images and text.
4. **Formal Representation of Medical Reasoning.** For GMAIs to perform advanced medical reasoning they need to have a formal representation of medical knowledge in the form of knowledge graphs or other techniques from symbolic AI, in conjunction with neural networks. This "hybridization" of GMAIs should correct one of the current problems in the outputs of LLMs called "hallucinations" whereby the systems generate improbable falsehoods. By applying rigorous medical formalisms, GMAIs will be able to check the plausibility and veracity of their outputs, and thus generate reliably factual statements.

[14] Reed, S. et al. A generalist agent. In Transactions on Machine Learning Research (2022).

Figure 3, below, shows how a GMAI model pipeline would look[15]. These models will need to be trained across a wide range of data modalities, such as images, EHR data, labs, multiomics, etc. They must also be capable of contextual self-learning by blending tokens from various modalities, e.g., audio, images, video, and text. Following that, the outputs will be checked against formal reasoning with multiple knowledge sources of medial knowledge, including medical textbooks, research papers, clinical notes, and knowledge graphs. This approach will furnish GMAIs with the ability to have dynamic and highly flexible interactions with medical doctors. For example, a doctor would be able to provide their reasoning on a patient case, as well as relevant data from various sources, and engage in a natural language conversation with the system, as they would with an expert human colleague. The range of applications for those future GMAI systems would be wide. Indeed, they will be used to build systems for interactive note taking, augmented surgical procedures, bedside decision support, chronic disease management applications and more. As a promising early example of a GMAI, CheXzero can detect dozens of diseases in X-rays without being trained on explicit labels for these diseases but rather by learning from natural language reports in clinical notes[16].

Figure 3.

Pipeline for a General Medical Artificial Intelligence (GMAI) System



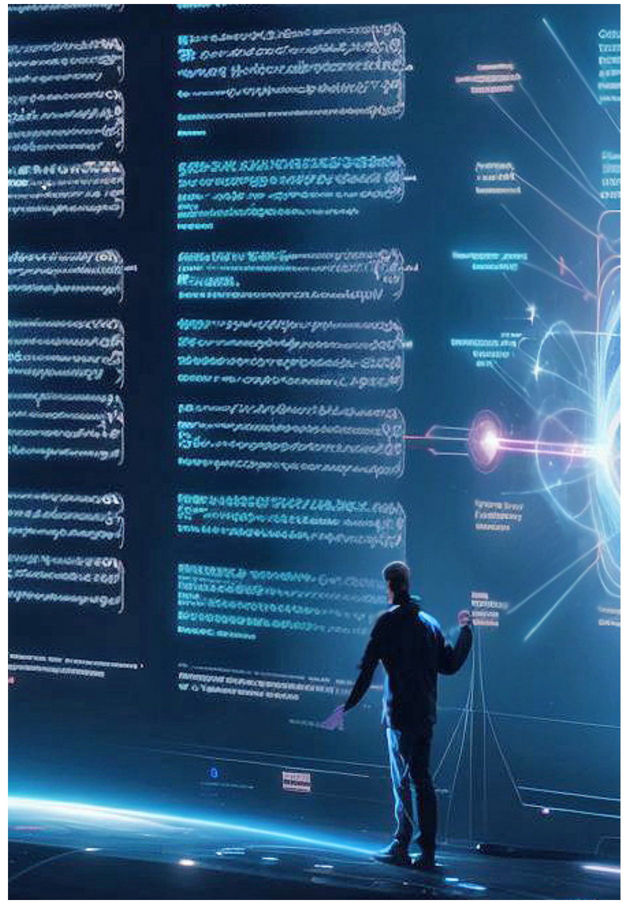
The GMAI is trained on multiple medical modalities, such as images, EHR data and multiomics. Formal medical reasoning is then applied using multiple knowledge sources, such as research papers, clinical notes, and knowledge graphs. As the GMAI self-learns in a medical context it becomes capable of flexible and dynamic interactions with human doctors.[15]

[15] Moor, M., Banerjee, O., Abad, Z.S.H. et al. Foundation models for generalist medical artificial intelligence. Nature 616, 259–265 (2023). https://doi.org/10.1038/s41586-023-05881-4.

