Chapter 2
AI in Medicine: Some Pertinent History

Edward H. Shortliffe and Nigam H. Shah

After reading this chapter, you should know the answers to these questions:

- What are the roots of artificial intelligence in human history, even before the general introduction of digital computers?
- How did computer science emerge as an academic and research discipline and how was AI identified as a component of that revolution?
- How did a medical focus on AI applications emerge from the early general principles of the field?
- How did the field of cognitive science influence early work on AI in Medicine (AIM) and how have those synergies evolved to the present?
- What were the early medical applications of AI and how were they received in the clinical and medical research communities?
- How has the focus of medical AI research and application evolved in parallel with AI itself, and with the progress in computing power, communications technology, and interactive devices?
- To what extent are the early problems and methods developed by early AIM researchers still relevant today? What has been lost and what has been gained?
- How have the advances in hardware and the availability of labeled data made certain forms of AI popular? How can we combine these recent advances with what we learned from the previous 40 years?
- How might we anticipate the further evolution of AI in medicine in light of the way the field has evolved to date and its likely trajectory?

E. H. Shortliffe (✉)
Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA
e-mail: ted@shortliffe.net

N. H. Shah
Stanford University School of Medicine, Stanford, CA, USA
e-mail: nigam@stanford.edu

Introduction

The history of artificial intelligence in medicine (AIM) is intimately tied to the history of AI itself, since some of the earliest work in applied AI dealt with biomedicine. In this chapter we provide a brief overview of the early history of AI, but then focus on AI in medicine (and in human biology), providing a summary of how the field has evolved since the earliest recognition of the potential role of computers in the modeling of medical reasoning and in the support of clinical decision making. The growth of medical AI has been influenced not only by the evolution of AI itself, but also by the remarkable ongoing changes in computing and communication technologies. Accordingly, this chapter anticipates many of the topics that are covered in subsequent chapters, providing a concise overview that lays out the concepts and progression that are reflected in the rest of this volume.

Artificial Intelligence: The Early Years

As was discussed in Chap. 1, AI is a diverse field that addresses a wide variety of topics regarding human intelligence and expertise, with an emphasis on how to model and simulate these topics in computer systems. Thus studies of how human beings reason are part of AI, but so are the creation of devices (such as robots) that incorporate human-like features. Viewed in this framework, notions relevant to AI emerged early in human history as people studied the workings of the human mind or imagined creations that might duplicate those capabilities.

For example, fantastical non-human intelligent entities were imagined as far back as Greek mythology. Hephaestus was a mythical blacksmith who manufactured mechanical servants, and there were even early tales that involved the concept of intelligent robots. But perhaps the most important early harbinger of AI was Aristotle's invention of **syllogistic logic** (a formal deductive reasoning system) in the fourth century BC.

Mechanical inventions that attempted the creation of human-like machines are known to have existed as early as the thirteenth century, when talking heads were created as novelty items and Al-Jazari, an Arab inventor, designed what is believed to be the first programmable humanoid robot (a boat carrying four mechanical musicians, powered by water flow). There are many other examples that could be mentioned from periods prior to the twentieth century.[1]

---

[1] For more discussion, see "A Brief History of AI" at https://aitopics.org/misc/brief-history (accessed August 13, 2022) and "History of Artificial Intelligence" at https://en.wikipedia.org/wiki/History_of_artificial_intelligence. (accessed August 13, 2022).

In the early twentieth century Bertrand Russell and Alfred North Whitehead published *Principia Mathematica*, which revolutionized formal logic [1]. Subsequent philosophers pursued the logical analysis of knowledge. The first use of the word "robot" in English occurred in a play by Karel Capek that was produced in 1921.[2] Thereafter a mechanical man, Electro, was introduced by Westinghouse Electricat at the New York World's Fair in 1939 (along with a mechanical dog named Sparko). It was a few years earlier (1936–37) that Alan Turing proposed the universal **Turing Machine** concept and proved notions of computability.[3] Turing's analysis imagined an abstract machine that can manipulate symbols on a strip of tape, guided by a set of rules. He showed that such a simple machine was capable of simulating the logic of any computer algorithm that could be constructed. Also relevant (in 1943) were the introduction of the term **cybernetics**, the publication by McCulloch and Pitts of *A Logical Calculus of the Ideas Immanent in Nervous Activity* (an early stimulus to the notion of **artificial neural networks**) [2], and Emil Post's proof that **production systems** are a general computational mechanism [3].

Especially important for AI was George Polya's 1945 book *How to Solve It*, which introduced the notion of **heuristic** problem solving [4]—a key influential concept in the AI community to this day. That same year Vannevar Bush published *As We May Think*, which offered a remarkable vision of how, in the future, computers could assist human beings in a wide range of activities [5]. In 1950, Turing published *Computing Machinery and Intelligence*, which introduced the **Turing Test** as a way of defining and testing for intelligent behavior [6]. In that same year, Claude Shannon (of **information theory** fame) published a detailed analysis showing that chess playing could be viewed as **search** (*Programming A Computer to Play Chess*) [7]. The dawn of computational artificial intelligence was upon us as computers became viable and increasingly accessible devices.

Modern History of AI

The history of AI, as we think of it today, began with the development of **stored-program** digital computers and the ground-breaking work of John von Neumann and his team at Princeton University in the 1950s. As the potential of computers

---

[2] Čapek K. *Rossumovi Univerzální Roboti* (Rossum's Universal Robots). It premiered on 25 January 1921 and introduced the word "robot" to the English language and to science fiction as a whole. https://en.wikipedia.org/wiki/R.U.R. (accessed August 13, 2022).

[3] Turing submitted his paper on 31 May 1936 to the London Mathematical Society for its *Proceedings*, but it was published in early 1937. https://en.wikipedia.org/wiki/Turing_machine. (accessed August 13, 2022).

began to be appreciated, academic engineering scientists began to pursue concepts that would evolve to be known as computer science. The history and capabilities of AI have subsequently been tied to the evolution of computers and their associated technologies.

As is mentioned in Chap. 1, it was at a conference at Dartmouth University in 1956 that a group of early computer scientists gathered to discuss the notion of simulating human reasoning by computer. One attendee, John McCarthy from Massachusetts Institute of Technology (MIT) (who later spent most of his professional life at Stanford University), coined a name for the developing field: *artificial intelligence*. At Carnegie Mellon University (then known as Carnegie Tech), psychologist Allen Newell, economist/psychologist Herbert Simon, and systems programmer (from the Rand Corporation) John Clifford Shaw introduced the Logic Theorist system[4]—arguably the first AI program—which was followed by their General Problem Solver in 1957.[5] At about the same time (1958), Frank Rosenblatt invented the **perceptron algorithm** at the Cornell Aeronautical Laboratory [8]. This introduced the notion of **connectionism** in AI, where networks of circuits or connected units were used to simulate intelligent behavior.

The notion of **machine learning** was first explored by Arthur Samuel (IBM) between 1958 and 1962 [9]. He developed a checker-playing program that learned strategy and novel methods by having it mounted on two machines and then having it play against itself thousands of times—resulting in a program that was able to beat the world champion. Another key development during that era (1958) was John McCarthy's creation of the **LISP** programming language[6]—which dominated as the basis for AI research and development for several decades (including in the medical AI community).

During the 1960s there was an explosion in creative AI work, initially at MIT and Carnegie Mellon, but later in the decade at other universities in the US. International explorations of AI were also underway, especially in the United Kingdom (where the first **Machine Intelligence** workshop was held in Edinburgh in 1966). By the end of the decade, as early computer science departments began to be formed, AI groups began to appear more broadly (with notable efforts underway at the University of California Berkeley and Stanford University). The first industrial robot company was formed (1962) and a series of influential AI PhD

---

[4] http://shelf1.library.cmu.edu/IMLS/MindModels/logictheorymachine.pdf. (accessed August 13, 2022).

[5] http://bitsavers.informatik.uni-stuttgart.de/pdf/rand/ipl/P-1584_Report_On_A_General_Problem-Solving_Program_Feb59.pdf. (accessed August 13, 2022).

[6] McCarthy's original paper is available at http://www-formal.stanford.edu/jmc/recursive.html. (accessed August 13, 2022).

dissertations emerged—particularly at MIT where the students of Marvin Minsky had a huge impact on the evolving field [10]. Also noteworthy was the invention of the mouse pointing device by Doug Engelbart at Stanford Research Institute (SRI[7]), which was to revolutionize the way in which human beings would interact with computers. In 1969, also at SRI, scientists developed "Shakey", a mobile robot that had a problem-solver embedded in addition to locomotion (wheels) and perception (cameras with image processing).[8] The first International Joint Conference on Artificial Intelligence (IJCAI) was held in Washington, DC in 1969. Meanwhile, that same year at MIT, Minsky and Seymour Papert published *Perceptrons* [11], an influential book that discussed the computational approach that Rosenblatt had introduced a decade earlier, outlining the limits of what perceptrons could do. This led to a decrease in interest in connectionist concepts and arguably held back the pace of development for what eventually became known as neural networks in the 1980s and in turn led to today's **deep learning** approaches (see Chap. 6).

AI research topics in the 1960s seem remarkably similar to those that dominate today. Machine learning, natural language processing, speech understanding, image analysis, robotics, and simulation of human problem solving were all major areas of research focus. Much of the funding for such research in the US came from the **Department of Defense (DOD)**, which envisioned eventual military applications of AI but provided extensive support for basic methodology development that had no immediate military application. The DOD also supported communications research, which in turn became a great facilitator of AI development work. Perhaps most notable was the introduction of a nationwide network for interconnecting major research computers that were located at academic institutions and in research centers for military contractors. The DOD's **Advanced Research Projects Agency (ARPA)[9]** supported much of the AI and communications research in the country. This network for research computers, was built on the notion of **packet switching** and became known as the ARPA Network or, simply, the **ARPAnet**. Collaborative AI research among universities became heavily dependent on this network, and the notion of electronic messaging among researchers across the various sites evolved into the email that we take for granted today. Similarly, the ARPAnet, and its packet switching technology, were eventually taken over by the National Science Foundation (NSF) and,

---

[7] See https://en.wikipedia.org/wiki/Douglas_Engelbart. (accessed August 13, 2022). SRI became an independent entity outside of Stanford University and is known today simply as SRI International.

[8] https://www.sri.com/hoi/shakey-the-robot/. (accessed August 13, 2022).

[9] Also often called **DARPA**, for Defense Advanced Research Projects Agency.

in turn, became a coordinated independent entity that is today known as the Internet.

AI Meets Medicine and Biology: The 1960s and 1970s

As AI was developing as a research discipline, it is not surprising that some of the challenging problems that attracted investigators were drawn from biomedical science. An early example from 1965 was MIT work by Joseph Weizenbaum who was exploring **chatbot** technology (conversational natural language processing and response generation; see Chap. 9). He developed a program known as "The Doctor", but more affectionately referred to as "Eliza", which attempted to provide psychiatric assessments of patients. The focus was on maintaining the conversation intelligently rather than actually reaching a psychiatric diagnosis. The program became a popular, easy-to-use "toy" at AI centers since it was available for conversations over the ARPAnet, and it did respond in ways that suggested, at some level, that it understood what the user was saying. A few years later, at Stanford, a psychiatrist on the medical school faculty, Ken Colby, worked with AI researchers to develop a conversational program, known as "Parry", that would simulate the behavior of a patient with paranoid schizophrenia. He undertook the work largely for educational purposes, and his students and residents enjoyed "interviewing" the program to learn about its thought disorder and to try to keep the "patient" from shutting down and refusing to communicate further. Of course, as Parry became known in the AI community, it was inevitable that people would begin to wonder how Eliza would handle a therapeutic session with Parry. Accordingly, in 1972, an ARPAnet link was created between Eliza at MIT (Cambridge, MA) and Parry at Stanford (Palo Alto, CA). Without human intervention, the two programs had a conversation,[10] and this somewhat hysterical match-up has become part of AI lore [12].

