Chapter 5
Clinical Cognition and AI: From Emulation to Symbiosis

Vimla L. Patel and Trevor A. Cohen

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

- How do contemporary AI systems differ from expert human decision makers?
- Why is understanding clinical cognition critical for the future of sustainable AI?
- What constraints on human decision making suggest a complementary role for AI in clinical decision making?
- How might AI enhance the safety of clinical practice?

Augmenting Human Expertise: Motivating Examples

One of the more controversial claims about AI systems in medicine is that they have the potential to replace the role of the physician, especially in perceptual domains such as radiology and pathology, in which interpretation of images is a prominent component of physician work. While it is natural that practitioners with a focus on image interpretation would consider the implications of current AI technologies for the professional viability of their fields (see, for example [1]), a strong counterargument to this claim is that these technologies may play a complementary role in the field and allow radiologists (and pathologists) to focus on assessment and

V. L. Patel (✉)
New York Academy of Medicine, New York, NY, USA
e-mail: vpatel@nyam.org

T. A. Cohen
University of Washington, Seattle, WA, USA
e-mail: cohenta@uw.edu

T. A. Cohen et al. (eds.), *Intelligent Systems in Medicine and Health*, Cognitive Informatics in Biomedicine and Healthcare,
https://doi.org/10.1007/978-3-031-09108-7_5

communication of AI-based image interpretations, and the positioning of these interpretations within a broader diagnostic workflow [2]. Alternatively, and in line with the main motivating argument for the current chapter, it has been proposed that physicians and AI systems might play a complementary role in diagnosis itself [1, 3, 4], though less attention has been paid to several other crucial areas of a clinician's task.

This chapter considers the proposal of complementary physician/AI systems from a **cognitive informatics** perspective, focusing on the strengths and weaknesses of the information processing systems concerned. Before proceeding to address these issues, the section below presents some examples from the published literature of systems that establish a case for the utility of human-machine collaboration in order to augment human abilities.

The burgeoning literature on AI-based diagnostic systems in radiology is replete with examples of hybrid human/AI systems outperforming either component taken alone in diagnostic tasks. For example, Lakhani and Sundaram report results from a combined human/AI workflow in which a board-certified cardiothoracic radiologist was enlisted to resolve disagreements between two convolutional network architectures trained to identify pulmonary tuberculosis in chest radiographs [5]. This arbitration process improved ensemble model specificity from 94.7% to 100% without loss in sensitivity, with the radiologist reviewing only those 13 of 150 test cases in which disagreement between models occurred. Patel and colleagues report results from a workflow in which images with low-confidence predictions for the presence or absence of pneumonia from a convolutional network were reconsidered by groups of radiologists in concert [6]. Probabilistic estimates from these experts were then used as an alternative to the model's original predictions, resulting in an approximately 10% improvement in accuracy over that obtained with deep learning alone.

In both cases, the combined human/AI system also outperformed its human component, an individual radiologist in the tuberculosis study, and a group of radiologists in the pneumonia study. Another common finding of interest is that the predominant mode of improvement with human oversight is an improvement in specificity. That is to say, the AI models alone tended toward overdiagnosis, which supports a pragmatic argument for the judicious use of human expertise to reduce false positive diagnoses in those cases in which uncertainty is identified either through disagreement between models, or through low-probability predictions from a single model.

Similar findings have been observed in dermatology diagnosis. Combined human/AI systems outperformed their independent components [7], with a 2.5% increase in specificity when enforcing the same level of sensitivity. Notably, some work in this area has also investigated the role of representation—advantages in performance for the human-computer collective were observed to be contingent upon the granularity (probabilities of differential diagnoses vs. global risk of malignancy) and cognitive demand of the representation used to convey predictions to

physicians [8]. These results illustrate the need to consider the constraints on human information processing when attempting to integrate AI into clinical decision making processes. While these results concern perceptual domains of medicine, it has also been argued that AI can play a complementary role in verbal domains by supporting the aggregation and synthesis of information required to reach a diagnostic conclusion [3].

These sorts of pragmatic motivating arguments for the consideration of human cognition are very different from those that motivated considerations of human information processing earlier in the development of AIM. With early systems, there was a desire to develop models that emulated procedures characteristic of human intelligence, with two early systems (INTERNIST-1 [9] and the Present Illness Program [10]) deliberately designed to model the generation and testing of a set of diagnostic hypotheses that cognitive studies had suggested were characteristic of the behavior of medical experts [11].

The section that follows considers the intersection among cognitive science, clinical cognition and AI, from earlier studies to current work, with a focus on the shared roots of these fields and the need for AI development to consider human cognition.

Cognitive Science and Clinical Cognition

Cognitive science, or the science of cognition, includes numerous subfields of psychology, philosophy, linguistics, cognitive anthropology, neuroscience and computer science. Basic research in cognitive science uses theories and methods from a combination of these domains to investigate problems, including clinical problems. For example, a program of research has used theories and methods from cognitive science to investigate clinical cognition and medical decision making (for examples see: [12–14]). Table 5.1, illustrates how research in basic cognitive sciences is related to our understanding of clinical cognition.

Similarly, our understanding of the reasoning processes and knowledge associated with diagnostic and patient management provides a basis for influencing the development of medical AI and decision support systems. For example, research in characterizations of expert and novice clinical knowledge organization in human memory can be used in creating representations of such knowledge in clinical AI systems. Table 5.2 shows the corresponding relationships between medical cognition and research in AI. The science of cognition provides the foundation needed to drive AI-based decision-support systems that can augment human behavior.

Research in clinical cognition draws on the theories, and methods developed in basic cognitive science, and contributes to applications in biomedical informatics in a number of ways. We are beginning to see a greater awareness of the concept of clinical cognition and its relationship to clinical support systems. A recent literature

Table 5.1 Correspondences between cognitive science and medical cognition

Cognitive science	Medical cognition
Knowledge organization and human memory	Organization of clinical and basic science knowledge
Problem solving, heuristics/reasoning strategies	Medical problem solving and decision making
Perception/attention	Interpretation of radiologic and other visual data
Diagrammatic reasoning	Perceptual processing of patient data displays
Text comprehension	Learning from medical texts
Dialog analysis	Medical discourse analysis
Distributed cognition	Collaborative practice in health care
Coordination of theory and evidence	Diagnostic and therapeutic reasoning
Natural intelligence	Expertise in clinical practice

Table 5.2 Correspondences between medical cognition and research in AI

Medical cognition	Medical AI
Organization of clinical basic science knowledge	Development and use of medical knowledge bases in intelligent systems
Medical problem solving and decision making	Medical artificial intelligence/decision support systems
Radiologic and dermatologic diagnosis	Visual data analytics/machine learning
Perceptual processing of patient data displays	Biomedical information visualization
Learning from medical texts/medical discourse analysis	Natural language processing
Collaborative practice in health care	Technology-supported collaborative environments
Diagnostic and therapeutic reasoning	Clinical support systems
Natural intelligence in clinical practice	Interactive environments for collaborative problem solving

evaluation from a biomedical informatics journal identified 57 articles that were related to cognitive informatics [15]. The topics of these articles ranged from characterizing the limits of clinician problem-solving and reasoning behavior and characterization of distributed clinical teams, to developing cognitively plausible interventions for supporting clinician activities. The reader is referred to Chap. 4 in Shortliffe and Cimino's textbook of Biomedical Informatics for comprehensive coverage of this topic [16].

Symbolic Representations of Clinical Information

Much of the research in late 1980s and 90s, such as the research in Patel's laboratory, fell into the **symbolic** tradition, and dealt with models of diagnostic reasoning. The theoretical foundation of cognitive modeling is the idea that cognition is a kind

of computation (where computation involves the manipulation of symbols). The claim is that what the mind does, in part, is to perform cognitive tasks by mental computing. This computational theory of mind provides the fundamental underpinning for most contemporary theories of cognitive science. The basic premise is that much of human cognition can be characterized as a series of computations on mental representations. In medical cognition, **mental representations** are internal states that reflect a clinician's hypothesis about a patient's condition. For example, noticing an abnormal enlargement of the neck region, which prompts the clinician to elicit further inferences about the patient's underlying condition (such as family history of a thyroid condition), may influence the physician's information-gathering strategies and contribute to an evolving problem representation.