[16] Tiu, E. et al. Expert-level detection of pathologies from unannotated chest X-ray images via self-supervised learning. Nat. Biomed. Eng. 6, 1399–1406 (2022).

5. Challenges and Risks

LLMs are still in their infancy, and many issues regarding access, validation, regulation, safety, and ethics must be resolved before they are widely adopted in healthcare practice. For instance, LLMs require a huge amount of data as well as massive computing power processing capabilities for their training. These requirements translate to only a handful of big companies being able to develop those highly sophisticated systems, thus posing barriers on how accessible advanced AI systems are, by whom, and under what terms of use. Moreover, the need of multimodal medical data for their training poses growing challenges on how we protect patient privacy while serving the greater social good. The advent of GMAIs is promising to revolutionize medicine and improve patient outcomes, but only if the challenges and the risks that those systems introduce can be resolved satisfactorily. Let's examine some of these challenges and discuss potential strategies for remedying the harmful use or misuse of AI systems in medicine.



Need for Curated, Multimodal, and Transparent Data

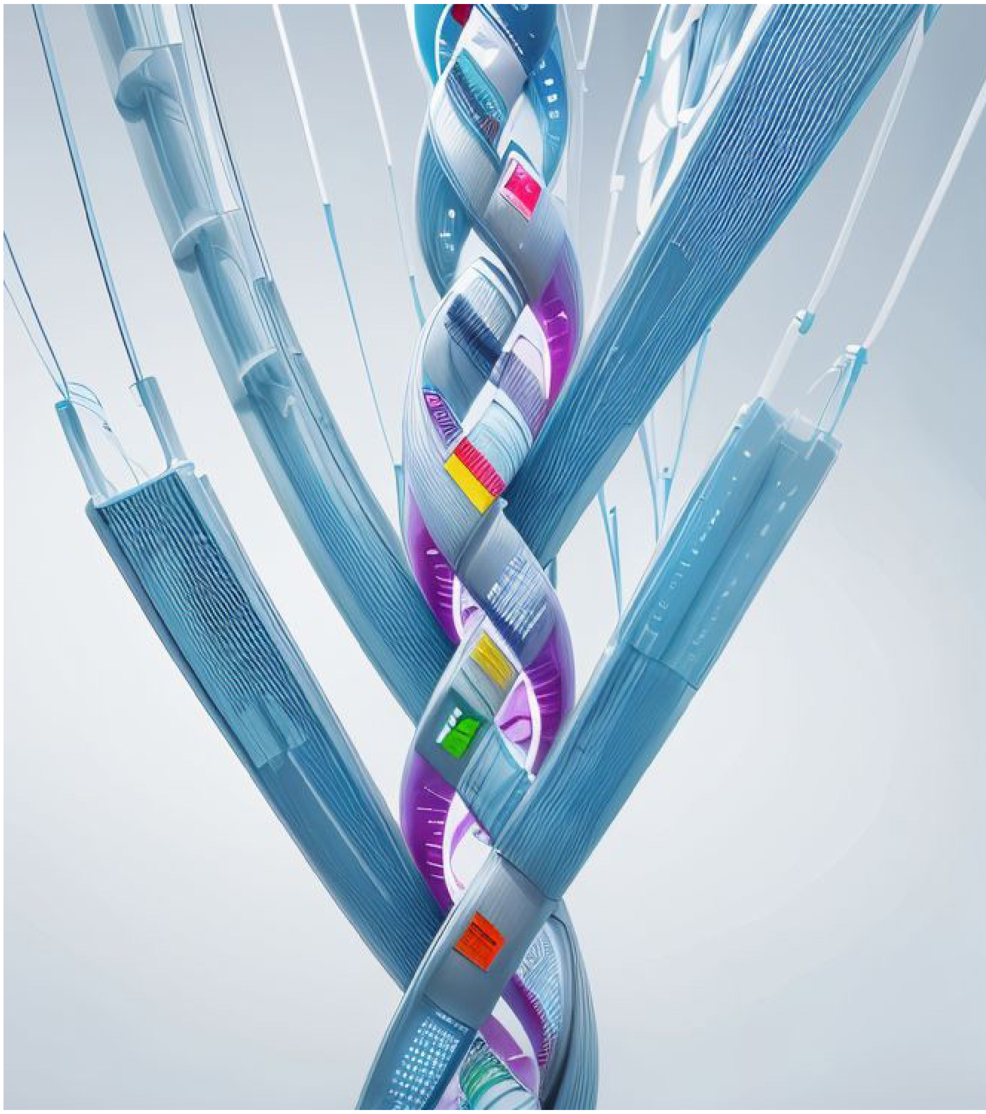
In mid-2023, a group of scientists at Microsoft published a paper that explored the behavior of GPT-4 and argued that the system exhibited signs of Artificial General Intelligence (AGI)[17]. The critique that the paper received was mostly about the non-transparent data that the system used for its training. Given that the training data set was proprietary to OpenAI, who developed the system, the publication's claims were impossible to peer review, as the latter would require access to the data sets to duplicate the paper's results. This event highlighted one of the key problems in the current state of development of LLMs, and soon of GMAIs as well. Training data sets must become transparent and accessible so that these systems can be independently reviewed and assessed. Specifically for medical applications, it is essential that multimodal, well-curated medical data are available, with rigorous data governance applied so that patients' rights are withheld and protected. Across all cases, the data must be able to be reviewed and rectified for implicit social biases, so that the resulting AI systems do not produce outputs that are prejudiced against minority groups. Given the diversity in human phenotype and genotype, medical AI systems must be trained on diverse medical data sets for best results that consider all human patients. By pre-training GMAIs in a diverse and multimodal corpus, it will be possible for these systems to perform effective and efficient in-context learning and adapt to the geographically and ethnically diverse environments of their application.