Emergence of AIM Research at Stanford University

A more serious and ground-breaking AI research effort in biomedicine was the Dendral Project at Stanford University. It began as an effort developed by a remarkable scientist, Joshua Lederberg, who had been attracted to Stanford as founding chair of their Department of Genetics in the late 1950s. He arrived shortly after receiving the Nobel Prize in Physiology or Medicine (at age 33!) for his ground-breaking work, at the University of Wisconsin, on genetic transfer between bacteria. Then, in the mid-1960s, a young researcher, Edward Feigenbaum, joined the faculty in Stanford's nascent computer science department, arriving from UC Berkeley

---

[10] See https://tools.ietf.org/html/rfc439 for a transcript of the interchange between the two programs. (accessed August 13, 2022).

after studying with Herbert Simon at Carnegie Tech (Carnegie Mellon University today). Lederberg and Feigenbaum teamed up with Carl Djerassi, an eminent professor in the Chemistry Department, who was an expert in organic and hormonal chemistry and who had been instrumental in the development of birth control pills a decade earlier.

Lederberg was himself an excellent programmer (in addition to his skills as a geneticist) who became fascinated with the challenge of determining organic compound structures from mass spectral data—a task mastered by very few organic chemists. He wondered if there might be a computational solution and felt that the first requirement was to consider all the possible structures consistent with a compound's chemical formula ($C^aH^bO^c$, where the superscripts indicate the number of carbon, hydrogen, and oxygen atoms in one molecule of the compound). As the number of atoms in a compound increases, the number of potential structures becomes very large. Lederberg developed an algorithmic approach, which he called the "dendritic algorithm,"[11] and wrote a program that could generate the entire exhaustive set of potential structures for any organic compound. Pruning that large space to define a couple of likely structures was guided by mass spectral analysis (**mass spectroscopy**) of the compound, and it was in this area that Djerassi had special expertise. With the addition of Feigenbaum and other computer scientists to the team, the Dendral Project thus sought to encode the rules used by organic chemists who knew how to interpret mass spectra in order to infer the small number of structures, from among all those generated by the dendritic algorithm, that were consistent with the spectral data. The focus on knowledge representation and the use of **production rules**, plus the capture and encoding of expertise, placed this early work solidly in the AI arena.

Another key contributor to this work in the early years was Bruce Buchanan,[12] a research scientist with computing expertise and formal training that included a PhD in Philosophy of Science. He stimulated and participated in efforts to view the Dendral work as research on **theory formation**. Although the system was initially based solely on rules acquired from Djerassi and other experts in interpretation of the mass spectra of organic compounds, Buchanan and others pursued the possibility that it might be possible to infer such rules from lots of examples of mass spectra and the corresponding compounds of known structure. This machine learning approach, which greatly enhanced the Dendral program's performance over time as new rules were added, became known as Meta-Dendral.

By the early 1970s, Dendral had become well known in computer science circles [13] and the biomedical focus had spawned methods that generalized for use in other domains—a phenomenon that was to occur many times in subsequent decades as biomedicine became a challenging real-world stimulus to novel approaches that were adopted broadly by AI researchers in areas beyond medicine. DENDRAL also

---

[11] The name was inspired by the expanding network of possible solutions that reminded him of a neuron's dendrites.

[12] See Buchanan's foreword to this volume.

spawned a dynamic research environment at Stanford, linking the school of medicine with the university's young computer science department. As other projects were developed that focused on capturing biomedical expertise in computer programs, Feigenbaum generalized the efforts in an overriding principle that had guided much of the work:

> *The key empirical result of DENDRAL experiments became known as the knowledge-is-power hypothesis (later called the Knowledge Principle), stating that knowledge of the specific task domain in which the program is to do its problem solving was more important as a source of power for competent problem solving than the reasoning method employed.*—Edward A. Feigenbaum, 1977 [14].

The process of capturing and encoding expert knowledge became known as **knowledge engineering.** See Chap. 4 for a focused discussion on knowledge-based systems, their subsequent evolution, and the current status of such work.

As DENDRAL grew and new projects were started at Stanford, it became clear that the computing facilities available for the research work were too limited. Furthermore, other medical AI projects were underway at a handful of other institutions and most researchers working on medical AI problems were feeling similar computational constraints. Lederberg accordingly submitted a successful proposal to the **Division of Research Resources (DRR)** at the **National Institutes of Health (NIH)**. He envisioned a major computing facility that would support medical AI research, not only at Stanford but at other universities around the US. The resulting shared resource was also granted one of the few remaining available connections to the ARPAnet—the first computer on the network that was not funded by the DOD. In this way the computer could be used by researchers anywhere in the country, using their own local connections to the ARPAnet to provide them with access to the computational power available at Stanford.[13] This shared computing resource, installed on the Stanford medical school campus in 1973, was known at the Stanford University Medical Experimental Computer for Artificial Intelligence in Medicine, more commonly referred to as **SUMEX-AIM**, or simply SUMEX. With grant renewals every 5 years, SUMEX served the national (and eventually the international) AI in Medicine community for 18 years.[14] With the departure from Stanford of Dr. Lederberg (who became President of Rockefeller University in New York City in the mid-1970s), Feigenbaum took over as Principal Investigator of SUMEX-AIM for several years.

---

[13] Since local area networking did not yet exist, most connections to the ARPAnet relied on dial-up modems with **acoustic couplers**. The network had local phone numbers for **terminal interface processors** (known as TIPs), scattered around the country, so investigators could generally access the network, and hence the computer at Stanford, with a local phone call.

[14] See "The Seeds of Artificial Intelligence. SUMEX-AIM." Published in 1980 by the Division of Resarch Resources at NIH. https://eric.ed.gov/?id=ED190109. (accessed August 13, 2022).

Three Influential AIM Research Projects from the 1970s

The notion of using computers to assist with medical diagnosis often traces its roots to a classic article that was published in *Science* in 1959 [15]. It was written by two NIH physician-scientists, one a dentist (Robert Ledley) and the other a radiologist (Lee Lusted). The paper laid out the nature of **Bayesian probability theory** and its relevance to medical diagnosis, arguing that computers could be programmed to assist with the Bayesian calculations and thus could serve as diagnostic aids. They acknowledged the challenges in deriving all the necessary probabilities and recognized the problem of conditional dependencies when applying Bayes' theorem for a real-world problem like medical diagnosis. However, their work stimulated a number of research projects that sought to use probability theory for diagnosis, with especially influential projects by Homer Warner and colleagues at the University of Utah [16] and by Timothy deDombal and his team at Leeds in the United Kingdom [17].

It was the challenges with statistical approaches, and their lack of congruence with the way in which human experts solved similar problems, that led scientists to consider whether AI methods might not be adapted for such clinical decision making problems. Three AIM research efforts from the 1970s are particularly well known and played key roles in the evolution of the field. Unlike DENDRAL, these projects were focused on clinical medicine, and two of them were created using the SUMEX-AIM resource. All three programs were envisioned as potential sources of consultative decision support for clinicians as they thought to diagnose and/or manage patients.

INTERNIST-1/QMR

One of the early SUMEX projects was developed over the ARPAnet from the University of Pittsburgh. There an esteemed physician leader, Dr. Jack Myers, had stepped down as Chair of Medicine and in the early 1970s became interested in sharing his clinical knowledge and experience in a novel way (rather than writing "yet another textbook"). Renowned as a master clinician and diagnostician, and a past President of the American College of Physicians and Chairman of the American Board of Internal Medicine, he collaborated with an MIT/Carnegie Tech-trained computer scientist, Harry E. Pople, Jr., PhD. Randolph A. Miller, then a second year Pitt medical student who had learned to program in machine language and a higher level language while in high school, joined the project in its second year. They worked together in an effort to create a program that would assist in the diagnosis of adult patients with problems whose diagnoses fell in the realm of internal medicine.
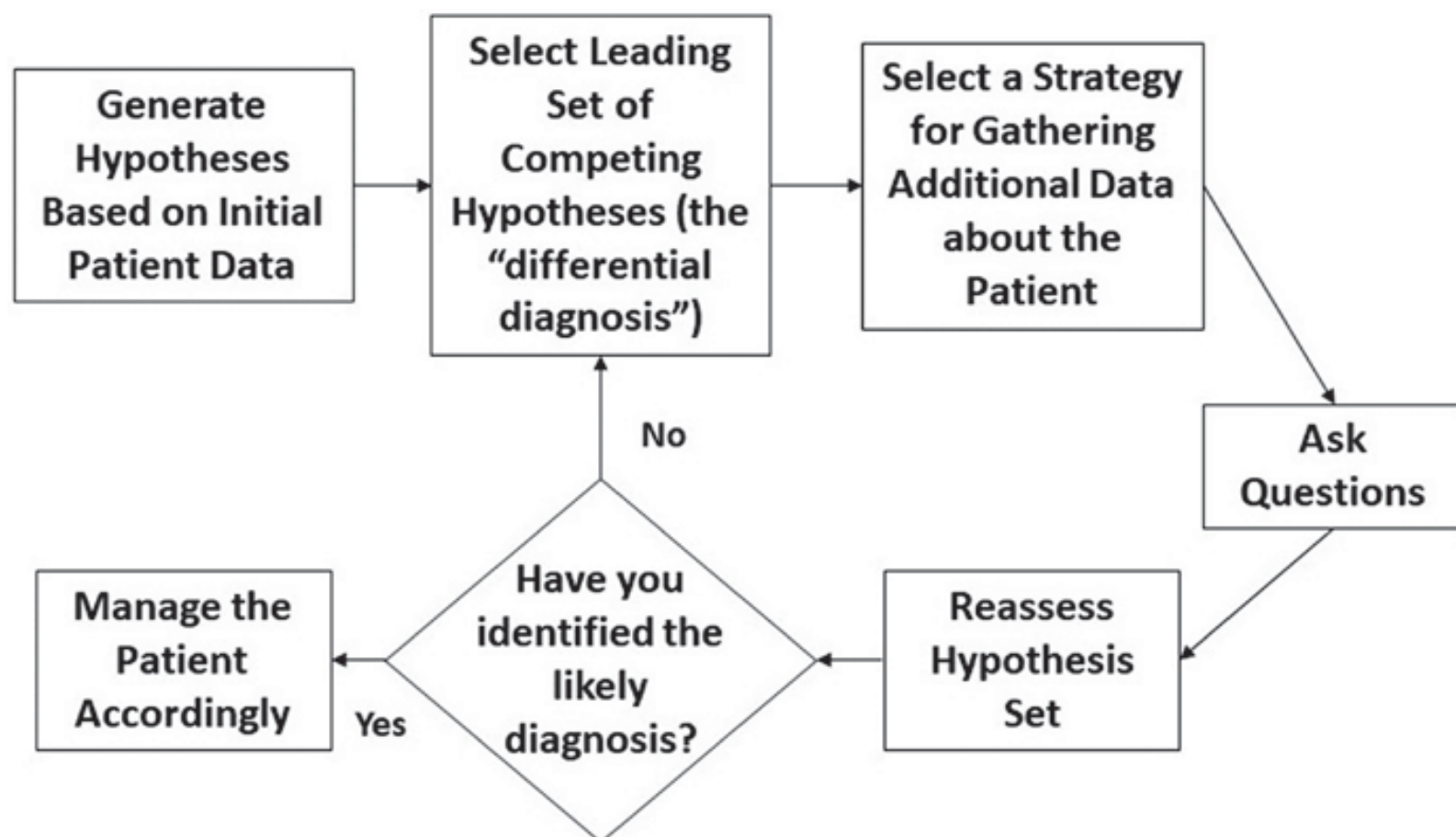
Fig. 2.1 The *hypothetico-deductive approach*, as applied to medical diagnosis. The Internist-1 program implemented these general notions in a program that tackled the diagnosis of essentially all diseases in internal medicine

The basic notion behind *INTERNIST-1*[15] was that it should be possible to simulate by computer the **hypothetico-deductive approach** that cognitive studies had shown were often used by expert clinicians as they attempted to diagnose challenging cases (Fig. 2.1) [18]. Myers invited medical students, including Miller and others, to spend medical school elective time conducting intensive analyses of the peer-reviewed literature on a disease topic of their choosing, which was then augmented by Myers' own experience. They thus characterized 650 disorders in internal medicine using 4500 possible patient descriptors. Miller took a sabbatical research year, working full time with Pople and Myers in 1974–75 to write the INTERNIST-1 Knowledge Base (KB) editor program.