In artificial intelligence, symbolic AI is an approach to AI based on the manipulation of knowledge represented in language-like (symbolic) structures in which all relevant semantics (meaning) is explicit in the syntax (formal structure). This also provids a framework for the study of human cognition as the manipulation of symbolic structures. It involves the explicit embedding of human knowledge and behavior rules into computer programs. This type of research in early decades has in recent years been superseded by connectionist AI (neural networks), though in cognitive science both symbolic and connectionist approaches have had periods of historical predominance [17]. All the steps in symbolic AI are based on human-readable representations of the problem that use formal logic. This reasoning process can be easily understood, and a symbolic AI program can therefore explain why a certain conclusion is reached, including the reasoning steps. An explanation that is understandable to human beings helps create a shared meaning of the reasoning process underlying clinical problem-solving, which is an important step in building trust (see Chap. 18).

As the investigations moved from laboratory conditions to realistic clinical environments, it became evident that cognitive factors alone did not account for all the variance in clinicians' performance. Besides cognition, other differences were found to influence decision making, due to socio-cultural, organizational and technological factors. This alerted researchers in their early work to consider the situated nature of the clinical environment in addition to human cognition [18, 19].

Clinical Text Understanding

Early research in language understating lead to development of an influential method of analyzing the process of text understanding or **text comprehension,** based on the assumption that text can be described at multiple levels, from surface codes (e.g., words and syntax) to a deeper level of semantics (meaning) [20, 21]. **Comprehension** refers to cognitive processes associated with understanding or deriving meaning from written text, conversation, or other informational resources. It involves the processes that people use when trying to make sense of

a piece of text, such as a sentence, or a verbal utterance, such as verbal exchanges during a conversation. This work influenced the studies of medical text understanding by physicians at various levels of expertise, where formal methods of natural language representations, such as propositional and semantic representations were used.

Propositions are a form of natural language representation that captures the essence of an idea (i.e., semantics) or concept without explicit reference to linguistic content. Propositional representations constitute an important construct in theories of comprehension. Propositional knowledge can be expressed using a predicate calculus formalism or as a semantic network.

The formalism is informed by an elaborate propositional language [22]. Patel and Frederiksen [23] and Patel and Groen [24] introduced the use of propositional analysis as a method of natural language representation in the clinical domain. The method provided the means to characterize the information clinicians and medical students understood from reading a text, based on their summaries or explanations of the patient problems. Figure 5.1 presents a schematic representation of natural language analysis of clinical text, using a propositional representation representing a **text-based** model and its relationship to semantic and conceptual level analysis, representing a **situational** model [25].

These studies have shown that individuals at different levels of expertise represent clinical text differently [26–29]. This means that these various representations

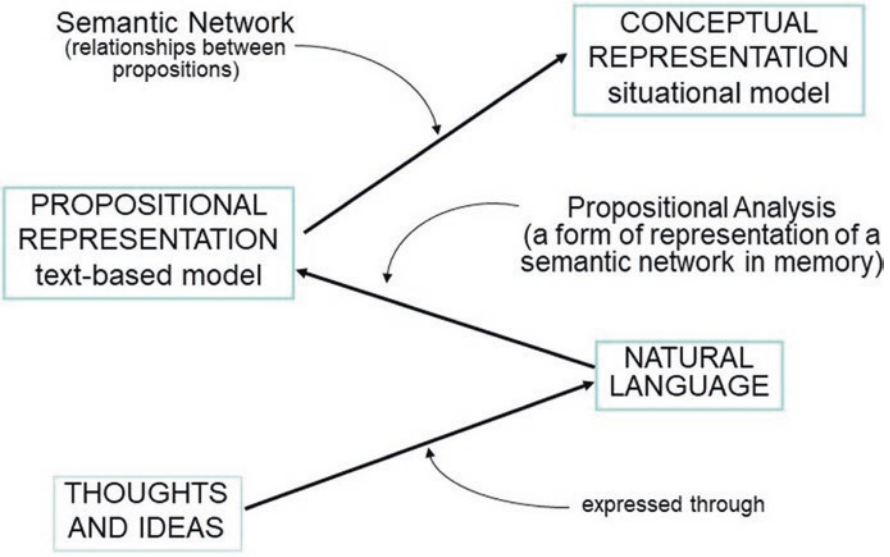

Fig. 5.1 Schematic representation of text (propositions with text-based model) using propositional analysis and its relationship to semantic structure and higher-level conceptual representation (situational model)

will lead to different interpretations of a patient's problem, leading to inconsistent diagnostic decisions. The details of the results show that expert physicians (Board certified in their domain of expertise), are able to separate relevant clinical information that can be used to inform the diagnostic decision-making process, from information that is not pertinent to this process. Non-experts remember considerably more information, but much of this is usually not relevant to the diagnostic decision at hand [26, 27]. Theories and methods of text comprehension have been widely used in the study of medical cognition and have been instrumental in characterizing the process of guideline development and interpretation (for examples see [30]).

Medical expertise is one of those areas of research where the importance of comprehension processes has been demonstrated [24]. Medical problem solving depends on understanding the problem because problem interpretation and analysis in medicine requires construction of appropriate clusters of information in long-term memory that match the current patient presentation. The construction-integration model was developed to account for the process of text comprehension [31, 32]. This model consists of a hybrid symbolic/connectionist architecture developed by Kintsch to account for the process of text comprehension. A model of diagnostic problem solving based on the construction-integration theory involves an interaction between the textbase and the long-term memory store, from which a situation model (Refer to Fig. 5.1) is derived through the cyclical process of construction and integration. A detailed account of how the construction-integration theory is used to explain some important aspects of expertise in medicine is given elsewhere [33]. The authors present a series of studies which serve as evidence for the validity of the construction-integration theory in accounting for the construction of schema during real-time diagnostic reasoning.

The study of medical cognition has been summarized in a series of articles [12, 34] and edited volumes (e.g., [35]). In more recent times, medical cognition is discussed in the context of informatics, generating a new field of investigation, cognitive informatics (for example, [13–15, 36]). Furthermore, foundations of cognition play a significant role in investigations of human computer interaction (HCI), including human factors and patient safety [37].

Clinical Cognition, Reasoning and the Evolution of AI

AI in medicine and medical cognition mutually influenced each other in several ways, including providing a basis for developing formal models of competence in problem-solving tasks. It is not necessary to replicate literally the human mind in order to exhibit intelligent behavior, and besides this may not always be desirable since human beings are error prone. However, in areas such as natural language understanding, commonsense reasoning and the ability to generalize effectively from small numbers of examples, human beings are still far superior to the best

contemporary AI systems. Learning the mechanisms underlying these human abilities could lead to advances in AI. Using techniques and insights drawn from cognitive psychology, more robust and comprehensive AI systems could be built, resulting in models motivated not only by mathematics and a desire to optimize performance, but also by learning from the strengths of human psychology.

Early studies in linking clinical cognition to intelligent systems in medicine began with Anthony Gorry's series of studies in the 70s, comparing a computational model of medical problem solving with the actual problem-solving behavior of physicians [38]. Drawing on this work, others [10] developed a clinical program, where they were guided by the nature and organization of expert knowledge—which was a central concern to both developers of clinical expert systems and researchers in clinical cognition. Medical expert consultation systems, such as INTERNIST-1 [9] and MYCIN [39], introduced ideas about knowledge-based reasoning strategies across a range of cognitive tasks. MYCIN, in particular, had a substantial influence on studies in clinical cognition (see Chap. 2).

A landmark publication that significantly influenced clinical cognition is Newell and Simon's *Human Problem Solving* [40], relating human problem solving to research in artificial intelligence. It described a theoretical framework, extended a language for the study of cognition, and introduced protocol-analytic methods [41] that have become prevalent and dominant methods in investigations of high-level cognition, including the use of this framework for knowledge elicitation techniques in the development decision support systems. This work provided a foundation for the formal investigation of symbolic-information processing (problem solving) approaches.

Protocol analysis is among the most commonly used methods. It refers to a class of techniques for representing verbal **think-aloud protocols**, which are the most common source of data used in studies of problem solving. In these studies, subjects are instructed to verbalize their thoughts as they perform an experimental task. Ericsson and Simon [41] specify the conditions under which verbal reports are acceptable as legitimate data. Data collected during **retrospective think-aloud** protocols, where the subject has had the opportunity to reconstruct the information in memory (with potential for memory distortion), are considered suspect. Think-aloud protocols recorded while collecting observational data in context, provide rich data for the characterization of cognitive processes. In studies of expertise, Patel and colleagues used the think-aloud paradigm to generate sparse data, showing that the use of specific probes could constrain data collection, where subjects were asked to provide explanations for a patient's pathophysiological condition.