AI Equity

The costs of developing, training, and maintaining LLMs are enormous. Each LLM requires thousands of Graphics Processing Units, or GPUs, offering the parallel processing power needed to handle the

[17] Bubeck S et al, Sparks of Artificial General Intelligence: Early experiments with GPT-4, [Submitted on 22 Mar 2023 (v1), last revised 13 Apr 2023 (this version, v5)], https://arxiv.org/abs/2303.12712

massive datasets these models learn from. The cost of the GPUs, alone, can amount to millions of dollars. According to a technical overview of OpenAI's GPT-3 language model, training required at least USD $5 million worth of GPUs[18]. These models require many training runs as they are developed and tuned, so the final cost is far greater than this figure. Up to 10 gigawatt-hour (GWh) power consumption is needed to train a single large language model like ChatGPT-3. This is, on average, roughly equivalent to the yearly electricity consumption of over 1,000 U.S. households. The electricity alone used to train GPT-3 is estimated to have cost USD 100 million.[18]



Training is getting more expensive as we move towards GMAIs. These high costs have created a global oligopoly of advanced AI systems with enormous political power. It is important for global development that all nations have equal access to these systems. For this to take place, more competition is necessary, for instance by making foundation models open-source. An example of this is RETFound, a pioneering open-source foundational model developed by the team at Moorfields and UCL in London. By learning generalisable features, RETFound can diagnose eye diseases with ten times less labelled data than previous approaches, and even anticipate systemic illnesses like heart disease[19]. The model's availability means it could serve as new public infrastructure, and researchers from around the world could fine-tune it to craft their own diagnostic tests. RETFound's inception was facilitated by access to the NHS's healthcare data.