Miller's programming enabled Pople's diagnostic algorithms to access and manipulate a KB that otherwise exceeded the computer system's available address space. The team developed a computational algorithm that used presenting history, symptoms, physical exam findings, and lab results from a patient to generate a set of diagnoses that could potentially explain the patient's problems. They also created a refinement process that selected a strategy and identified additional questions that would allow the program to distinguish between competing hypotheses and to generate new ones.

INTERNIST-1 could accurately diagnose many difficult cases. In addition, it could deal with multiple concurrent disorders in the same patient. It was ultimately tested with some of the most difficult diagnostic challenges in the clinical literature

---

[15] The community often referred to the program simply as *Internist*, although this simpler name was legally unavailable for ownership/copyright reasons.

(Clinical Pathological Conferences published in the *New England Journal of Medicine*) where it correctly diagnosed more of the cases than did the physicians who had actually cared for the patients [18].

While the evaluation of INTERNIST-1 showed the potential of the heuristic AI approach to assist human beings with diagnosis, it also uncovered a number of shortcomings that showed that the system was not suitable for widespread clinical use [19]. After Miller joined the Pitt faculty, he observed that INTERNIST-1 was of great interest to medical students and faculty clinicians. However, it was also clear that the system was impractical to use—especially because it required the user to take an hour or more to enter all information about the patient, and then to respond to queries from the program. Recognizing this, he decided that the most useful element of INTERNIST-1 was its knowledge base.

Beginning in 1983, he began working on a different approach to diagnostic assistance—one that recognized the human clinician-user was the most knowledgeable intelligence in the diagnostic consultation process. The doctor knew the patient far better than the computer system could. The new diagnostic assistant system, Quick Medical Reference (QMR), ran on the newly available personal computers. Miller felt that QMR should support the clinician's problem-solving as efficiently as possible. He worked with colleagues to develop QMR as a toolkit to assist clinicians with about a dozen specific diagnostic assistance tasks, which the user could select individually or chain together serially to address the dilemmas that had puzzled them. The user could invoke QMR quickly on a personal computer in the office. For example, it allowed questions such as "What is the differential diagnosis of finding x?" or "How can I best screen a patient for disease y?" QMR allowed the user to rank and influence the differential diagnosis produced, and to determine the mode for generating questions, in a way that had not been possible with INTERNIST-1. Eventually, over the course of a decade, QMR was marketed as a commercial product.

One lesson of this work, and other medical systems to be described shortly, was that consultative decision aids were not likely to be used if they did not integrate well into clinicians' existing workflow [19] (see also Chap. 17 for more discussion of this issue). It took the revolution in networking and electronic health records, which introduced new ways of accessing pertinent patient data, for such programs to be more realistically used, even though their early capabilities were impressive.

CASNET

Another center of excellence for research on medical AI in the 1970s was based at Rutgers University in New Brunswick, New Jersey. Their computer science department, chaired by Saul Amarel, had recruited a young faculty member, Casimir Kulikowski, who had applied his computer science expertise to medical problems during his training and early postdoctoral work. Amarel and Kulikowski successfully proposed a second computing resource for applied artificial intelligence in
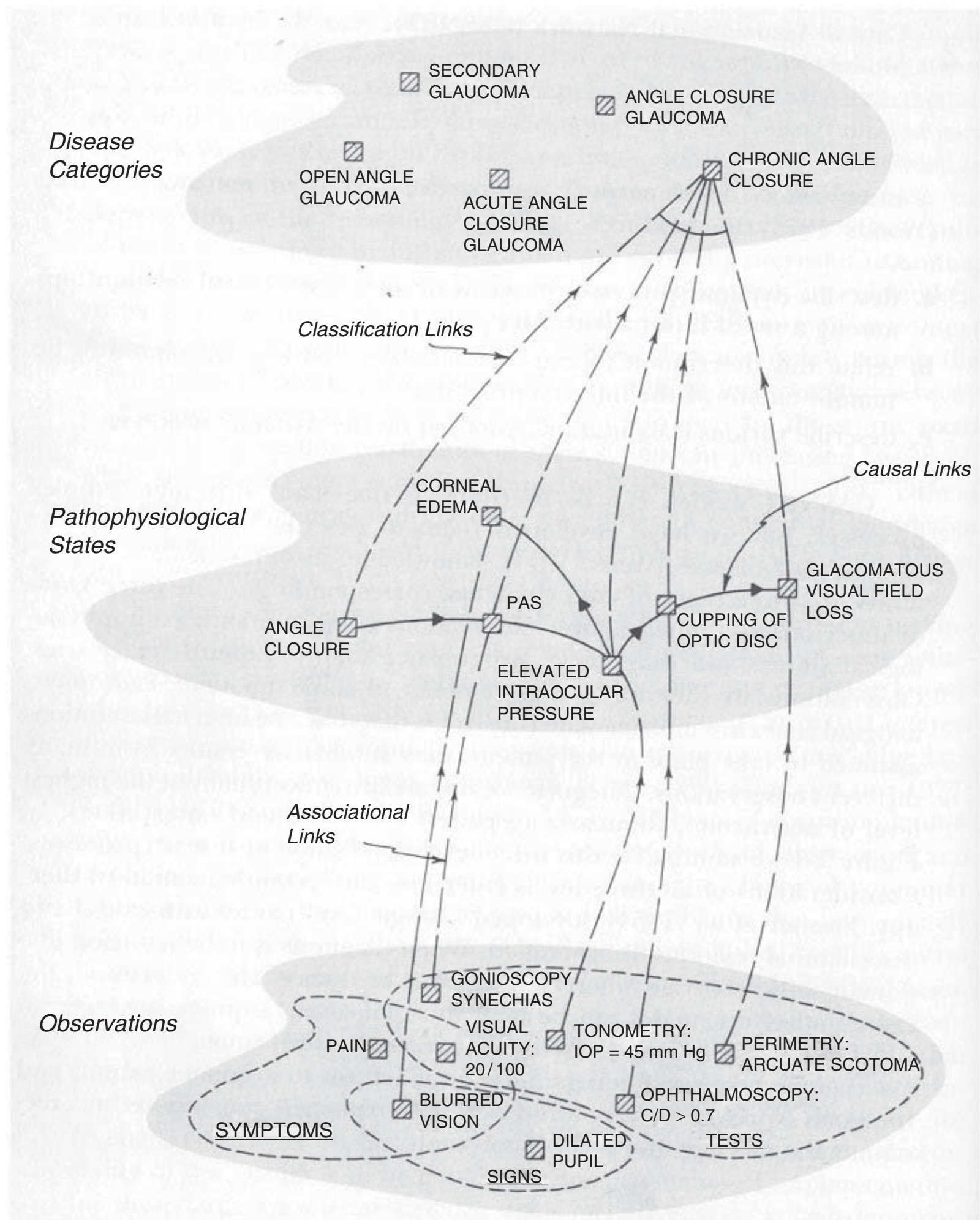
Fig. 2.2 CASNET's three-level description of a disease process. Note the *causal links* at the level of pathophysiological states. Observations (symptoms, signs, or tests) could be associated with either pathophysiological states or disease categories. (Reproduced with permission from C. Kulikowski and S. Weiss)

medicine. Like SUMEX, the Rutgers Resource was funded by the Division of Research Resources at NIH and was in time connected to the ARPAnet. Their initial major project involved a collaboration with Dr. Arin Safir, an ophthalmologist at Mt.

Fig. 2.3 An example of a MYCIN rule. Rules were encoded using the LISP programming language (at the top). Given the standardized approach to representing the knowledge, it was possible to write code to translate the rules into English (at the bottom). This provided transparency during interactions with clinical users

```
PREMISE:   ($AND  (SAME CNTXT GRAM GRAMPOS)
                  (SAME CNTXT MORPH COCCUS)
                  (SAME CNTXT CONFORM CLUMPS))
ACTION:    (CONCLUDE CNTXT IDENT STAPHYLOCOCCUS TALLY 700)
```

or (translated into English):

IF: 1) The gramstain of the organism is grampos

         2) The morphology of the organism is coccus

         3) The growth conformation of the organism is clumps

THEN: There is suggestive evidence (.7) that the identity of the organism is staphylococcus.

Sinai Medical Center in New York City, who provided the necessary clinical expertise.

This system focused on modeling **causal reasoning** using a network-based representation of the pertinent domain knowledge. The program, known as CASNET (for *causal associational network*) assisted with the diagnosis of various forms of glaucoma. Their networked approach modeled the ability of expert clinicians to reason from observations about a patient to the delineation of existing physiological states (Fig. 2.2), which in turn helped to distinguish among potential diagnostic explanations for the findings. This important work was pursued with involvement of a talented PhD student, Shalom Weiss, who made portions of the project the focus of his doctoral dissertation [20].