Bridging Cognition to Medical Reasoning

The study of expertise is one of the principal paradigms in problem-solving research, which has been documented in a number of volumes in the literature [42–45]. Comparing experts to novices provides us with the opportunity to explore the

aspects of performance that undergo change and result in increased problem-solving skill. A goal of this approach has been to characterize expert performance in terms of the knowledge and cognitive processes used in comprehension, problem solving, and decision making, using carefully developed laboratory tasks [46].

The origin of medical problem-solving research on medical thinking is associated with the seminal work of Elstein and colleagues, who studied the problem-solving processes of physicians by drawing on then-contemporary methods and theories of cognition, based on psychology [11]. Their highly publicized research findings led to an elaborated model of **hypothetico-deductive reasoning**, which proposed that physicians reasoned by first generating and then testing a set of hypotheses to account for clinical data (i.e., reasoning from hypothesis to data). This model of problem-solving has had a substantial influence on studies of medical education. These authors were the first to use experimental methods and psychological theories to investigate problem solving in medicine. Patel and colleagues studied the knowledge-based solution strategies of expert cardiologists as evidenced by their **pathophysiological explanations** of a complex clinical problem [24]. The results indicated that expert physicians who accurately diagnosed the problem, employed a **forward (data-driven) reasoning** strategy—using patient data to lead toward a complete diagnosis (i.e., reasoning from data to hypothesis). This contrasts with subjects who misdiagnosed or partially diagnosed the patient problem. They tended to use a **backward or hypothesis-driven reasoning** strategy. Figure 5.2 shows a diagrammatic representation of data-driven reasoning. From the presence of puncture wound mark on the arm to a young unemployed male, (clinical findings on the *left side of figure*), the physician reasons forward to conclude the diagnosis of infection (*right side of the figure*). Figure 5.3 shows a representation of hypothesis-driven reasoning. When making the diagnosis of myxedema, the physician explains an inconsistent finding of respiratory failure to be the result of a hypometabolic state of the patient.

Although expert clinicians, in their own domain of expertise, typically use data-driven reasoning or general heuristics during clinical tasks, this type of reasoning sometimes breaks down, and the physician must resort to hypothesis-driven

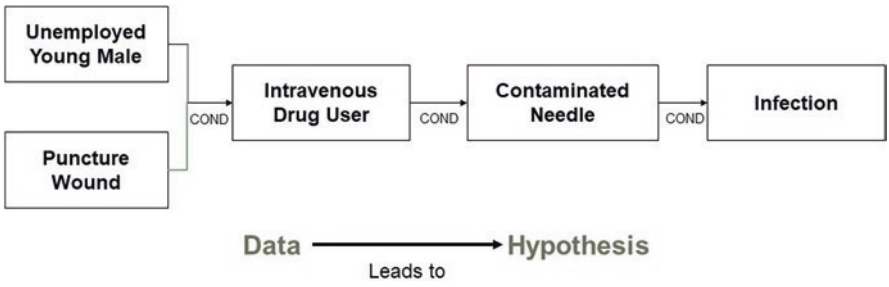

Fig. 5.2 A diagrammatic representation of data-driven reasoning when an unemployed young male presents with fever and a puncture wound mark on the arm. Presenting signs and symptoms through data-driven inferences, indicated likelihood of this patient being an intravenous drug user, with possible use of a contaminated needle, leading to infection. *COND* refers to a conditional relation, based on propositional analysis. Arrows indicate directionality

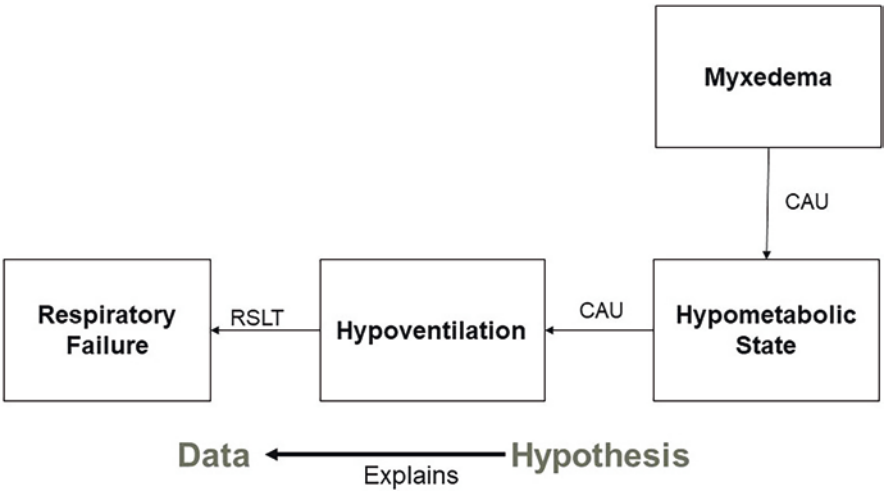

Fig. 5.3 A diagrammatic representation of hypothesis-driven reasoning. An anomalous finding of respiratory failure, which is inconsistent with the main diagnosis (myxedema), is accounted for as a result of a hypometabolic state of the patient, in a backward-directed inference. *CAU* indicates a causal relation, and *RSLT* identifies a resultive relation in propositional analysis. Arrows indicate directionality

reasoning. In everyday practice, both types of reasoning are used. Forward directed reasoning was found to be the hallmark of expertise, as shown in other knowledge-based domains, such as physics [47]. Although data-driven reasoning is highly efficient, it is often error-prone in the absence of adequate domain knowledge, since there are no built-in checks on the legitimacy of the inferences that a person makes. In contrast, hypothesis-driven reasoning is slower and may make heavy demands on working memory, because one must keep track of goals and hypotheses. It is, therefore, most likely to be used when there is uncertainty, domain knowledge is inadequate, or the problem is complex. This type of reasoning is not used in regular time-constrained practice because details interfere with the utility of efficient decision making. Other chapters in the book discuss the concepts of forward and backward chaining in systems (Chaps. 3 and 4). It should be noted that forward reasoning by expert systems consists of straightforward chaining of rules, whereas the forward reasoning of human experts invariably has missing steps in the inferencing process [28]. This indicates that forward reasoning may be generated by a process considerably more complex than the simple chaining of rules.

Hypothesis-driven reasoning is usually exemplary of a *weak method* of problem solving in the sense that is used in the absence of relevant prior knowledge and when there is uncertainty about a problem solution. In problem-solving terms, strong methods engage knowledge, whereas weak methods refer to general strategies that do not. Weak does not necessarily imply ineffectual in this context. Furthermore, hypothesis-driven reasoning may be more conducive to the novice learning experience in that it can guide the organization of knowledge [48]. Causal reasoning as part of the backward reasoning is an indispensable part of human thought, and it has been argued that formalizing it is a prerequisite to achieving human-level machine

intelligence [49]. These types of reasoning relate to Kahneman's "fast" and "slow", models of reasoning [50], where the author proposes two types of reasoning corresponding to two different components of the human brain. There are identified as **System 1** and **System2**. System 1 processes information fast, but is slow to learn, since it learns through experience—often through sensory perception and pattern matching strategies—and it is error prone. System 2 processes information slowly, but is fast to learn. It learns from theory through explanatory processes with a logical inference engine, and is relatively reliable because it has built in error checks. This process is effortful and is triggered under uncertain conditions. The characterization of the two systems is not unlike the forward and backward reasoning in medical decision making developed by Patel and Groen and described above. The authors showed a formal relationship between comprehension and problem solving [51] in clinical medicine. The recognition of the relationship between the cognitive studies in clinical comprehension and problem solving, and AI dates back at least to 1991, when the two keynote presentations at the *Artificial Intelligence-Europe* meeting in Maastricht, Netherlands discussed the two topics and their synergies [28, 52]. These relationships show that collaboration among cognitive science, AI and neuroscience can produce an understanding of the mechanisms in the brain that generate human cognition. Thus, it is important to build AI systems with the ability to understand, think, reason and learn flexibly and rapidly, which will require deeper understanding of how the human mind functions as we do our tasks.

Models of Medical Reasoning

It is generally accepted there are two basic forms of reasoning: **deductive reasoning**, which in medicine consists of deriving a diagnosis (conclusion) from diagnostic category or a pathophysiological process (hypothesis). The other form is **inductive reasoning**, which consists of generating a diagnosis (conclusion), from patient data. However, reasoning in the "real world" does not fit neatly into any of these basic reasoning types. A third form of reasoning was identified as best capturing the generation of clinical hypotheses, where deduction and induction are intermixed. This is termed **abductive reasoning** [53], which is based in philosophy and is illustrated by the clinician generating a plausible explanatory hypothesis through a process of heuristic rule utilization (see for example, [54]).