Alignment of Goals

LLMs and GMAIs are extremely complex systems that build their own internal reasoning mechanisms based on their training data. Unlike traditional software systems where a software engineer can review the code for errors and take corrective action, if necessary, these highly advanced AI systems are virtual "black boxes" whose emergent behaviour is essentially unpredictable and can only be observed from the outside. To validate and verify the safety of such systems, we need controlled experiments and rigorous testing protocols like those used in clinical

[18] Li C, Open AI's GPT-3 Language Model: A technical Overview, (June 3, 2020), Lamba Labs, Accessed via: https://lambdalabs.com/blog/demystifying-gpt-3

[19] Zhou, Y., Chia, M.A., Wagner, S.K. et al. A foundation model for generalizable disease detection from retinal images. Nature (2023). https://doi.org/10.1038/s41586-023-06555-x

trials[19] and the inclusion of "red teams" to identify potential weaknesses, vulnerabilities, and areas for improvement. Such protocols require the combined skillsets of clinicians, data scientists, AI engineers, as well as protocol designers and ethicists. We also need a common language with precise and shared terminology to facilitate communication among this wide range of experts. Moreover, a big question is being debated around whether those future, super intelligent systems, will continue to have goals aligned with the goals of human society, or not. Will GMAIs aspire to improve, and save, patient lives by "abstaining from all intentional wrong-doing and harm", as every junior doctor vows when taking the Hippocratic Oath? Or will their goals and ours diverge? It is important to understand that this central dilemma in advancing Artificial General Intelligence (AGI) rests on the premise of "system autonomy"; in other words, on the assumption that these AI systems will be capable of developing their own goals, possibly independently of humanity. To mitigate this risk – often referred to as the "existential risk" of AI - it is imperative that we limit the autonomy of GMAIs, and that we build these systems to work alongside human doctors, rather than to replace them.

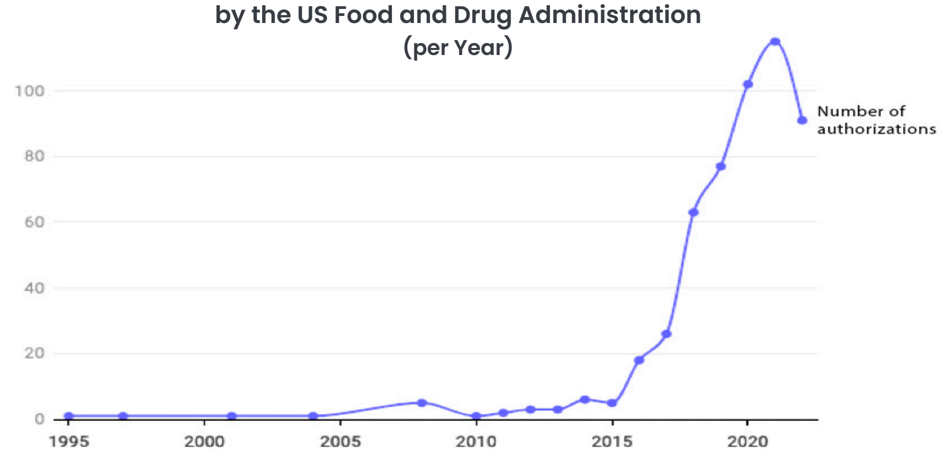
Regulation and Ethics

The need for multi-modal data to power GMAIs, as well as their potential to upend current practices in medicine, creates a host of new ethical issues to grapple with. Patient privacy must be respected and always maintained with the use of de-identified data or synthetic training data sets only, and strict data governance must be applied when those systems access data, whether centralized or federated. Current technology enables data to be accessed remotely in "clean rooms" where data policies can be enacted so that data and machine learning pipelines comply with local and international regulations. These technologies will need to advance further to consider AI systems with the ability to build their own pipelines.

With that future possibility in mind, the proposed Artificial Intelligence Act by the European Union aims to classify AI systems according to "risk", and to strengthen existing rules on data quality, transparency, human oversight, and accountability. Nevertheless, enforcing any kind of regulation will be a challenge due to the sheer volume of AI systems expected to proliferate in the years to come.