MYCIN

This Stanford project began as doctoral research for a medical student who was also pursuing a PhD in what today would be called **biomedical informatics**. Edward Shortliffe had come to Stanford to study medicine in 1970—partly because of the school's flexibility (which would permit a medical student to pursue a simultaneous second degree in a computer-related discipline), but also because of the advanced biomedical computing environment that Lederberg and others had created. He quickly got to know AI researchers in the computer science department on the main campus, and especially those who were involved with the Dendral project. Guided by medical school faculty (Stanley Cohen, then Chief of Clinical Pharmacology and a genetics researcher,[16] and Stanton Axline, an infectious disease expert), Shortliffe built on the Dendral notion of encoding expert knowledge in production rules. His

---

[16] Cohen, who succeeded Lederberg as chair of genetics, briefly served as Principal Investigator of the SUMEX-AIM resource upon the Lederberg's departure for Rockefeller University. After a year, Feigenbaum took over that role until he was succeeded by Shortliffe.
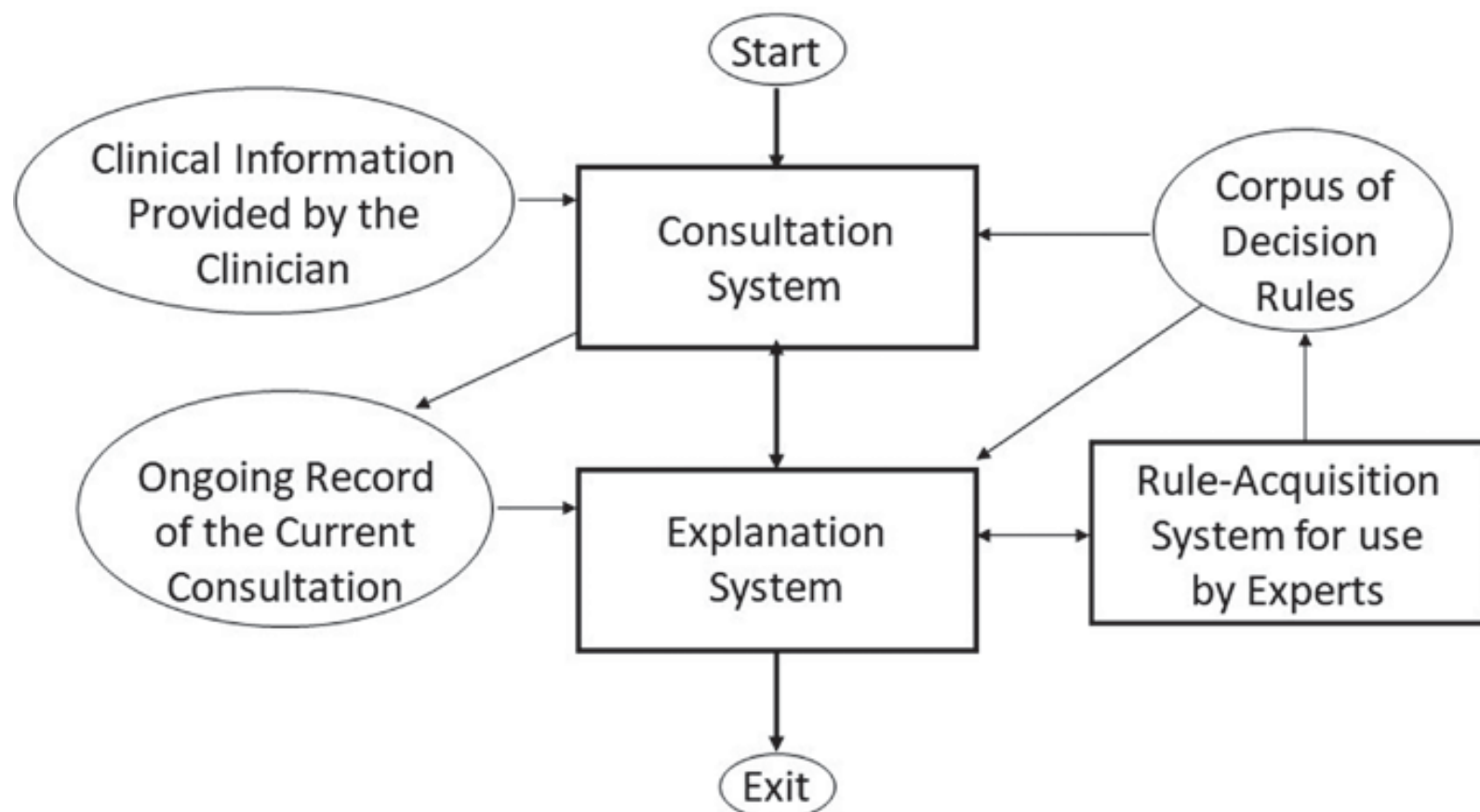
Fig. 2.4 This diagram provides an overview of the MYCIN system, identifying the three subsystems (rectangles), the corpus of decision rules, and the dynamic information that was generated during the consideration of a specific case. See text for details

principal computer science colleague was Bruce Buchanan. The idea was to develop a consultation program that would advise physicians on the selection of antimicrobial therapy for patients with severe infections. The resulting project was known as MYCIN, with Cohen serving as Shortliffe's dissertation advisor [21].

MYCIN used a collection of decision rules, acquired from Cohen, Axline, and others as the research group discussed actual cases taken from Stanford's wards. These rules were then encoded and stored in a growing collection (Fig. 2.3).

The rules were then kept separate from the actual program, which had three components (see rectangles in Fig. 2.4). The primary focus was the Consultation Program, which obtained patient data and offered advice, but also important was the Explanation Program, which could offer English-language explanations of why questions were being asked and why the program had offered its recommendations. The program itself knew how to handle a consultative interaction, but knew nothing about the domain of infectious diseases. All such knowledge was stored in the corpus of decision rules. A third subsystem, the Rule-Acquisition Program, was developed to allow experts to offer new rules or to edit existing ones. By running a challenging case through the Consultation Program, and using the Explanation Program to gain insight into why the program's performance might have been inappropriate for a given case, the expert could use the Rule-Acquisition Program to update the system's knowledge – entering new rules (for translation from English into LISP-coded versions) or editing existing ones. By re-running the case, the expert could see if MYCIN's advice had been suitably corrected.

MYCIN was formally evaluated in a blinded experiment that had infectious disease experts compare its performance with nine other prescribers who were
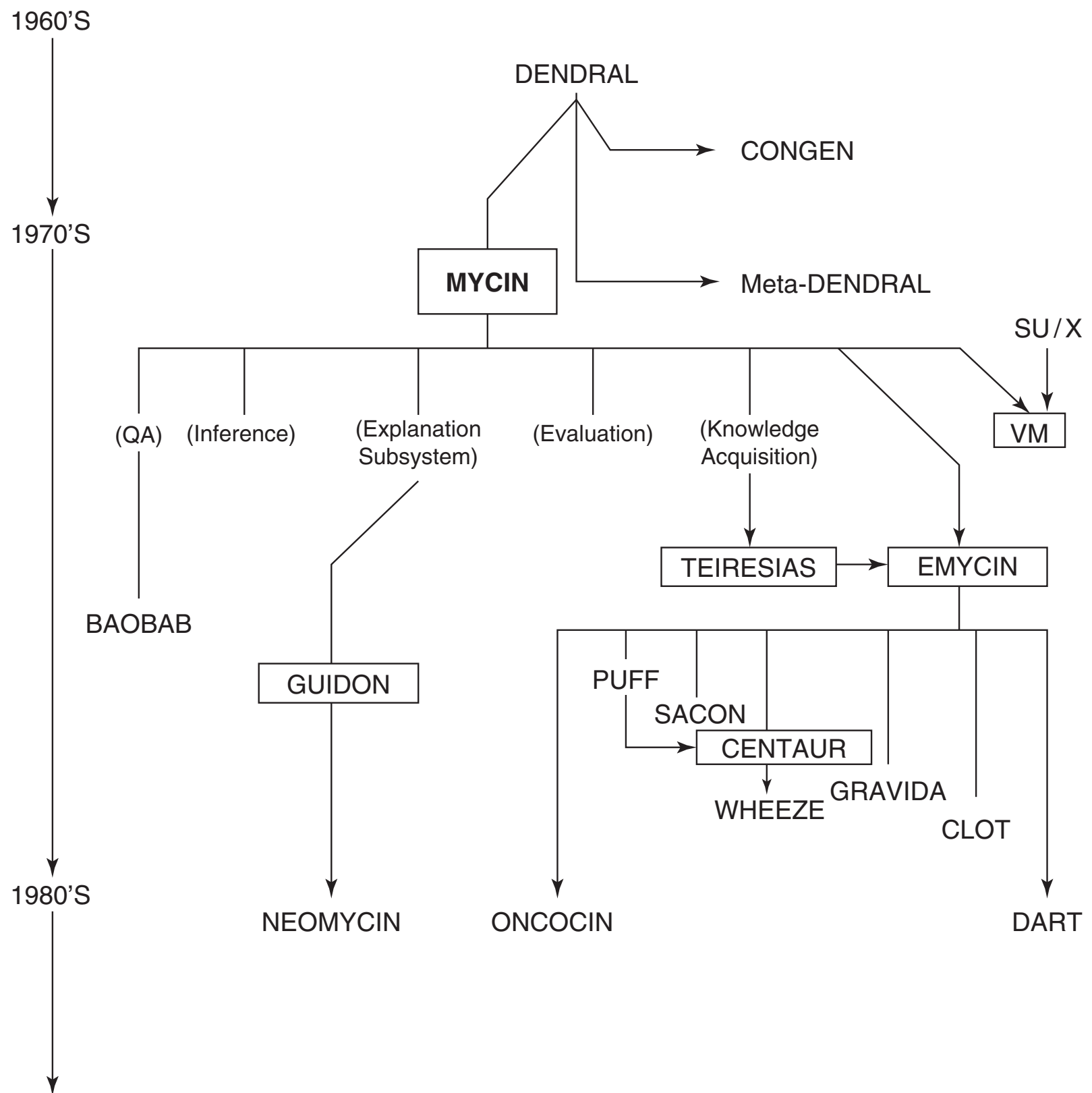
Fig. 2.5 Just as MYCIN drew inspiration from the earlier DENDRAL work, several other Stanford research projects built on the methods and concepts that MYCIN had introduced. This diagram shows many of these projects and their ancestry. Those projects depicted in rectangles were themselves the basis for computer science doctoral dissertations (VM: LM Fagan; TEIRESIAS: R Davis; EMYCIN: W van Melle; GUIDON: WJ Clancey; CENTAUR: J Aikins)

presented with the same ten cases [22]. The comparison group included the actual therapy given to the patient, Stanford infectious disease faculty members and a fellow, a medical resident, and a medical student. MYCIN was shown to perform at the top of the comparison group, as judged by the evaluators (who did not know which advice had been offered by the program).

The AI approach developed for MYCIN became known as a **rule-based expert system**. The architecture was attractive because the knowledge base was kept in rules that were separate from the program, offering the possibility that the system could provide advice in a totally different domain if the infectious disease rules

were removed and a new set of rules was substituted. The program without the rules was termed "empty MYCIN" or "essential MYCIN"—generally simply referred to as EMYCIN [23]. This work provided further support for Feigenbaum's *knowledge is power* aphorism, previously mentioned. MYCIN also stimulated several other research projects in what became known as the Stanford Heuristic Programming Project, many of which were also focused on medical topics and were doctoral dissertations in computer science (Fig. 2.5). This diagram conveys the way in which Stanford's AIM science advanced over two decades, with each project introducing methods or concepts on which subsequent research could build. An important lesson is that AIM research is about more than building systems in the engineering sense. Equally important is its dependence on the scientific method, with experiments offering lessons that generalize and can feed back into the evolution of the field [24].

Cognitive Science and AIM

As the 1970s progressed, AIM researchers became aware of the synergy between their work to capture and convey clinical expertise and the work of researchers in educational psychology and cognitive science, many of whom were focused on medical problem solving. Since AIM researchers were seeking to encode clinical expertise and to produce systems that could reason using that knowledge, they were naturally drawn to work that studied clinicians as they solved problems. An esteemed physician at Yale University's medical school, Alvan Feinstein, had published an influential volume in 1967, *Clinical Judgment* [25]. Feinstein is commonly viewed as the founder of the field of clinical epidemiology, and the focus of his volume was on defining and teaching clinical thinking. The work inspired others to pursue related aspects of clinical expertise, and several groups tackled tasks in medical problem solving, using methods from psychology and cognitive science.