Abductive reasoning is thought of as a cyclical process of generating possible explanations (i.e., identification of a set of hypotheses that are able to account for the clinical case on the basis of the available data) from a set of data and testing those explanations (i.e., evaluation of each generated hypothesis on the basis of its expected consequences) for the abnormal state of the patient at hand [11, 55–57]. Abductive reasoning is a data-driven process and dependent on domain knowledge. Within this generic framework, various models of diagnostic reasoning may be constructed. Following Patel and Ramoni [58], we can distinguish between two major models of diagnostic reasoning: **heuristic classification** [59] and **cover and differentiate** [60]. However, these models can be seen as special cases of a more general model: the **select and test** model [57], where the processes of hypothesis

generation and testing can be characterized in terms of four types of processes: abstraction, abduction, deduction, and induction.

During **abstraction,** pieces of data in the data set are selected according to their relevance for the problem solution and chunked in **schemas** representing an abstract description of the problem at hand (e.g., abstracting that an adult male with hemoglobin concentration less than 14 g/dL is an anemic patient). Following this, hypotheses that could account for the current situation are related through a process of *abduction*, characterized by a "backward flow" of directed inferences. This model of reasoning can be used to explain the medical diagnostic process. Expert clinicians are selective in the data they collect (**abstraction**), focusing only on the data that are relevant to the generated hypotheses, while ignoring other less-relevant data [24, 27]. Successful clinicians focus on the fewest pieces of data and are better able to integrate these pieces of data into a coherent explanation for the problems [61]. Typically, physicians generate a small set of hypotheses very early in the case (**abduction**), as soon as the first pieces of data become available, as was first shown by Elstein's group [11], and later corroborated by other researchers (For example, [62, 63]). Physicians sometimes make use of the hypothetico-deductive process (**deduction**), which involves four stages: cue acquisition, hypothesis generation, cue interpretation, and hypothesis evaluation [11]. The reader is referred to the comprehensive summary of the research in clinical reasoning provided by Patel and colleagues in a recent book chapter [34]. The complex nature of clinical reasoning and decision making illustrates why is it so difficult to develop intelligent systems that can behave like human beings.

Knowledge Organization, Expert Perception and Memory

The discussion so far has focused more on expertise and the processes of diagnostic reasoning. Research has also revealed differences in knowledge representation with levels of expertise. A recurring finding from studies of expertise is that experts represent knowledge at a higher level of abstraction than their less experienced counterparts [64]. For example, Norman and colleagues investigated the ability of clinicians of different levels of dermatology expertise to make clinical diagnoses based on images presented as slides. Experts were more accurate in their diagnoses, and also exhibited a tendency to categorize slides at higher levels of abstraction. A similar finding was found in the study of expertise in radiology: less experienced subjects focused on surface anatomical features, while experienced radiologists developed deeper, more principled problem representations [65]. While this was not unexpected in *visual* domains of medicine, Patel and her colleagues identified an analogous difference in levels of abstraction in *verbal* problem solving, with expert physicians tending to represent case information from written scenarios at a higher level of abstraction than novice physicians [33]. Specifically, experts are distinguished by their emphasis on the *facet* level [66], which represents intermediate solutions to diagnostic problems. An example might be the cluster of symptoms

associated with congestive cardiac failure—once these are recognized a specific *diagnosis* that explains the cause of the congestive cardiac failure can be sought. For experts, these facet-level pre-diagnostic hypotheses serve as intermediate steps in a diagnostic process, narrowing down the space of possible solutions to mediate effective problem solving. In addition, the aggregation of information into larger, meaningful units allows expert problem solvers to represent complicated cases within the laboratory-determined constraints on working memory capacity (famously, 7+-2 units of information) [67]. Such patterns of knowledge organization have immediate implications for the design of AIM systems. Adler-Milstein and her colleagues use the analogy of "wayfinding" to describe the use of AI to support the process of diagnosis by gathering, organizing and prioritizing information that is germane to the solution of a diagnostic problem [3]. How then, should the information be organized once gathered? The section on "AI, Machine Learning, and Human Cognition" considers how what is known about clinical knowledge organization and decision making might be used to guide this process.

Understanding Clinical Practice for AI Systems

The Role of Distributed Cognition

The work discussed in previous sections has focused on the cognitive processes of individual decision makers, often captured in laboratory experiments. However, toward the turn of the twenty-first century, a new paradigm of cognitive research emerged, known as **distributed cognition** [68]. Distributed cognition broadens the focus of cognitive research, moving from the study of individuals in laboratory settings to the study of groups of individuals at work in naturalistic environments. For example, Hutchins, a seminal figure in the field, conducted his influential work on navigation aboard naval vessels at sea [68]. A pragmatic advantage of this approach to research is that while representations in the mind (**internal representations**) cannot be observed directly, representations that occur in the work environment (**external representations**) can be recorded and studied. A famous example of an external representation concerns the "speed bug", a positionable plastic pointer that slides around the edge of the speedometer and can be used to demarcate appropriate landing speeds once these have been retrieved from a reference book [69]. This example is illustrative of a fundamental idea in distributed cognition: that an individual (or team of individuals) in a work environment constitute a composite cognitive system—a symbol processing system—with greater functionality than any of its individual components. From this perspective, the reference book of acceptable speeds is part of the long-term memory of the system, and the speed bug—a **cognitive artifact**—is part of its working memory [69]. In previous research, a significant paradigm shift was seen from a focus on individual cognition to collaborative and distributed cognition in healthcare. A special issue of the journal *AI in Medicine* included five original articles by prominent scholars that present complementary

approaches to collaboration and distributed cognition in health and medicine, emphasizing situations where collaboration is between human and computer or facilitated by computers [70]. On account of the prominent role of cognitive artifacts such as whiteboards and different sorts of clinical notes in clinical practice, distributed cognition has proved to be an informative way to characterize such settings [71, 72], and identify opportunities to design tools that support their cognitive work [73].

As an illustrative example, Cohen and his colleagues used the distributed cognition paradigm to characterize the distribution of cognitive work in a psychiatry emergency department [71]. The work revealed ways in which cognition was distributed across teams and cognitive artifacts (such as written notes, see Fig. 5.4), and also over time, with these cognitive artifacts serving as bridges to maintain the continuity of cognitive tasks despite frequent staffing changes.

Considering a clinical environment from this perspective can lead to a more holistic picture of the ways in which AI technologies can offer support than the prevailing approaches of automated diagnostic decision making or prediction of adverse outcomes, including support for such cognitive tasks as information search, aggregation and synthesis [74].

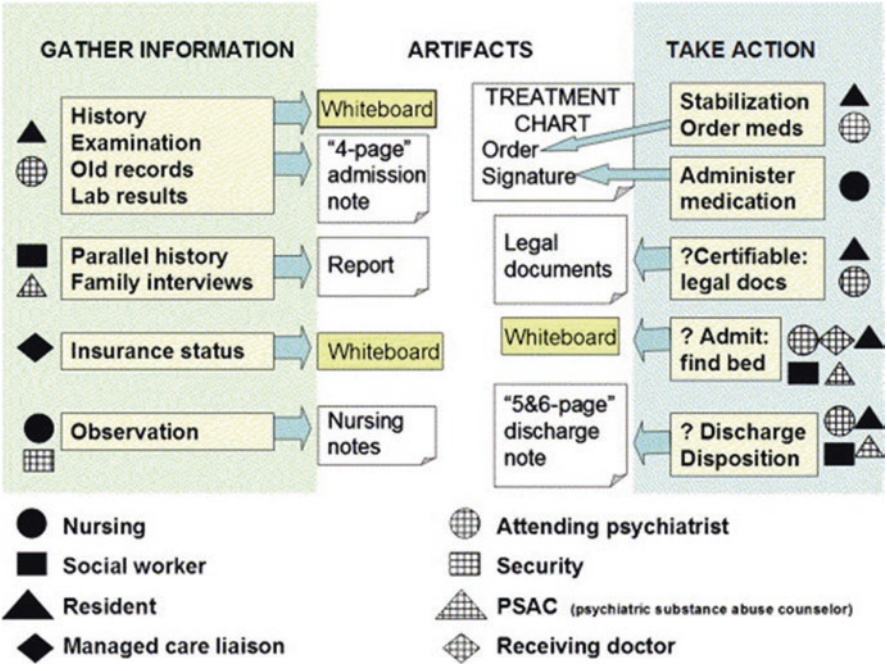

Fig. 5.4 Distribution of cognitive tasks in a psychiatric emergency department. The tasks, broadly categorized into information gathering tasks, and those involving actions taken on the basis of this information, are supported by a range of cognitive artifacts such as specific document types and the departmental whiteboard. Both the internal (mental) representations of the staff members and the external physical representations on these artifacts support the cognitive work required