Regulators are already struggling with an exponential rise of AI in medicine (see Figure 4). As of January 2023, the US Food and Drugs Administration (FDA) had approved 520 AI and ML algorithms, most of them (75%) in medical imaging and computer vision. The FDA first approved the use of AI for medical purposes in 1995, and only 50 other algorithms were approved over the next 18 years. Yet between 2019 and 2022, over 300 were approved, with a further 178 granted FDA approval in 2023 so far. This illustrates the increasing size

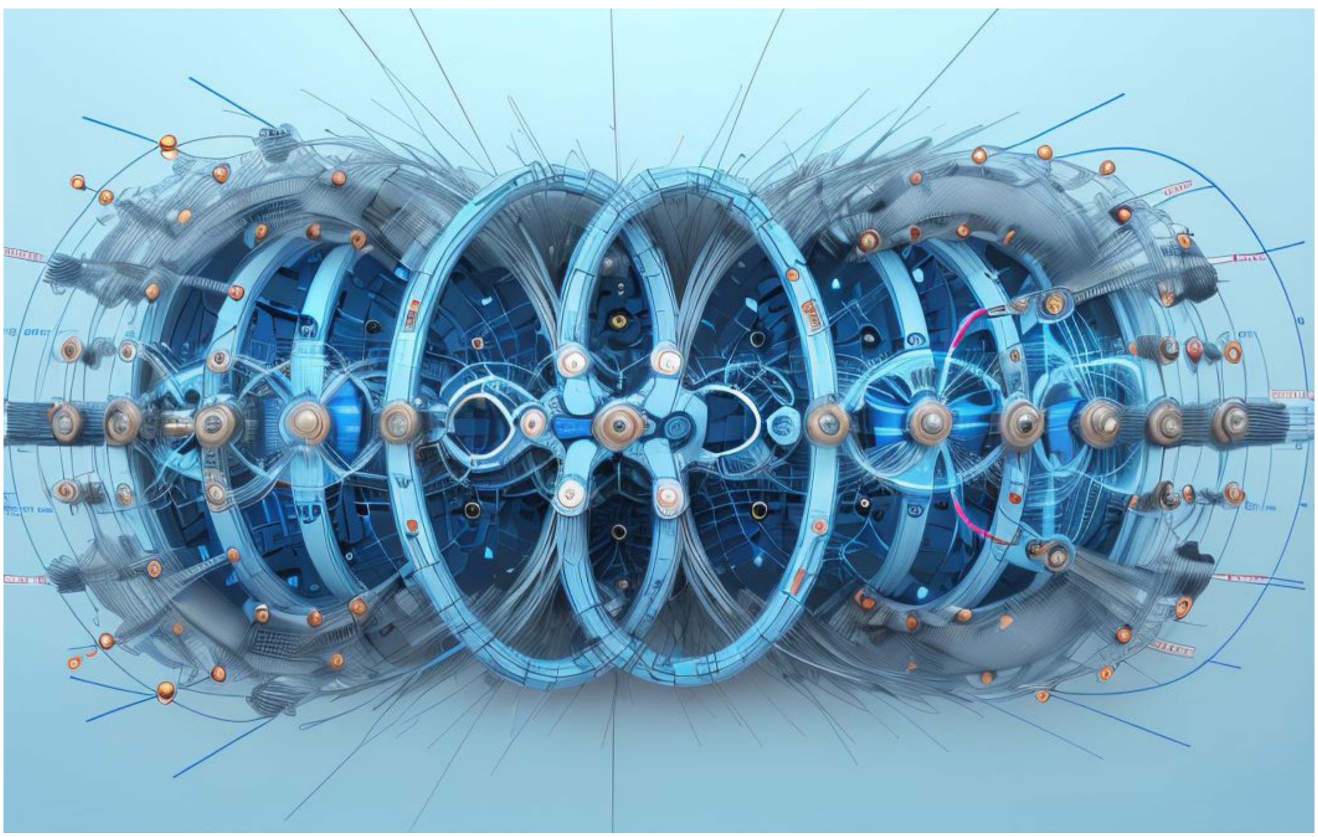
Figure 4. **Number of Approvals and Clearances by the US Food and Drug Administration (per Year)**



Source: FDA

of the task, which will only get bigger in the years to come. This will especially become true with AI systems that will have a high degree of autonomy in making clinical decisions, in which case they will be regarded as "medical devices" by the FDA and "high risk" by EU's AI Act, and as such they will have to be regulated.