Particularly influential was a volume by Elstein, Shulman, and Sprafka, educational psychologists at Michigan State University [26]. They performed a variety of studies that sought to apply contemporary psychological theories and methods to address the complexity of problem solving in cases derived from real-life clinical settings. Their work influenced the thinking of AIM researchers, who were seeking to capture elements of medical reasoning, even if their programs were not formally modeling the workings of the human mind.

Meanwhile, at Tufts New England Medical Center in Boston, two nephrologists were becoming interested in the nature of medical problem solving and the role that computers might play in capturing or simulating such reasoning. William Schwartz had published a thoughtful piece in 1970 that anticipated the future role that computers might play in medicine and the impact that such changes might impose on clinical practice and even on the types of people who would be drawn to becoming

a physician [27]. The second nephrologist, Jerome Kassirer,[17] had developed a collaboration with a computer science graduate student at MIT, Benjamin Kuipers, and they performed and published a number of experiments (further discussed in Chap. 5) that offered insights into clinical reasoning processes, including a classic paper on causal reasoning in medicine that appeared in 1984 (by which time Kuipers had joined the faculty at the University of Texas in Austin) [28].

The interest in expert reasoning in medicine, shared by Schwartz and Kassirer, also attracted a Tufts cardiologist, Stephen Pauker, and a computer scientist at MIT, Anthony Gorry. Pauker also knew how to program and this group sought to develop an experimental program that explicitly simulated the cognitive processes that they had documented in studies of expert physicians who were solving problems. This led to the development of the *Present Illness Program* (PIP - see also Chap. 4) which leveraged early cognitive science and AI and was arguably the first AIM research project to be published in a major clinical journal [29]. When Gorry departed MIT for Rice University, he was succeeded later in the decade by Peter Szolovits, himself a leader in AIM research and knowledge-based systems (see Chap. 4).

By the early 1980s there was pertinent related work underway at McGill University. Vimla Patel and Guy Groen were examining the relationship between comprehension of medical texts or descriptions with approaches to problem solving by individuals with varying levels of expertise [30]. This body of work, which extended throughout the next decade as well, provided an additional set of cognitive insights that informed the work of the AIM research community, while attracting the McGill group to become interested in how their work might influence the development of computational models of clinical expertise (see Chap. 5).

The work described briefly in this section laid the groundwork for subsequent work on expert reasoning and cognition that accounts for this book's emphasis on the interplay between AIM and cognitive science. These relationships were further solidified by the close interactions, and attendance at one another's meetings, between members of the AIM community and those in the Society for Medical Decision Making (SMDM).[18] The emergence of cognitive informatics as a specialty area within AIM research was built upon this early work and also on the growing recognition of the importance of cognitive issues in related areas of computer science, including computer-based education and human-computer interaction.

Reflecting on the 1970s

By the end of the decade, medical AI was having a significant impact on AI more generally. The top journal in the field, *Artificial Intelligence*, devoted an entire issue to AIM research [31], and the field of expert systems was being applied broadly in

---

[17]Years later Kassirer became the Editor-in-Chief of the *New England Journal of Medicine* (1992–2000).

[18]https://smdm.org/. (accessed August 13, 2022).

Fig. 2.6 The RX project was an early example of machine learning in the form of data mining under AI control. The goal was to use existing knowledge, plus real world data, to support the discovery of hypotheses that could in turn be formally explored using large amounts of data and statistical methods – thereby adding to current knowledge. For more details see https:// www.bobblum.com/ ESSAYS/COMPSCI/ rx-project.html. (Figure reproduced with the permission of R.L. Blum)
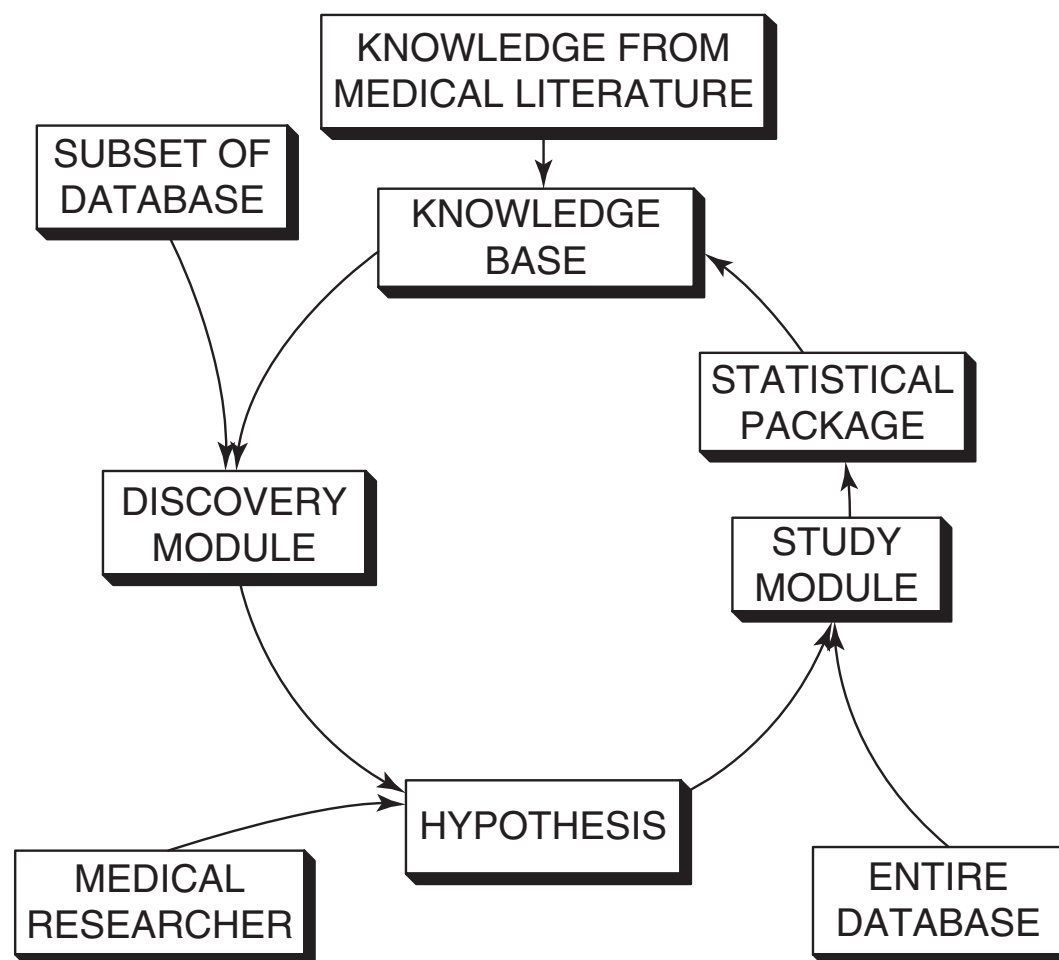
other areas of society. A community of medical AI researchers had come together to hold annual meetings (dubbed AIM Workshops) and to form new collaborations while embracing an increasing number of research projects. The first Symposium on Computer Applications in Medical Care (SCAMC), held in Arlington, VA in 1977,[19] had an entire session devoted to AIM research projects. AIM research was heavily cited in computer science research papers outside the field of medicine.

There was also important exploratory machine learning research in the medical arena, inspired in part by the Meta-Dendral work mentioned earlier. As clinical databases became available in specialized areas of medicine [32], it was natural to explore how computers might be able to learn, or discover new relationships. Blum pursued such work, proposing a cycle for discovery and clinical studies through the principled examination of such databases (Fig. 2.6) [33]. His RX program ultimately discovered and analyzed an association between prednisone and cholesterol that was published in a major clinical journal [34].

By the end of the 1970s, the AIM field was devoted to the notion that knowledge representation and use was the key to intelligent behavior by computer programs. As we describe in subsequent sections, the *knowledge is power* aphorism has been somewhat forgotten in today's AI research and application communities—arguably to their detriment.

---

[19] SCAMC went on to become the major US meeting in the field of biomedical informatics, merging with other organizations in 1989–1990 to create today's American Medical Informatics Association (AMIA), https://www.AMIA.org, (accessed August 13, 2022).

Evolution of AIM During the 1980s and 1990s

The next two decades were characterized by substantial evolution of AI and AIM, partly because of the remarkable changes in computing technology, but also because of the ups and downs of academic, industrial, and government interest in AI and its potential.

## *AI Spring and Summer Give Way to AI Winter*

By the early 1980s there was rapidly growing interest in AI, medical applications, and especially in expert systems [35]. Companies began to recruit AI scientists and commercial expert systems were introduced to the marketplace or used for internal purposes [36]. Cover stories on AI and expert systems began to appear in major popular news magazines, often with prominent featuring of medical programs such as the ones described in the previous sections of this chapter. They tended to make wild predictions about the impact AI would soon be having on society, much of which ironically did not align well with what the system developers believed to be reasonable. However, the enthusiasm continued for several years and led, for example, to a major investment by the Japanese Ministry of International Trade and Industry which formed their *Fifth Generation Computer Project*[20] starting in 1982.

Early in the decade new companies, such as Teknowledge and Intellicorp, were also created specifically to commercialize expert systems. In parallel, hardware companies such as Symbolics, LISP Machines Inc., and Xerox Corporation introduced single-user machines that were designed to run the LISP programming language, to offer graphical user interfaces with mouse pointing devices, and to support the development of expert systems and other AI-related applications. Note that these machines appeared only shortly after the introduction of the first personal computers (e.g., the Apple II in the late 1970s followed by the first IBM PC and the Apple Macintosh a few years later). In parallel, the first commercial **local area networking** products were introduced (e.g., Ethernet from Xerox Corp and a competitor known as Wangnet), which had a profound effect on the ways in which computers and programs were designed to interact and share data.

The rapid change in the early 1980s continued throughout the decade. For example, it was not long until the first general-purpose workstations running the Unix operating system were introduced (e.g., by SUN Microsystems), and these rapidly made the notion of a LISP machine obsolete. The LISP machine market disintegrated and "Unix Boxes" (high-end workstations that were much more powerful than the existing personal computers) began to dominate in the AI research community.

---

[20] See https://en.wikipedia.org/wiki/Fifth_generation_computer. (accessed August 13, 2022).