AI, Machine Learning, and Human Cognition

The AI of today is a natural evolution of what we have seen over recent decades. For example, the deep neural networks currently used to classify images in radiology and other medical domains originated in the twentieth century [75, 76]. The changes, the reasons we are seeing AI in every aspect of life, appear to be less about AI advancement itself than they are about data generation and our current ability to leverage advanced computational power. However, there are certain barriers to the rapid growth of AI that are unlikely to be overcome by data and computational power alone. These barriers demonstrate that the path to the advancement of AI can be tricky and challenging. Present AI systems do not have a deep understanding—an understanding that integrates new observations with prior structured knowledge—but, rather, a shallow intelligence, that is the ability to emulate and, in the context of constrained tasks, sometimes even to improve upon some human pattern recognition and perception abilities. One cannot deny that there is intelligence in AI systems, but it does not follow the same rules as humans do.

The major goal of AI is to push forward the frontier of machine intelligence. Before going any further, it may be important to introduce a few terms. Machine learning and **deep learning** are two subsets of artificial intelligence which have garnered a lot of attention over the past few years. Many machine learning applications aim to allow computers to analyze and act with less human intervention by learning from training data. Deep learning—itself a type of machine learning—aims to support analyses that use multilayered structures inspired by the neural connectivity of the human brain (see Chap. 6). While many other machine learning methods require less training data and computing power than deep learning, deep learning methods typically need less human intervention because they have the capacity to learn useful representations of incoming data by themselves, obviating the need for these to be engineered manually. Deep learning can be viewed as a statistical technique for recognizing patterns in sample data, using neural networks with multiple layers, where there is an attempt at imitating (albeit superficially) the structure and function of neural networks in the human brain. An important advantage of deep neural networks is that they are able to learn useful representations while training. For example, in image processing a deep learning model may propagate data through different layers of the network, with each layer successively learning to recognize higher level image features that collectively suggest a label, as learned from training data. This is similar in some ways to how expert problem solvers work—using abstraction to relate their observations to previously learned hierarchies of concepts and relations in order to find an answer. However, there are important differences between these processes.

Consider the case of text comprehension. Human beings, as they process texts, frequently derive a wide range of inferences, as explained earlier. Deep learning currently struggles with open-ended inference based on real-world knowledge at the level of human accuracy [77]. Furthermore, human reasoners have the capability to explain the sequences of inferences that drive their decision making processes.

However, the propagation of representations from layer to layer of a deep neural networks, en route to a prediction, defies explanation in human terms. This transparency issue is a fundamental concern when using deep learning for problem domains like medical diagnosis, where clinicians need to understand how a given system made a decision. Problems that have to do with commonsense reasoning are usually outside the scope of deep learning. Human beings solve even simple problems by integrating knowledge across vastly disparate sources. In medicine these may include observational data, knowledge of clinical science, laboratory data and so forth. This is not true for the majority of deep learning models, which learn complex statistical correlations among input and output features, but with no inherent representation of causality or associated domain knowledge. We need to reach human-level cognitive flexibility if we are to see AI models reach human-like performance. These issues are well addressed in recent scholarly literature [77–79]. However, such flexible human-like performance is not a prerequisite to improving healthcare with AI. Contemporary AI methods can already perform constrained tasks with human-like accuracy, and have other capabilities—such as the ability to process large amounts of data quicky—that can be leveraged to support human decision makers.

Reinforcing the Human Component

Artificial intelligence is poised to transform the healthcare industry. By developing new data analytics, intelligent clinical systems can analyze large and varied data sets, and clinicians can easily access the information they need to deliver care to their patients. AI and **augmented intelligence** have similar goals but differ in the way of achieving them. Augmented intelligence is like AI in that both fields use machine learning to enhance performance. However, instead of replacing human intelligence, augmented intelligence aims to use AI methods to build upon it in an assistive role. This change in emphasis has broad implications. Technologies mediate human performance, and influence the way people behave as they interact with them. This goes beyond merely supporting, enhancing or expediting performance. Tools, and artifacts not only enhance people's ability to perform tasks but also change the way in which they do so. The following sections provide some examples of how AI systems can be used to augment human cognition in medicine.

Augmenting Clinical Comprehension

One approach to leveraging what is known about medical cognition to inform the design of AIM systems involves using approaches that deliberately emulate the knowledge organization of expert clinicians. As an illustrative example, Fig. 5.5 shows one of four views of a narrative text discharge summary (from a fictional

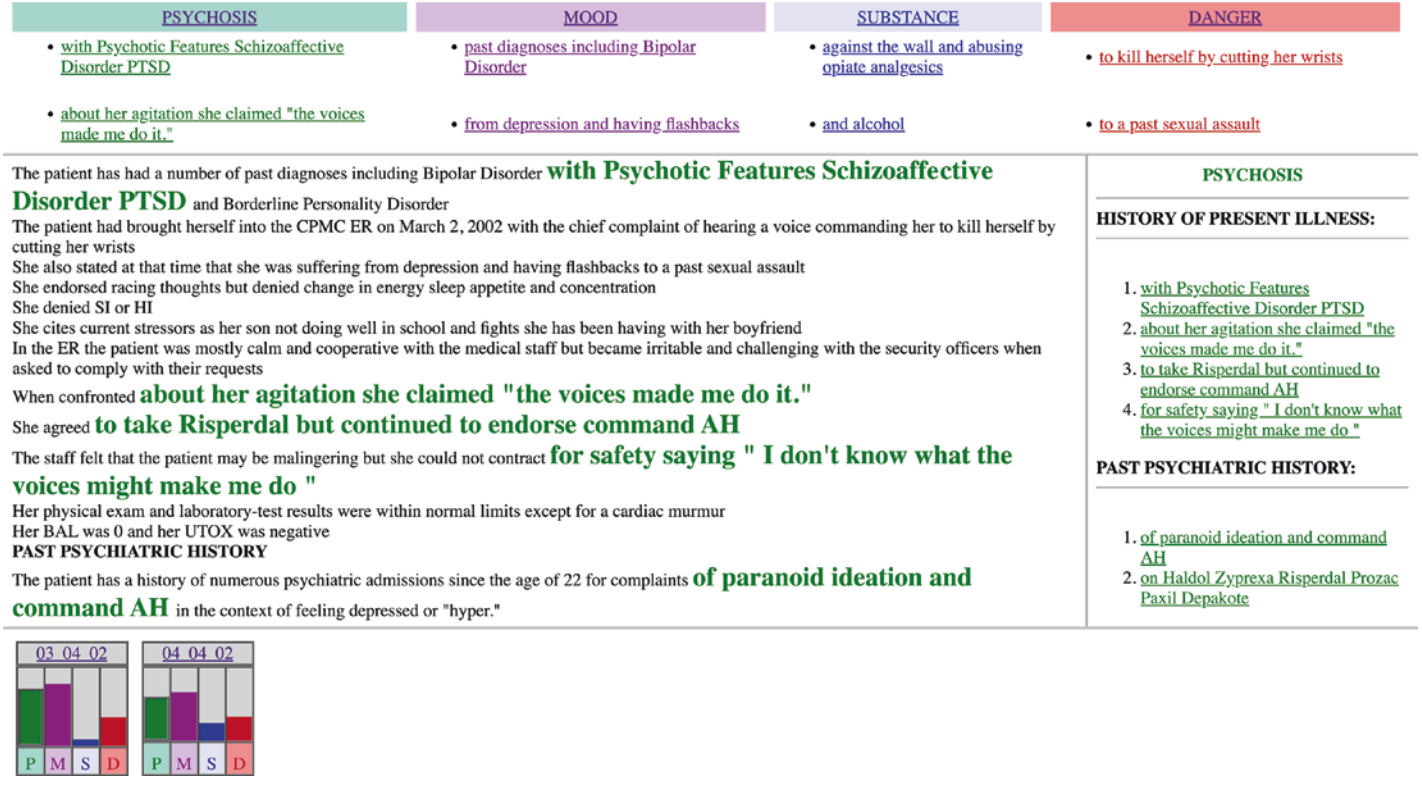

Fig. 5.5 View of a psychiatry discharge summary emphasizing psychosis-related elements. The summary was adapted from a text-based narrative developed in prior research, based on a case example from a textbook [80, 81]