Indicatively, under the US 21st Century Cures Act, most software and AI tools are exempt from FDA regulatory approval "as long as the healthcare provider can independently review the basis of the recommendations and doesn't rely on it to make a diagnostic or treatment decision." Ensuring that human doctors remain in the loop of clinical practice will ensure not only alignment of goals between AI and humans, but also the faster adoption of GMAIs by healthcare organizations across the world[20]. As GMAIs come online, regulators will recognize that it is a Sisyphean task to try and regulate the unpredictable. This realization will hopefully introduce more innovation in regulation practices, for example by establishing "regulatory sandboxes" for GMAIs where real-world experiments can be carried over and their implications observed before a regulation is enacted.



[20] *The Presidio Recommendations on Responsible Generative AI, June 2023*, World Economic Forum, accessed via: https://www.weforum.org/publications/the-presidio-recommendations-on-responsible-generative-ai/

6. A Strategy for AI in Medicine

AI is advancing at an incredible rate and affecting medical practice in multiple profound ways. It is vital for healthcare organizations to take the initiative and establish successful strategies for adopting AI in an ethical and appropriate way. They must exploit the enormous opportunities that technology provides for reducing costs and improving patient outcomes while mitigating the associated risks. Healthcare organizations should consider a "three pillar" AI strategy, as shown in Figure 5, that takes the following aspects into consideration:

AI Leadership

Adopting a continuously improving technology in any organization requires the shaping of a shared vision, as well as long-term commitment to translating that vision into actionable steps and milestones. It is imperative that a healthcare organization has an executive AI champion who will lead the development of an AI culture, support teams in the execution of tasks, ensure that failures are seen as opportunities for learning, and generally steer the organization towards a rate of AI adoption that is focused on the organization's needs. This executive sponsor must be surrounded by a team of operational leaders who are experts in various fields, including medicine, medical ethics, information technology, systems architecture, privacy and security, compliance, and user experience. This team will provide the necessary collective leadership for the successful adoption of AI.

Data, Systems and Pilots

This is the first pillar for executing on the shared AI vision. Data is the fuel of the AI revolution, and nothing can be achieved without them. Healthcare organizations should invest in EHR systems and establish well-documented processes so that clinicians adhere to the discipline of updating information in those systems and the use of international standards (e.g., ICD) to code diagnoses and procedures. Organizations should not limit themselves to clinical data, but invest in secure data infrastructures for all the data that flows in their operations, including labs, radiology, multiomics, billing, finance, etc. There is a plethora of cloud-based AI systems available that organizations can use in conjunction with their data to explore applications and conduct pilots. AI leadership should decide what pilots to prioritize based on the areas that have the best available data and promise the highest return on investment (ROI). For example, an organization may wish to start using chatbots to improve patient experience, or an AI system to perform scheduling.

Skills

This second pillar is pivotal in developing an AI culture across the organization. Medical doctors must be trained in understanding and using analytics tools across multiple data for analysis and insights. The future of the medical profession appears to be one where doctors will be working alongside GMAI systems. For instance, imagine a scenario of a surgical team using an AI system in a procedure and posing it queries like: "We cannot find the intestinal rapture. Check whether we missed a view of any intestinal section in the visual feed of the past 15 minutes" [15]. GMAI systems would then carry out visualization tasks, annotating video streams, read relevant medical literature, and advise the surgical team in seconds when encountering rare anatomical phenomena. GMAIs will also be used in training doctors, in conjunction with Virtual or Augmented Reality equipment.