Fig. 2.7 Graphic shows
the two periods often
called *AI Winter*, one at the
end of the 1970s (which
had little impact on AIM)
and the second in the late
1980s and early 1990s
(which did affect AIM
work for several years)
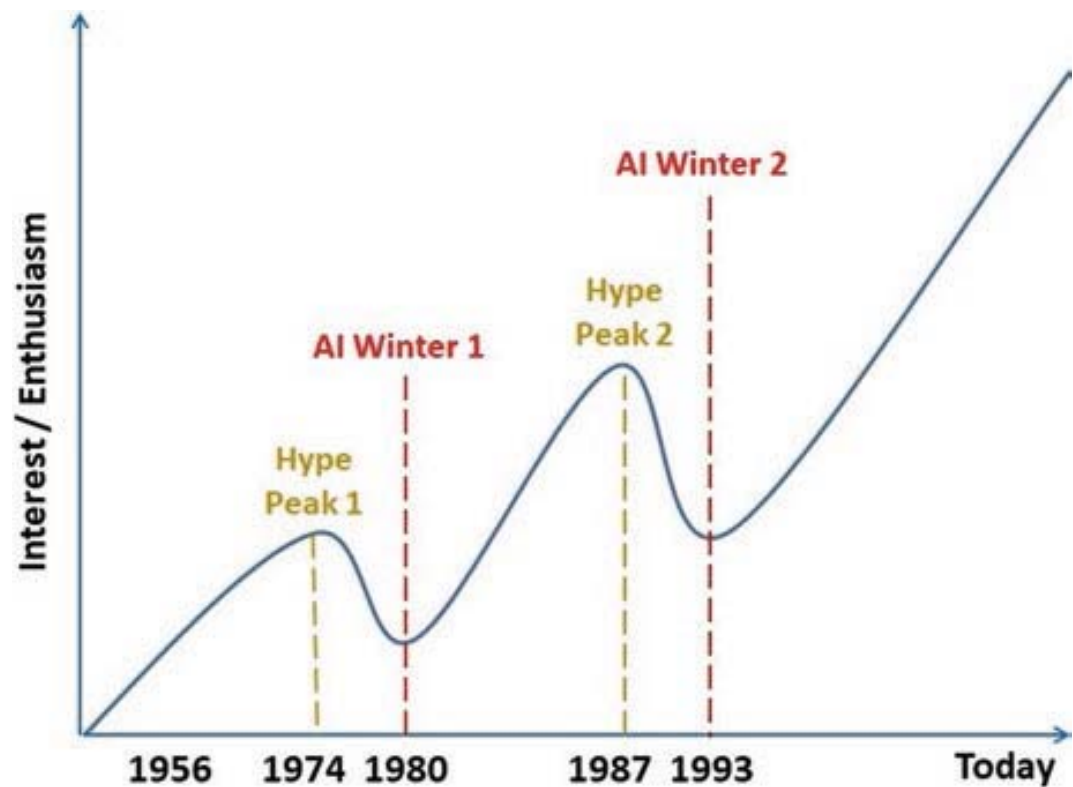
As the decade proceeded, the AI "luster" also began to fade, as highly touted
systems tended to fail to live up to their expectations. Companies often found that
the systems were expensive to maintain and difficult to update. They generally had
no machine learning component, so maintenance was crucial in order to incorporate
new knowledge into them. Performance was accordingly viewed as "brittle".

In the early 1980s ARPA had again begun to fund its support of general AI
research, with an emphasis on knowledge-based systems—no doubt encouraged to
do so by the major Japanese investment in their own project in the area. ARPA had
lowered its enthusiasm for AI research in the mid- to late-1970s, even as the AIM
activities were taking off. But AIM researchers were not supported by DARPA but
rather by NIH or, in a few cases, by the National Science Foundation (NSF), and
their work and impact had continued apace as described previously. Support for
AIM research also continued during the 1980s, while ARPA was ramping up its
own support for AI generally. However, as the decade came to an end, the AI com-
munity faced a clear diminution in the enthusiasm that had been strong only a few
years earlier. Thus, there was again a dip in funding support for AI as the 1990s
began, and some of this affected the AIM research community as well.

The dips in support for AI, and in belief for its potential, occurred in the late
1970s and again in the period between 1987 and 1993. These two drops in funding
and interest have been called **AI Winter** #1 and #2[21] (see Fig. 2.7). During these
periods it became unhelpful for companies or researchers to emphasize that they
were working on AI problems. It was hard to attract interest from collaborators or
funding agencies at a time when AI was viewed as having been oversold and having
failed to demonstrate the utility that had been promised. By the early 1990s, those
working in AI areas, including AIM, often sought new terms for what they were

---

[21] See https://en.wikipedia.org/wiki/AI_winter. (accessed August 13, 2022).

doing, hoping they would avoid the taint of the AI label. For example, work on knowledge base and terminology development often fell under the term **ontology** research,[22] and some types of machine learning research were often called **knowledge discovery in databases (KDD).**[23]

As is shown in Fig. 2.7, there has been no downturn in enthusiasm for AI and its promise for almost 30 years. Those who lived through the early AI winters often wonder if the extreme enthusiasm for AI today, with remarkable investment in almost all areas of science (and medicine/health), is a harbinger of what could become a third period of disenchantment. However, most observers feel that the field has greatly matured and that current approaches are better matched to the state of computing and communications technology than was possible when earlier research, and commercial experiments, were being undertaken. As you read this book, you should develop your own sense of whether today's enthusiasm is well matched to the reality of what is happening, especially in AIM, and whether we can be optimistic about ongoing progress and impact. We return to this topic in Chaps. 19 and 20.

AIM Deals with the Tumult of the 80s and 90s

The expert systems fervor in the 1980s, which had been driven in part by medical AI projects that offered new methods and models for analyzing data and offering advice, put the AIM community in a highly visible position. AI in Medicine had become a worldwide phenomenon, with some medical focus in Japan during the Fifth Generation Computer Project. The major new source of AIM research energy, however, was in Europe, where a medical AI community began to coalesce. The first European meeting that focused on AIM (1985) was organized by Ovo de Lotto and Mario Stefanelli as a 2-day conference in Pavia, Italy. The meeting's success led to the decision to hold such meetings biannually under the name *Artificial Intelligence in Medicine Europe* (AIME). They quickly attracted an audience from the US and other parts of the world, so eventually the meeting name was adjusted to be simply *Artificial Intelligence in MEdicine*, continuing the AIME acronym.[24]

A retrospective paper analyzing three decades of trends in the content of AIME meetings, published in 2015, provides some instructive insights on how the field evolved over that time [37]. At the first meeting in 1985, essentially all the papers dealt with knowledge-based systems and knowledge engineering, reflecting the expert systems phenomenon. However, the number of papers in those categories

---

[22] See https://en.wikipedia.org/wiki/Ontology_(information_science). (accessed August 13, 2022).

[23] See https://www.techopedia.com/definition/25827/knowledge-discovery-in-databases-kdd. (accessed August 13, 2022).

[24] The first meeting hosted outside of Europe (by the University of Minnesota) was held in 2020, although it needed to be held virtually due to the COVID-19 pandemic that prevented most travel to conferences.
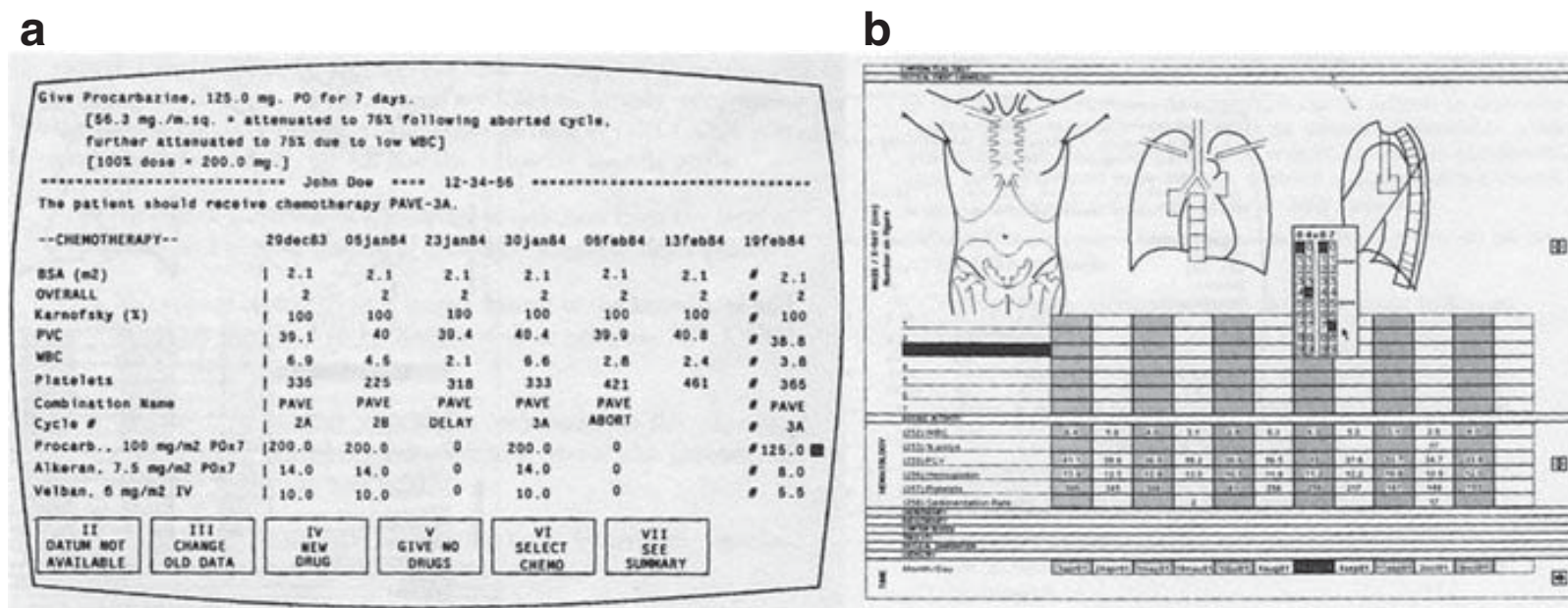
Fig. 2.8 The original ONCOCIN interface used a simple ASCII terminal and all interaction was through a computer keyboard (**a**). Within a few years, with reimplementation on a LISP machine, the program offered a greatly improved interface to clinician users (**b**)

decreased substantially over time while major new areas of emphasis were ontologies and terminologies, **temporal reasoning**, natural language processing (see Chap. 7), guidelines/protocols (see Chap. 10), management of uncertainty, and image/signal processing (see Chap. 12). The largest increase, which began slowly in the 1990s, was in the area of machine learning. By 2013 it had surpassed knowledge engineering as the most dominant topic at the meetings when measured cumulatively over three decades. This is not surprising given the AI emphasis on machine learning that today makes it the most active subfield of the discipline (see Chap. 6).

By the end of the 1980s, there was consensus that the AIM field was so active and productive that it warranted its own journal. *Artificial Intelligence in Medicine* was first published in 1989 with Kazem Sadegh-Zadeh, from the University of Műnster in Germany, serving as founding editor [38]. This journal, published by Elsevier, is a major source of current research results in the field to this day. Several other peer-reviewed journals also publish AIM methodologic research papers,[25] and the more applied work has appeared in a variety of clinical, public health, and general science journals.

The rapid evolution in networking, hardware capabilities, and computing power during the 1980s also had a major influence on AIM research and capabilities during that decade. As an example, consider the ONCOCIN program, which was developed to apply knowledge-based methods to provide advice to oncologists caring for patients enrolled in cancer chemotherapy clinical trials [39]. The program was initially conceived to run on an computer terminal attached to a mainframe computer running a LISP programming environment (Fig. 2.8a). The terminal could display

---

[25] Examples include the *Journal of the American Medical Informatics Association* (JAMIA, Oxford University Press), *Journal of Biomedical Informatics* (JBI, Elsevier), and *Intelligence-Based Medicine* (Elsevier).