patient encounter developed for research purposes) provided by a system that combines supervised machine learning with semantic word vector representations to draw connections between phrases in text and the diagnostically and prognostically important facet-level constructs of psychosis, mood, substance abuse and dangerousness [82, 83]. The figure shows a view that emphasizes phrases related to psychosis, such as those mentioning auditory hallucinations or paranoid ideation (as well as phrases mentioning antipsychotic medications such as Risperdal). Relevant phrases are also presented in the top and rightmost panels, and both these phrases and the four facets in the top panel serve as links to accommodate navigation, and switch perspective to emphasize the facet concerned. The interface also provides a graphical summary (bottom panel) of *other* narrative text records that indicates the extent to which content from each facet is represented, to facilitate exploration of historical narratives at a conceptual level that is conducive to problem solving. Evaluation of the interpretation of two case scenarios by 16 psychiatry residents revealed that the interface supported clustering of case-relevant information, with more detailed case recollection and better diagnostic accuracy in the more complex of the two scenarios when the interface was used [84]. In addition, residents using the interface better attended to clinically relevant elements of the case that had been neglected by non-expert participants in previous work [80], including important indicators of potential dangerousness to self and others. Qualitative evaluation of verbal think-aloud protocols captured during the process of exploring the cases using the interface revealed patterns of navigation used by residents to explore hypotheses at the facet level. These studies demonstrate the potential for AI to augment human decision making by simulating expert knowledge organization to reveal patterns in clinical data, rather than making decisions or predictions directly. From a distributed cognition perspective, the simulations of the knowledge and **retrieval structures**—structures that would typically support efficient decision making in the minds of the experts—are part of a larger cognitive system that includes residents, the interface and the AIM models that underlie it.

Supporting Specific Cognitive Tasks

The preceding section describes a system that was developed to support trainees (residents) by simulating knowledge organization and retrieval structures that are characteristic of expert medical cognition, and expertise in general. It is also possible to design systems to support the thought processes underlying a specific task, that have been characterized using cognitive methods. For example, Baxter and his colleagues describe the use of a **cognitive task analysis**—a systematic approach for collecting information about the mental processes underlying a particular task [85]—to inform the development of an expert system named FLORENCE to support decision making about ventilator settings in the context of neonatal respiratory distress [86]. This work involved a detailed characterization of the tasks, actors, communication events, documents and instruments in the neonatal intensive care unit concerned, resulting in a number of design implications for the system. These

included practice recommendations for staff that identified contingencies in which the system's suggestions may be unreliable, the need for a distinctive alarm that would stand out from those already prevalent in the environment, and the incorporation of mnemonic devices already used by staff into the wording of the system's recommendations. These design implications were all informed by what had been learned about the cognitive capabilities of the team in the unit: their ability to recognize anomalous data that may lead to untrustworthy recommendations, the potential for their awareness of one alert to be drowned out by others, and the aids they use to remember procedural tasks developed to preemptively address potential causes of faulty readings that could mislead FLORENCE.

Mental Models of AI Systems

Interestingly, many of the design implications that emerged from the aforementioned cognitive task analysis concerned devising ways for human team members to recognize or preempt conditions under which an AIM system is likely to be incorrect. This requires having a mental model the system, akin to those shown to enhance learning to use devices in general [87]. Bansal and his colleagues provide empirical evidence that an accurate mental model of such conditions is fundamental to effective team performance in AI-advised decision making [88]. In these experiments, which were conducted with crowdsourced workers in the context of a simulated AI-advised task, better overall team performance was observed when using systems with error-prone conditions that were easier to understand because they depended upon fewer data features, and consistently led to a system error. The benefits of consistent model performance have also been shown in prior work by this group related to updating machine learning models, which was shown to have detrimental effects on overall team performance when it led to changes in decision-making on previously-observed examples [89]. These findings are also consistent with subsequent work showing that more accurate mental models of AI systems lead to better collaborative performance on word games [90]. Related work has investigated mediation of the development of accurate mental models of AI systems [91], and how such mental models are revised in response to surprising behavior [90]. While these findings mostly emerged from work outside the medical domain, they have clear implications for the development of AIM systems, and characterization of healthcare provider's mental models of AIM is an important area for future cognitive informatics research.

Conclusion

The influence of technology is not best measured quantitatively alone, since it is often qualitative in nature. The importance of cognitive factors that determine how human beings comprehend information, solve problems, and make decisions cannot

be overstated. Investigations into the process of medical reasoning have been one such area where advances in cognitive science have made significant contributions to AI.

At the AI in Medicine conference in Amsterdam in 2009, researchers raised the question of whether we have forgotten about the role of the human mind as we perform our tasks in the evolution of AIM research [92]. This question is still salient today, perhaps more salient given that technological advances have surpassed our understanding of human behavior in such complex socio-technical environments. Today, a new question is whether we are getting the most out of our AIM inventions. It is time to reshape the current innovative technologies to serve human beings and augment our activities. In the clinical world, such augmented intelligence can provide clinicians with additional assistance they need to deliver a better quality of care for their patients.

Questions for Discussion

- Discuss, with examples, how the knowledge of cognitive science foundations can provide a better understanding of human-technology collaboration for developing contemporary AI systems for clinical practice. Can you think of principles of some of the component subfields of cognitive science that may also be valuable in such collaborative efforts?
- What are the ways to augment human intelligence for safer clinical practice, given what we know about current medical AI systems? Consider known limitations of human cognition, such as a propensity toward bias in diagnostic decision making and constraints on attentions span and working memory, how these limitations may manifest as vulnerabilities to medical error, and how AI methods may be used to preempt these patient safety concerns.
- Consider the potential and limitations of symbolic representation of knowledge in AI systems, and ways to circumvent these limitations with more contemporary approaches. Conversely, consider the limitations of contemporary deep learning models. How might the limitations of these approaches be addressed through incorporation of symbolic approaches, and vice versa?

Further Reading

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(1):5–17.

- The section on “Clinical Cognition, Reasoning and the Evolution of AI” of this paper argues for the importance of cognitive factors in the design of medical AI systems, and introduces many of the topics developed in this chapter.

Clark A. Natural-born cyborgs? International conference on cognitive technology. Berlin: Springer; 2001. p. 17–24.

- This book provides a readable introduction to the framework of distributed cognition, and its role in technology design.

Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, Janda M, Lallas A, Longo C, Malvehy J, Paoli J, Puig S, Rosendahl C, Soyer HP, Zalaudek I,

Kittler H. Human–computer collaboration for skin cancer recognition. Nat Med. 2020;26(8):1229–34.

- This paper goes a step beyond establishing the benefits of the human-AI collaborative diagnosis in perceptual aspect of cognition, by investigating relationships between representation of AI output and diagnostic accuracy.

Patel VL, Kaufman DR. Cognitive science and biomedical informatics. In: Shortliffe EH, Cimino JJ, Chiang, M, editors. Biomedical informatics: computer applications in health care and biomedicine. 5th ed. Chap 4. New York: Springer; 2021.

- This chapter introduces cognitive research in healthcare and informatics, a discipline referred to as cognitive informatics. It presents the basic theoretical underpinnings of cognitive science with a focus on information-processing, natural language representation and distributed cognition frameworks.