Healthcare organizations are encouraged to become early adopters of such novel training methods as they will accelerate the cultivation and nurturing of a shared AI culture.

Networks and Ecosystems

This third pillar recognizes that no organization can be successful on their own and emphasizes the importance of becoming part of a wider ecosystem and collaborating with other organizations that share a common vision. For LLMs and GMAIs to deliver on their promise, we need massive medical data sets from various modalities that can come about only through innovative data sharing solutions that allow cross-border and cross-institutional data sharing. Such solutions can enable the centralization of de-identified data from various institutions and countries, as well as the creation of synthetic data using these central or federated repositories. Platform-enabled ecosystems are the most appropriate for power-boosting the AI vision and ambitions of a healthcare organization. Such platforms usually provide ways to share and use data across organizations ethically and securely, as well as sandboxes for quickly and easily piloting applications in collaboration with other data providers and innovative technology partners[21].

Ethics & Regulatory Compliance

This is the foundation of a successful AI strategy for healthcare organizations. The role of Hospital Ethics Committees (HEC) should be expanded to include AI systems and to ensure compliance with national and international ethical and regulatory guidelines[22]. They should support the determination of pilots and resolve ethical challenges that may occur. Additionally, HECs may expand their role to include direct engagement with patients and patient advocacy groups, so that the end beneficiaries of AI systems are included and involved in shaping the future of medical AI systems too.

Figure 5.

[21] *Zarkadakis G, Data Trusts could be the key to better AI, Harvard Business Review, Nov 10, 2020, accessed via: https://hbr.org/2020/11/data-trusts-could-be-the-key-to-better-ai*

[22] *Hajibabaee F, Joolaee S, Cheraghi MA, Salari P, Rodney P. Hospital/clinical ethics committees' notion: an overview. J Med Ethics Hist Med. 2016 Dec 18;9:17. PMID: 28523118; PMCID: PMC5432947.*

7. Conclusions

AI systems in medicine have finally come of age and are transforming medical practice around the world. Foundation models are showing signs of "Artificial General Intelligence."

As such, it is expected that soon the first "General Medical Artificial Intelligence" (GMAI) systems will become available, capable of processing multiple data modalities, supporting clinicians in a variety of essential tasks, reducing health inequity by making high-quality healthcare available to many more patients at a lower cost, and minimizing the time spent by clinicians on administrative tasks so that they may dedicate more time to their patients.

Nevertheless, there are several challenges and risks in adopting AI systems in everyday clinical practice. Foremost is the decision on how far those systems should be allowed to go in making autonomous decisions about patients, and thus "replacing" human doctors in various application areas or circumstances. The current trend in regulating AI systems in the US and EU considers fully autonomous AI systems as "high risk", and as such those systems require protocols of very rigorous testing, akin to controlled clinical trials, to gain regulatory approval. It is therefore likely that most AI systems in medicine will be "augmenting" rather than "replacing" human doctors. This is something to be welcomed and celebrated. Leaders in healthcare organizations should double down on their efforts to cultivate an AI culture in their organization and lead a shared vision for the execution of a comprehensive AI strategy.

About Syndesis Health

Syndesis Health accelerates research and innovation for healthcare and life sciences companies. The company was founded on the potential for real-world clinical data to advance medical research, improve patient health outcomes, and inform healthcare policy decisions. The Syndesis Health Network enables secure collaboration amongst member organizations and is powered by Syntium, the Syndesis data platform. Visit https://syndesis.com to learn more.

About the Author

George Zarkadakis is the Chief Innovation Officer at Syndesis Health. He has a PhD in Artificial Intelligence for Medicine, and several years of experience in business consulting, Artificial Intelligence, startups, and media. He is the author of ***In Our Own Image**: **will Artificial Intelligence Saves Us or Destroy Us?***, (Rider Books, 2016), and ***Cyber Republic**: **Reinventing Democracy in the Age of Intelligent Machines***, (MIT Press, 2020).

SYNDESIS HEALTH