Table 2.1 Some Questions Asked by AIM Researchers in an Online Discussion Forum - 1991

Question	Brief response (details in ref. [39])
1. Are AIM systems intended to address user needs?	AIM research is driven by desire to develop tools that address clinical needs, but like all basic research, there can be a long trajectory before reaching those goals
2. Has AIM research contributed to AI? To cognitive psychology? To clinical medicine?	Basic methodological innovation is often required, even when working on applied goals. There is ample evidence that AI, computer science, and cognitive psychology have all been affected by AIM research. Contributing to clinical medicine remains a future goal, although the work has already stimulated much discussion in the clinical world
3. Is AIM part of information science, computer science, AI, engineering, or biomedicine?	Created at the intersection of all these fields, AIM is a key component of the field of biomedical informatics (BMI), which itself merges those disciplines
4. Are AIM researchers adequately trained? To what extent is there a problem with inbreeding?	Daunting as it may be to study broadly the many fields that come together in BMI and AIM, it has been the lack of expertise *at that intersection* which has accounted for many of the problems in developing and implementing computer-based systems for biomedicine. Hence training in BMI is optimal. Inbreeding will be less of an issue as more BMI training programs are introduced
5. Why is it so difficult for AIM systems to be tested in clinical settings with regard to process or outcome of clinical care?	This question reveals unfamiliarity with the many AIM system evaluations that have been published. Most focus on the quality of decisions. Demonstrations of impact on the quality of care and patient outcome remain to be undertaken
6. Why isn't AIM research better funded?	All AI-related research funding has suffered during the downturn in interest in AI. But it will rebound as we continue to demonstrate the value and innovation in what we do
7. Why have AIM systems been so difficult to transport from site to site successfully?	With the demise of the "Greek Oracle" (consultation model) [19], integration with local information systems has become the crucial delivery mechanism. In the lack of standards or facile access to such systems, it is very hard to move a developed system to a new integrated environment

only ASCII characters[26] and all interactions were by computer keyboard. Within a few years, with the introduction of Xerox LISP machines that were self-contained for single users and included both a mouse pointing device and high quality graphical capabilities, ONCOCIN was ported to a LISP device that provided a greatly improved interface that was intuitive for clinicians to use (Fig. 2.8b).

The democratization of the Internet, which occurred during the late 1980s and early 1990s (with the commercialization of its management and creation of the domain system), created opportunities for collaboration at a distance as well as the

---

[26]American Standard Code for Information Interchange (ASCII), is a character encoding standard for electronic communication. ASCII codes represent text in computers, telecommunications equipment, and other devices.

emergence of communities with shared interests. At a time when AI Winter was affecting the AIM research community, it is not surprising that forums for sharing opinions, asking questions, and providing pointers to information of interest would emerge. One such list, simply called *ai-medicine@stanford.edu*, had been created in advance of the AIME meeting held in Maastricht, The Netherlands in August 1991. A keynote presentation at that meeting assessed a variety of soul-searching questions that AIM researchers had been asking one another on the list server. Later published in the *AI in Medicine* journal, the paper looked at AI in Medicine's "adolescence" and anticipated its future directions [40]. Table 2.1 summarizes seven key questions and briefly provides the response from the article, although interested readers should peruse the full paper. Many of the questions (and answers) are still relevant today, some 30 years later. Fifteen years after the Maastricht meeting, the AIME meeting, held in Amsterdam in 2007, provided a panel that reassessed the questions and answers from 1991, while adding thoughts about how the field had evolved in the intervening years [41].

As AIM's first four decades came to an end (with the century), work on advanced systems was continuing apace, with improved funding and enthusiasm. With growing implementation of electronic health records (EHRs) and creation of digital imaging databases, coupled with the general availability of enhanced computational power, machine learning (ML) research was gaining in interest and impact. The ML revolution was on the horizon and today has been a dominant element in AI in general and in AIM. In the next section, we briefly examine the two decades that led to the present.

The Last 20 Years: Both AI and AIM Come of Age

The early 2000s were dominated by the completion of the human genome project and the associated rise of interest in **bioinformatics**, while the adoption of EHRs continued silently in the background at a slow pace. Several techniques in supervised machine learning were first applied to large biomedical datasets in the context of **genomics** and bioinformatics work [42, 43].

Meanwhile, in computer science, there were two major developments underway: (1) the availability of commodity **graphical processing units** (GPUs),[27] beginning in about 2001, for efficiently manipulating image data–which at their core comprise an array of numbers, and (2) the availability of large, labeled datasets (such as the introduction of ImageNet[28] in 2010) to support efforts to learn increasingly complex classifiers via **supervised machine learning**. The availability of ImageNet and the recognition that GPUs could be as flexible as CPUs (but orders of magnitude faster in array operations) led to accelerated progress in image recognition—partly due to

---

[27] https://en.wikipedia.org/wiki/GeForce_3_series. (accessed August 13, 2022).
[28] https://en.wikipedia.org/wiki/ImageNet#History_of_the_database. (accessed August 13, 2022).

the creation of annual contests using shared datasets.[29] The computing ability offered by GPUs accelerated the adoption of artificial neural networks (which, as was mentioned earlier in this chapter, had been explored since the 1960s, initially inspired by the concept of perceptrons). Ideas put forward by Geoffrey Hinton, Yann LeCun, and Yoshua Bengio for deep neural networks [44] became widely adopted beginning in 2006 (earning the trio the 2018 ACM A.M. Turing Award[30]). A landmark was reached in 2012, when a deep convolutional neural network called AlexNet achieved a 16% error rate in the ImageNet challenge (the previous best performance had hovered at around 25%). That same year, Andrew Ng and Jeff Dean (both at Google) demonstrated the feasibility of **unsupervised machine learning** (see Chap. 6) by training a computer to recognize over 20,000 object categories, such as cat faces and human faces, without having to label images as containing a face or a cat [45].

The developments in computer science percolated to medicine, initially in the form of image analysis advances in radiology and pathology. For a few years expert systems (and knowledge-based approaches in general) took a back seat given the challenges in acquisition of patient data in electronic form to enable the machine learning approaches. Adoption of electronic medical records regained momentum after the passage of the Health Information Technology for Economics and Clinical Health (**HITECH**) Act in 2009. By 2012 the powerful compute capabilities (in the form of **cloud computing**) were readily accessible for a nominal fee; machine learning using neural networks had proved its worth in image, text and speech processing; and patient data in electronic form were available in large amounts–leading to a renewed enthusiasm about the potential of AI in Medicine.

As a result, the application of supervised machine learning to medical datasets became commonplace, leading to rapid advances in learning classifiers using large amounts of labeled data. Computers approximated human ability in reading retinal images [46], X-rays [47], histopathology slides [48], and the entire medical record to provide diagnostic as well as prognostic outputs [49]. However, as mentioned earlier, in the hype around "deep learning" the *knowledge is power* aphorism was often forgotten and, on occasion, re-discovered [50].

It is too soon to tell if this third AIM wave will deliver on the hype or lead to another, and potentially more severe, AI winter (Fig. 2.7). However, old concerns around explainability and trustworthiness of AI systems in medicine [51] are again being actively discussed (see Chaps. 8 and 18), with a keen focus on prevention of bias and ensuring fairness in their use in medical decision making [52, 53].

Given today's massive amount of activity in the field, there are several ongoing debates. For example, it is unclear if the unstructured content from clinical notes holds much value in improving diagnostic or prognostic systems given the high prevalence of copy-and-pasting, use of templates, and pressures to over-document

---

[29] https://www.image-net.org/challenges/LSVRC/. (accessed August 13, 2022).

[30] Often called the "Nobel Prize in Computer Science". See https://awards.acm.org/about/2018-turing. (accessed August 13, 2022).

in light of billing concerns (see Chaps. 10 and 11). As another example, there is increasing tension between the need to share data for training AI systems and the desire to ensure patient privacy (see Chap. 18). Once considered a forward-thinking piece of legislation, the **Health Insurance Portability and Accountability Act (HIPAA)** from 1996 is increasingly considered a hindrance to building AI systems[31] while also being inadequate to protect patient privacy [54].

While the media hype around AI in medicine continues, there are several exciting possibilities to integrate the advances from the pre-2000s with recent developments. A particularly noteworthy direction is on combining symbolic computing with deep neural networks [44] (see Chap. 6). As Bengio, Lecun and Hinton note, it was a surprise that the simple approach (creating networks of relatively simple, non-linear neurons that learn by adjusting the strengths of their connections) proved so effective when applied to large training sets using huge amounts of computation (thanks to GPUs!). It turned out that a key ingredient was the depth of the networks; shallow networks did not work as well, but until the last decade or so we lacked the computational power to work with neural networks that were "deep". In outlining the promising future directions for AI research, these authors reflect in their Turing lecture [43] on the role that the symbolic AI research from the twentieth century might play in guiding how we structure and train neural nets so they can capture underlying causal properties of the world. In the same vein, we encourage the reader to reflect again on the rich history of symbolic reasoning systems built by AIM researchers in the twentieth century (as presented earlier in this chapter and recapitulated in some detail in Chap. 4). It is exciting to consider how that earlier work might be complementary to the machine learning developments in the last 20 years. As we suggested in Chap. 1 and earlier in this chapter, future work may demonstrate that combining the two paradigms, with a better focus on the role of cognitive science in designing ML systems (see Chaps. 5, 6 and 20), might catalyze rapid progress in the core diagnostic and prognostic tasks of AI in Medicine.

Today's cutting edge research will be tomorrow's history. The following chapters provide a glimpse of how current research and practice may evolve as both methods and computational capabilities continue to advance.

Questions for Discussion

- How would you characterize the notion of "intelligence", first as a characteristic of human beings (or other organisms) and second as a feature of modern computing? How do those characterizations diverge from one another? In what sense are devices that you use every day "intelligent".
- What has been the role of communications technology in advancing both artificial intelligence research and its applications in biomedicine?
- Given the explosive interest in expert systems, including their potential use in biomedicine, to what do you attribute their failure to meet early expectations and

---

[31] https://hai.stanford.edu/news/de-identifying-medical-patient-data-doesnt-protect-our-privacy. (accessed August 13, 2022).

the emergence of the AI Winter of 1987–1993? Consider inherent characteristics of the approach as well as the then-current communications and computational technologies.
- What accounts for the slow progress in machine learning (despite some impressive early examples) until the last two decades?
- Do we need a resurgence of expertise in the area of knowledge engineering for the development of medical AI systems? Why or why not?
- What uses might unsupervised learning have in medicine?
- How might prior medical knowledge, codified in knowledge structures such as ontologies, be provided to deep neural networks to improve their performance?
- What are the principal barriers that you envision in the ongoing effort to develop, test, and implement medical AI systems that interact directly with clinicians? With patients?

Further Reading

Dyson G. Turing's Cathedral: The Origins of the Digital Universe. New York: Vintage Books, 2012.

- A historical description of scientific innovation, told in the context of work by a team of young mathematicians and engineers, led by John von Neumann at Princeton's Institute for Advanced Study, who applied the ideas of Alan Turing to develop the fastest electronic computer of its era. That work also introduced the concept of RAM (random access memory) that we still use in most computers today. See also Alice Rawsthorn's book review, "Genius and Tragedy at Dawn of Computer Age" (New York Times, March 25, 2012).

Simon HA. The Sciences of the Artificial (3rd edition). Cambridge, MA: MIT Press, 1996.

- Originally published in 1968, this is a classic volume by a Nobel Laureate (Economics) who was also an early luminary in the field of AI. His assessment of AI includes topics that include not only his thoughts as a cognitive psychologist, but also analyses of the organization of complexity, the science of design, chaos, adaptive sysstems, and genetic algorithms.

Clancey WJ, Shortliffe EH. Readings in Medical Artificial Intelligence: The First Decade. Reading, MA: Addison-Wesley, 1984.

- This book is a compendium of classic papers describing the first generation of AIM systems, including MYCIN, CASNET and INTERNIST-1. It provides a detailed account of the methods underlying the development of these systems, including methods for the elicitation of expert knowledge and probabilistic inference procedures.

Shortliffe EH. Artificial intelligence in medicine: Weighing the accomplishments, hype, and promise. *IMIA Yearbook of Medical Informatics* 2019;28(01):257–62, (https://doi.org/10.1055/s-0039-1677891).