References

1. Chan S, Siegel EL. Will machine learning end the viability of radiology as a thriving medical specialty? Br J Radiol. 2018;92(1094):20180416. https://doi.org/10.1259/bjr.20180416.
2. Jha S, Topol EJ. Adapting to artificial intelligence: radiologists and pathologists as information specialists. JAMA. 2016;316(22):2353–4. https://doi.org/10.1001/jama.2016.17438.
3. Adler-Milstein J, Chen JH, Dhaliwal G. Next-generation artificial intelligence for diagnosis: from predicting diagnostic labels to "wayfinding". JAMA. 2021;326(24):2467–8. https://doi.org/10.1001/jama.2021.22396.
4. Dreyer KJ, Geis JR. When machines think: radiology's next frontier. Radiology. 2017;285(3):713–8. https://doi.org/10.1148/radiol.2017171183.
5. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology. 2017;284(2):574–82. https://doi.org/10.1148/radiol.2017162326.
6. Patel BN, Rosenberg L, Willcox G, Baltaxe D, Lyons M, Irvin J, Rajpurkar P, Amrhein T, Gupta R, Halabi S, Langlotz C, Lo E, Mammarappallil J, Mariano AJ, Riley G, Seekins J, Shen L, Zucker E, Lungren MP. Human–machine partnership with artificial intelligence for chest radiograph diagnosis. NPJ Digit Med. 2019;2(1):1–10. https://doi.org/10.1038/s41746-019-0189-7.
7. Hekler A, Utikal JS, Enk AH, Hauschild A, Weichenthal M, Maron RC, Berking C, Haferkamp S, Klode J, Schadendorf D, Schilling B, Holland-Letz T, Izar B, von Kalle C, Fröhling S, Brinker TJ, Schmitt L, Peitsch WK, Hoffmann F, et al. Superior skin cancer classification by the combination of human and artificial intelligence. Eur J Cancer. 2019;120:114–21. https://doi.org/10.1016/j.ejca.2019.07.019.
8. Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, Janda M, Lallas A, Longo C, Malvehy J, Paoli J, Puig S, Rosendahl C, Soyer HP, Zalaudek I, Kittler H. Human–computer collaboration for skin cancer recognition. Nat Med. 2020;26(8):1229–34. https://doi.org/10.1038/s41591-020-0942-0.
9. Miller RA, Pople HE, Myers DJ. Internist-I, an experimental computer-based diagnostic for general internal medicine. In: Clancey WJ, Shortliffe EH, editors. Readings in medical artificial intelligence. Reading, MA: Addison-Wesley; 1984. p. 190–209.
10. 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.

11. Elstein AS, Shulman LS, Sprafka SA. Medical problem solving: an analysis of clinical reasoning. Cambridge: Harvard University Press; 1978.
12. Patel VL, Arocha JF, Kaufman DR. Diagnostic reasoning and medical expertise. In: Medin DL, editor. The psychology of learning and motivation: advances in research and theory, vol. 31. San Diego: Academic Press, Inc.; 1994. p. 187–252.
13. Patel VL, Kaufman DR, Cohen T. Cognitive informatics in health and biomedicine: case studies on critical care, complexity and errors. London: Springer; 2014.
14. Patel VL, Kaufman D, Cohen T, editors. Cognitive informatics in health and biomedicine: case studies on critical care, complexity and errors with preface by Vimla L. Patel. London: Springer; 2014.
15. Patel VL, Kannampallil TG. Cognitive informatics in biomedicine and healthcare. J Biomed Inform. 2015;53:3–14.
16. Patel VL, Kaufman DR. Cognitive science and biomedical informatics. In: Shortliffe EH, Cimino JJ, Chiang M, editors. Biomedical informatics: computer applications in health care and biomedicine. 5th ed., Chap 4. New York: Springer; 2021.
17. Medler DA. A brief history of connectionism. Neural Comput Surveys. 1998;1:18–72.
18. Patel VL, Kaufman DR, Arocha JF. Steering through the murky waters of a scientific conflict: situated and symbolic models of clinical cognition. Artif Intell Med. 1995;7:413–38.
19. Patel VL, Kaufman DA, Arocha JF. Emerging paradigms of cognition and medical decision making. J Biomed Inform. 2002;35:52–75.
20. Kintsch W. Comprehension: a paradigm for cognition. Cambridge/New York: Cambridge University Press; 1998.
21. van Dijk TA, Kintsch W. Strategies of discourse comprehension. New York: Academic; 1983.
22. Frederiksen CH. Representing logical and semantic structure of knowledge acquired from discourse. Cogn Psychol. 1975;7(3):371–458.
23. Patel VL, Frederiksen CH. Cognitive processes in comprehension and knowledge acquisition by medical students and physicians. In: Schmidt HG, de Volder MC, editors. Tutorials in problem-based learning. Assen, Holland: van Gorcum; 1984. p. 143–57.
24. Patel VL, Groen GJ. Knowledge based solution strategies in medical reasoning. Cogn Sci. 1986;10(1):91–116.
25. Kintsch W. The representation of meaning in memory. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers. 1974.
26. Patel VL, Groen GJ. Developmental accounts of the transition from medical student to doctor: some problems and suggestions. Med Educ. 1991;25(6):527–35.
27. Patel VL, Groen GJ. The general and specific nature of medical expertise: a critical look. In: Ericsson KA, Smith J, editors. Toward a general theory of expertise: prospects and limits. New York: Cambridge University Press; 1991. p. 93–125.
28. Patel VL, Groen GJ. Real versus artificial expertise: the development of cognitive models of clinical reasoning. In: Stefanelli M, Hasman A, Fieschi M, Talmon J, editors. Lecture notes in medical informatics (44). Proceedings of the third conference on artificial intelligence in medicine. Berlin: Springer; 1991. p. 25–37.
29. Patel VL, Kaufman DR. Medical informatics and the science of cognition. J Am Med Inform Assoc JAMIA. 1998;5(6):493–502.
30. Peleg M, Gutnik LA, Snow V, Patel VL. Interpreting procedures from descriptive guidelines. J Biomed Inform. 2006;39:184–95.
31. Kintsch W. The role of knowledge in discourse comprehension: A construction integration model. Psychological Review. 1988;95:163–82.
32. Kintsch W, Welsch DW. The construction-integration model: a framework for studying memory for text. In: Hockley WE, Lewandowsky S, editors. Relating theory to data: essays on human memory in honor of Bennet Murdock. Hillsdale, NJ: Lawrence Erlbaum Associates; 1991. p. 367–85.
33. Arocha JF, Patel VL. Construction-integration theory and clinical reasoning. In: Weaver CA, Mannes S, Fletcher CR, editors. Discourse comprehension: essays in honor of Walter Kintsch; 1995. p. 359–81.

34. Patel VL, Kaufman DR, Kannampallil TG. Diagnostic reasoning and expertise in healthcare. In: Ward P, Schraagen JM, Gore J, Roth E, editors. The Oxford handbook of expertise: research & application. Oxford University Press; 2018.
35. Evans DA, Patel VL, editors. Cognitive science in medicine: biomedical modeling. Cambridge, MA: MIT Press; 1989.
36. Zheng K, Westbrook J, Kannampallil T, Patel VL, editors. Cognitive informatics: reengineering clinical workflow for more efficient and safer care. London: Springer; 2019.
37. Patel VL, Kaufman DR, Kannampallil TG. Diagnostic reasoning and decision making in the context of health information technology. In: Marrow D, editor. Reviews of human factors and ergonomics, vol. 8. Thousand Oaks, CA: Sage; 2013.
38. Gorry GA. Computer-assisted clinical decision-making. Methods Inf Med Suppl. 1973;7:215–30. PMID: 4617100.
39. Shortliffe EH. Computer-based medical consultations: MYCIN. New York: Elsevier; 1976.
40. Newell A, Simon HA. Human problem solving. Englewood Cliffs, NJ: Prentice-Hall. 1972.
41. Ericsson KA, Simon HA. Protocol analysis: verbal reports as data. Rev. ed. Cambridge, MA: MIT Press; 1993.
42. Ericsson KA. The road to excellence: the acquisition of expert performance in the arts and sciences sports and games. Mahwah: Lawrence Erlbaum Associates; 1996.
43. Ericsson KA. The Cambridge handbook of expertise and expert performance. Cambridge/New York: Cambridge University Press; 2006.
44. Ericsson KA, Smith J. Toward a general theory of expertise: prospects and limits. New York: Cambridge University Press; 1991.
45. Ericsson KA, Hoffman RR, Kozbelt A, Williams AM, editors. The Cambridge handbook of expertise and expert performance. Cambridge University Press; 2018.
46. Chi MTH, Glaser R. (1981). Categorization and representation of physics problems by experts and novices. Cognitive Science, 1981;5:121–52.
47. Larkin JH, McDermott J, Simon DP, Simon HA. Models of competence in solving physics problems. Cogn Sci. 1980;4(4):317–45. https://doi.org/10.1207/s15516709cog0404_1.
48. Patel VL, Evans DA, Kaufman DR. Reasoning strategies and use of biomedical knowledge by students. Med Educ. 1990;24:129–36.
49. Pearl J, Mackenzie D. The book of why: the new science of cause and effect. Basic Books; 2018.
50. Kahneman D. 35. Two selves. In: Thinking, fast and slow. New York: Farrar, Straus & Giroux; 2011.
51. Groen GJ, Patel VL. Relationship between comprehension and reasoning in medical expertise. In: Chi M, Glaser R, Farr M, editors. The nature of expertise. Hillsdale, NJ: Lawrence Erlbaum; 1988. p. 287–310.
52. Shortliffe EH. The adolescence of AI in medicine: will the field come of age in the 90s? Artificial Intelligence in Medicine 1993;5:93—106.
53. Peirce CS. Abduction and induction. In C. S. Peirce & J. Buchler (Eds.), Philosophical writings of Peirce. New York, NY: Dover; 1955a. pp. 150–6.
54. Magnani L. Abduction, reason, and science: processes of discovery and explanation. Dordrecht: Kluwer Academic; 2001.
55. Joseph GM, Patel VL. Domain knowledge and hypothesis generation in diagnostic reasoning. Med Decis Mak. 1990;10:31–46.
56. Kassirer JP. Diagnostic reasoning. Ann Intern Med. 1989;110:893–900.
57. Ramoni M, Stefanelli M, Magnani L, Barosi G. An epistemological framework for medical knowledge-based systems. IEEE Trans Syst Man Cybern. 1992;22(6):1361–75.
58. Patel VL, Ramoni MF. Cognitive models of directional inference in expert medical reasoning. In: Feltovich PJ, Ford KM, Hoffman RR, editors. Expertise in context: human and machine. Cambridge: The MIT Press; 1997. p. 67–99.
59. Clancey WJ. Heuristic classification. Artif Intell. 1985;27:289–350.
60. Eshelman L. MOLE: a knowledge acquisition tool for cover-and-differentiate systems. In: Marcus SC, editor. Automating knowledge acquistion for expert systems. Boston: Kluwer; 1988. p. 37–80.