- This paper can be considered as the third in a series that includes refs. [40, 41] in that it describes the state of the field approaching the year 2020. The paper provides a historically-informed perspective on recent developments in machine learning methodology, takes stock of achievements to date and considers challenges that remain for clinically impactful deployment of AIM systems.

Szolovits P. (ed.) (1982). Artificial Intelligence in Medicine. (AAAS Selected Symposium). Boulder, CO: Westview Press. (https://www.google.com/books/edition/Artificial_Intelligence_In_Medicine/8tmiDwAAQBAJ)

- This book is an edited volume, published originally by the American Association for the Advancement of Science (AAAS), with chapters summarizing much of the medical AI research of the 1970s. It includes an especially important paper by Harry Pople describing the evolution of the INTERNIST-1 system.

Bengio Y, Lecun Y, Hinton G. Deep Learning for AI. Communications of the ACM 2021;64(7):58–65 (https://doi.org/10.1145/3448250).

- Yoshua Bengio, Yann LeCun, and Geoffrey Hinton are recipients of the 2018 ACM A.M. Turing Award for breakthroughs that have made deep neural networks a critical component of computing. This commentary describes their reflections on the progress to date in building deep neural networks and their thoughts on the future of deep learning, including the role of symbolic AI.

References

1. Whitehead AN, Russell B. Principia Mathematica. 2nd ed. Andesite Press; 2015, Originally published in 1910.
2. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys. 1943;5:115–33. https://link.springer.com/article/10.1007/BF02478259
3. Post EL. Formal reductions of the general combinatorial decision problem. Am J Math. 1943;65(2):197–215. https://doi.org/10.2307/2371809.
4. Polya G. How to solve it: a new aspect of mathematical method. Princeton, NJ: Princeton University Press; 2015, Originally published in 1945.
5. Bush V. As we may think. The Atlantic, July 1945. https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/.
6. Turing AM. Computing machinery and intelligence. Mind. 1950;LIX(236):433–60. https://doi.org/10.1093/mind/LIX.236.433.
7. Shannon CE. Programming a computer for playing chess. Computer Chess Compendium. 1943:2–13. https://doi.org/10.1007/978-1-4757-1968-0_1.
8. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev. 1958;65(6):386–408. https://doi.org/10.1037/h0042519.
9. Samuel AL. Some studies in machine learning using the game of checkers. IBM J Res Dev. 1959;44:206–26. https://ieeexplore.ieee.org/document/5389202
10. Minsky M. Semantic information processing. Cambridge, MA: MIT Press; 1960.
11. Minsky M, Papert SA. Perceptrons: an introduction to computation geometry. Cambridge, MA: MIT Press; 1987, Originally published in 1969.
12. Garber M. When PARRY met ELIZA: a ridiculous chatbot conversation from 1972. The Atlantic, June 9, 2014. https://www.theatlantic.com/technology/archive/2014/06/when-parry-met-eliza-a-ridiculous-chatbot-conversation-from-1972/372428/.

13. Lindsay RK, Buchanan BG, Feigenbaum EA, Lederberg J. Applications of artificial intelligence for organic chemistry: the DENDRAL project. New York: McGraw-Hill (advanced computer science series); 1980. https://collections.nlm.nih.gov/catalog/nlm:nlmuid-101584906X7379-doc

14. Feigenbaum EA. The art of artificial intelligence: themes and case studies of knowledge engineering. Proceedings of the fifth international joint conference on artificial intelligence, Cambridge, MA; 1977. https://www.ijcai.org/proceedings/1977-2.

15. Ledley RS, Lusted LB. Reasoning foundations of medical diagnosis: probability, logic, and medical diagnosis. Science. 1959;130(3380):892–930. https://doi.org/10.1126/science.130.3366.9.

16. Warner HR, Toronto AF, Veasy L. Experience with Bayes' theorem for computer diagnosis of congenital heart disease. Ann N Y Acad Sci. 1964;115:558–67. https://doi.org/10.1111/j.1749-6632.1964.tb50648.x.

17. de Dombal FT, Leaper DJ, Staniland JR, McCann AP, Horrocks JC. Computer-aided diagnosis of acute abdominal pain. Br Med J. 1972;1:376–80. https://doi.org/10.1136/bmj.2.5804.9.

18. Miller R, Pople H, Myers J. INTERNIST-1: an experimental computer-based diagnostic consultant for general internal medicine. N Engl J Med. 1982;307:468–76. https://doi.org/10.1056/nejm198208193070803.

19. Miller RA, Masarie FE. The demise of the "Greek Oracle" model for medical diagnostic systems. Methods Inf Med. 1990;29(1):1–2. https://doi.org/10.1055/s-0038-1634767.

20. Weiss SM, Kulikowski CA, Amarel S, Safir A. A model-based method for computer-aided medical decision-making. Artif Intell. 1978;11:145–72. https://doi.org/10.1016/0004-3702(78)90015-2.

21. Shortliffe EH. Computer-based medical consultations systems: MYCIN. New York: American Elsevier; 1976.

22. Yu VL, Fagan LM, Wraith SM, Clancey WJ, Scott AC, Hannigan J, Blum RL, Buchanan BG, Cohen SN. Antimicrobial selection by a computer: a blinded evaluation by infectious disease experts. JAMA. 1979;242:1279–82. https://doi.org/10.1001/jama.1979.03300120033020.

23. van Melle W. A domain-independent system that aids in constructing knowledge-based consultation programs. PhD dissertation, Computer Science Department, Stanford University, 1980. Published as van Melle W, System aids in constructing consultation programs. Ann Arbor, MI: UMI Research Press; 1981.

24. Buchanan BG, Shortliffe EH. Rule-based expert systems: the MYCIN experiments of the Stanford heuristic programming project. Reading, MA: Addison-Wesley; 1984.

25. Feinstein AR. Clinical judgment. Malabar, FL: Krieger Publishing Co.; 1967.

26. Elstein AS, Shulman LS, Sprafka SA. Medical problem solving: an analysis of clinical reasoning. Cambridge, MA: Harvard University Press; 1978.

27. Schwartz WB. Medicine and the computer: the promise and problems of change. N Eng J Med. 1970;283(23):1257–64. https://doi.org/10.1056/nejm197012032832305.

28. Kuipers BJ, Kassirer JP. Causal reasoning in medicine: analysis of a protocol. Cogn Sci. 1984;8:363–85. https://doi.org/10.1016/S0364-0213(84)80007-5.

29. Pauker SG, Gorry GA, Kassirer JP, Schwartz WB. Towards the simulation of clinical cognition: taking a present illness by computer. Am J Med. 1976;60(7):981–96. https://doi.org/10.1016/0002-9343(76)90570-2.

30. Patel VL, Groen GJ. Knowledge-based solution strategies in medical reasoning. Cogn Sci. 1986;10:91–116. https://doi.org/10.1207/s15516709cog1001_4.

31. Sridharan NS. Guest editorial: special issue on artificial intelligence in biomedicine. Artif Intell. 1978;11(1–2):1–4. https://doi.org/10.1007/s13755-017-0040-y.

32. Fries JF. The chronic disease databank: first principles to future directions. J Med Philos. 1984;9:161–80. https://doi.org/10.1093/jmp/9.2.161.

33. Blum RL. Discovery and representation of causal relationships from a large time-oriented clinical database: the RX project. In: Lecture notes in medical informatics, vol. 19. Berlin: Springer; 1982. https://doi.org/10.1007/978-3-642-93235-9_1.

34. Blum RL. Computer-assisted design of studies using routine clinical data: analyzing the association of prednisone and cholesterol. Ann Intern Med. 1986;104(6):858–68. https://doi.org/10.7326/0003-4819-104-6-858.

35. Hayes-Roth R, Waterman DA, Lenat DB. Building expert systems. Reading, MA: Addison-Wesley; 1983.

36. Feigenbaum EA, McCorduck P, Nii HP. The rise of the expert company: how visionary companies are using artificial intelligence to achieve higher productivity and profits. New York: Times Books; 1988.

37. Peek N, Combi C, Marin R, Bellazzi R. Thirty years of artificial intelligence in medicine (AIME) conferences: a review of research themes. Artif Intell Med. 2015;65:61–73. https://doi.org/10.1016/j.artmed.2015.07.003.

38. Sadegh-Zadeh K. Machine over mind (editorial). Artif Intell Med. 1989;1:3–10. https://doi.org/10.1016/0933-3657(89)90012-2.

39. Hickam DH, Shortliffe EH, Bischoff MB, Scott AC, Jacobs CD. A study of the treatment advice of a computer-based cancer chemotherapy protocol advisor. Ann Intern Med. 1985;103:928–36.

40. Shortliffe EH. The adolescence of AI in medicine: will the field come of age in the 90s? Artif Intell Med. 1993;5:93–106. https://doi.org/10.1016/j.artmed.2008.07.017.

41. Patel VL, Shortliffe EH, Stefanelli M, Szolovits P, Berthold MR, Bellazzi R, Abu-Hanna A. The coming of age of artificial intelligence in medicine. Artif Intell Med. 2009;46:5–17.

42. Ding CHQ, Dubchak I. Multi-class protein fold recognition using support vector machines and neural networks. Bioinformatics. 2001;17(4):349–58. https://doi.org/10.1093/bioinformatics/17.4.349.

43. Furey TS, Cristianini N, et al. Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics. 2000;16(10):906–14. https://doi.org/10.1093/bioinformatics/16.10.906.

44. Bengio Y, Lecun Y, Hinton G. Deep learning for AI. Commun ACM. 2021;64(7):58–65. https://doi.org/10.1145/3448250.

45. Quoc VL, Ranzato MA, et al. Building high-level features using large scale unsupervised learning. *arXiv* 2012;1112.6209v5. https://arxiv.org/abs/1112.6209.

46. Gulshan V, Peng L, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316(22):2401–10. https://doi.org/10.1001/jama.2016.17216.

47. Irvin J, Rajpurkar P, et al. CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison. Proc AAAI Conf Artif Intell. 2019;33:590–7. https://doi.org/10.1609/AAAI.V33I01.3301590.

48. Komura D, Ishikawa S. Machine learning methods for histopathological image analysis. Comput Struct Biotechnol J. 2018;16:34–42. https://doi.org/10.1016/j.csbj.2018.01.001.

49. Rajkomar A, Oren E, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med. 2018;1:18. https://doi.org/10.1038/s41746-018-0029-1.

50. Nestor, B, McDermott MBA, et al. Proceedings of the 4th machine learning for healthcare conference, PMLR. 2019;106:381–405. https://proceedings.mlr.press/v106/nestor19a.html.

51. Markus AF, Iors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. J Biomed Info. 2021;113:103655. https://doi.org/10.1016/j.jbi.2020.103655.

52. Char DS, Shah NH, Magnus D. Implementing machine learning in health care: addressing ethical challenges. N Engl J Med. 2018;378:981–3. https://doi.org/10.1056/NEJMp1714229.

53. McCradden MD, Joshi S, et al. Ethical limitations of algorithmic farness solutions in health care machine learning. Lancet Digital Health. 2020;2(5):E221–3. https://doi.org/10.1016/S2589-7500(20)30065-0.

54. Mandl KD, Perakslis ED. HIPAA and the leak of "deidentified" EHR data. N Engl J Med. 2021;384:2171–3. https://doi.org/10.1056/NEJMp2102616.