61. Groves M, O'Rourke P, Alexander H. Clinical reasoning: the relative contribution of identification, interpretation and hypothesis errors to misdiagnosis. Med Teach. 2003;25(6):621–5. https://doi.org/10.1080/01421590310001605688. PMID: 15369910.
62. Feltovich PJ, Johnson PE, Moller JH, Swanson DB. The role and development of medical knowledge in diagnostic expertise. In: Clancey WJ, Shortliffe EH, editors. Readings in medical artificial intelligence: the first decade. Reading: Addison Wesley; 1984. p. 275–319.
63. Patel VL, Evans DA, Kaufman DR. Cognitive framework for doctor-patient interaction. In: Evans DA, Patel VL, editors. Cognitive science in medicine: biomedical modeling. Cambridge, MA: The MIT Press; 1989. p. 253–308.
64. Glaser R, Chi MTH, Farr MJ. The nature of expertise. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
65. Lesgold A, Rubinson H, Feltovich P, Glaser R, Klopfer D, Wang Y. Expertise in a complex skill: diagnosing x-ray pictures. In: Chi MTH, Glaser R, Farr MJ, editors. The nature of expertise. Hillsdale: Lawrence Erlbaum Associates; 1988. p. 311–42.
66. Evans DA, Gadd CS. Managing coherence and context in medical problem-solving discourse. In: Evans DA, Patel VL, editors. Cognitive science in medicine: biomedical modeling. Cambridge, MA: MIT Press; 1989. p. 211–55.
67. Miller GA. The magic number seven plus or minus two: Some limits on our capacity for processing information. Psychological review. 1956;63:91–97.
68. Hutchins E. Cognition in the wild. Cambridge, MA: MIT Press; 1995.
69. Hutchins E. How a cockpit remembers its speeds. Cogn Sci. 1995;19:265–88.
70. Patel VL, editor. Distributed and collaborative cognition in health care: Implications for systems development. Special issue of Artificial Intelligence in Medicine. 1998;12(2).
71. Cohen T, Blatter B, Almeida C, Shortliffe E, Patel V. A cognitive blueprint of collaboration in context: distributed cognition in the psychiatric emergency department. Artif Intell Med. 2006;37:73–83.
72. Hazlehurst B, McMullen CK, Gorman PN. Distributed cognition in the heart room: how situation awareness arises from coordinated communications during cardiac surgery. J Biomed Inform. 2007;40:539–51.
73. Nemeth C, O'Connor M, Cook R, Wears R, Perry S. Crafting information technology solutions, not experiments, for the emergency department. Acad Emerg Med. 2004;11(11):1114–7. https://doi.org/10.1197/j.aem.2004.08.011.
74. Kannampallil TG, Franklin A, Mishra R, Almoosa KF, Cohen T, Patel VL. Understanding the nature of information seeking behavior in critical care: implications for the design of health information technology. Artif Intell Med. 2013;57(1):21–9. https://doi.org/10.1016/j.artmed.2012.10.002.
75. Fukushima K, Miyake S. Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition. In: Competition and cooperation in neural nets. Berlin/Heidelberg: Springer; 1982. p. 267–85.
76. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. 1998;86(11):2278–324.
77. Marcus G. Deep learning: A critical appraisal. arXiv preprint arXiv:1801.00631. 2018.
78. Coiera E. The cognitive health system. Lancet. 2020;395(10222):463–66. https://doi.org/10.1016/S0140-6736(19)32987-3. Epub 2020 Jan 7. PMID: 31924402.
79. Kuang C. Can A.I. be taught to explain itself? The New York Times Magazine, Feature article. 2017. https://www.nytimes.com/2017/11/21/magazine/can-ai-be-taught-to-explain-itself.html.
80. Sharda P, Das AK, Cohen TA, Patel V. Customizing clinical narratives for the electronic medical record interface using cognitive methods. Int J Med Inform. 2006;75:346–68.
81. Spitzer RL, Gibbon ME, Skodol AE, Williams JBW. DSM-IV casebook: a learning companion to the diagnostic and statistical manual of mental disorders. 4th ed. American Psychiatric Association; 1994.
82. Cohen T. Augmenting expertise: toward computer-enhanced clinical comprehension. PhD dissertation. Columbia University; 2007.

83. Cohen T, Blatter B, Patel V. Simulating expert clinical comprehension: adapting latent semantic analysis to accurately extract clinical concepts from psychiatric narrative. J Biomed Inform. 2008;41(6):1070–87. https://doi.org/10.1016/j.jbi.2008.03.008.
84. Dalai VV, Khalid S, Gottipati D, Kannampallil T, John V, Blatter B, Patel VL, Cohen T. Evaluating the effects of cognitive support on psychiatric clinical comprehension. Artif Intell Med. 2014;62(2):91–104. https://doi.org/10.1016/j.artmed.2014.08.002.
85. Schraagen JM, Chipman SF, Shalin VL. Cognitive task analysis. Psychology Press; 2000.
86. Baxter GD, Monk AF, Tan K, Dear PRF, Newell SJ. Using cognitive task analysis to facilitate the integration of decision support systems into the neonatal intensive care unit. Artif Intell Med. 2005;35(3):243–57. https://doi.org/10.1016/j.artmed.2005.01.004.
87. Kieras DE, Bovair S. The role of a mental model in learning to operate a device. Cogn Sci. 1984;8(3):255–73. https://doi.org/10.1016/S0364-0213(84)80003-8.
88. Bansal G, Nushi B, Kamar E, Lasecki WS, Weld DS, Horvitz E. Beyond accuracy: the role of mental models in human-AI team performance. Proc AAAI Conf Hum Comput Crowdsourc. 2019;7:2–11.
89. Bansal G, Nushi B, Kamar E, Weld DS, Lasecki WS, Horvitz E. Updates in human-AI teams: understanding and addressing the performance/compatibility tradeoff. Proc AAAI Conf Artif Intell. 2019;33(01):2429–37. https://doi.org/10.1609/aaai.v33i01.33012429.
90. Gero KI, Ashktorab Z, Dugan C, Pan Q, Johnson J, Geyer W, Ruiz M, Miller S, Millen DR, Campbell M, Kumaravel S, Zhang W. Mental models of AI agents in a cooperative game setting. In: Proceedings of the 2020 CHI conference on human factors in computing systems. Association for Computing Machinery; 2020. p. 1–12. https://doi.org/10.1145/3313831.3376316.
91. Kulesza T, Stumpf S, Burnett M, Kwan I. Tell me more? The effects of mental model soundness on personalizing an intelligent agent. In: Proceedings of the SIGCHI conference on human factors in computing systems; 2012. p. 1–10. https://doi.org/10.1145/2207676.2207678.
92. Patel VL, Shortliffe EH, Stefanelli M, Szolovits P, Berthold MR, Bellazzi R, Abu-Hanna A. The coming of age of artificial intelligence in medicine. Artificial Intelligence in Medicine. 2009;46:5–17.