AI in Medicine Comes of Age: Blossoming or Languishing?

Edward H. Shortliffe, MD, PhD

Professor of Biomedical Informatics
Arizona State University

Professor of Basic Medical Sciences and of Medicine
University of Arizona College of Medicine

Phoenix, AZ USA

ICAART 2009

International Conference on Agents and Artificial Intelligence

Porto, Portugal

January 19, 2009

Historical Perspective

- Earliest broad recognition of statistical issues in diagnosis and the potential role of computers occurred in the late 1950s
  - "Reasoning foundations in medical diagnosis": Classic article by Ledley and Lusted appeared in Science in 1959
- Computers began to be applied in biomedicine in the 1960s
  - Most applications dealt with clinical issues
- Bayesian diagnosis systems and statistical pattern recognition

1960's

- By the late 1960s it became apparent that statistical approaches had limitations and bore little resemblance to the way in which expert clinicians solved difficult medical problems
- Early AI work in non-medical domains
  - Production rules (Newell and Simon)
  - General problem solving systems
  - Theory formation and early machine learning
- AI in organic chemistry
  - Dendral modeled the expertise of organic chemists who could interpret mass spectroscopy of unknown compounds
  - Production rules derived from N&S approach

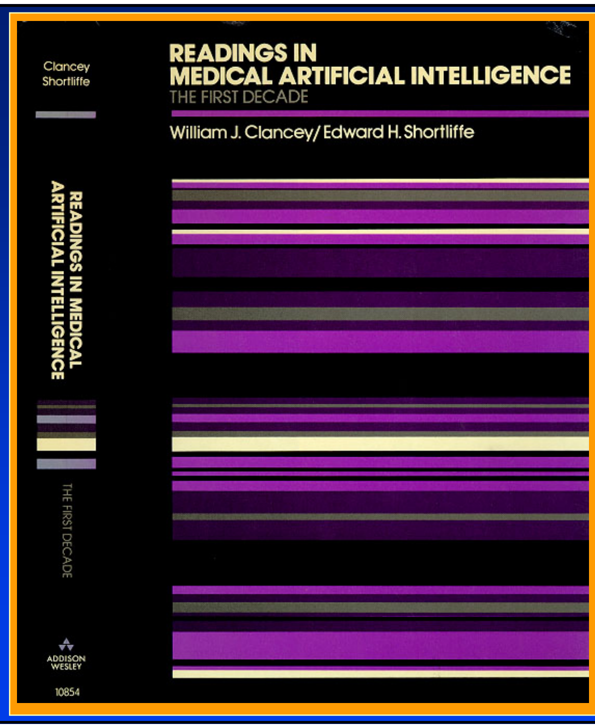
1970's

- Applications of flowcharting, logical diagrams, and complex algorithms
- Decision-analysis programs
- Mathematical modeling
- Early clinical work on modeling expertise using AI methods
  - Internist-1: hypothesis-directed reasoning
  - MYCIN: goal-directed reasoning using rules
  - CASNET: causal association networks

Readings in Medical Artificial Intelligence: The First Decade

Reading, MA: Addison-Wesley

1984



Computer-Assisted Decision Support

Examples of functionalities

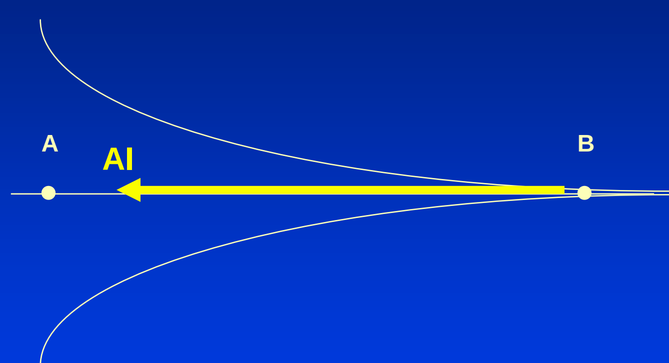
- Generic information access tools (e.g., Medline)
- Patient-specific consultation systems
  - Diagnosis, workup, therapy or patient management
  - Critiques: reactions to users' impressions or plans
- Browsing tools that mix generic and patient-specific elements (e.g., "electronic textbooks of medicine")
- Monitoring tools that generate warnings or advice as needed (advice as a byproduct of patient care and data recording)

Proactive Computer-Assisted Decision Support

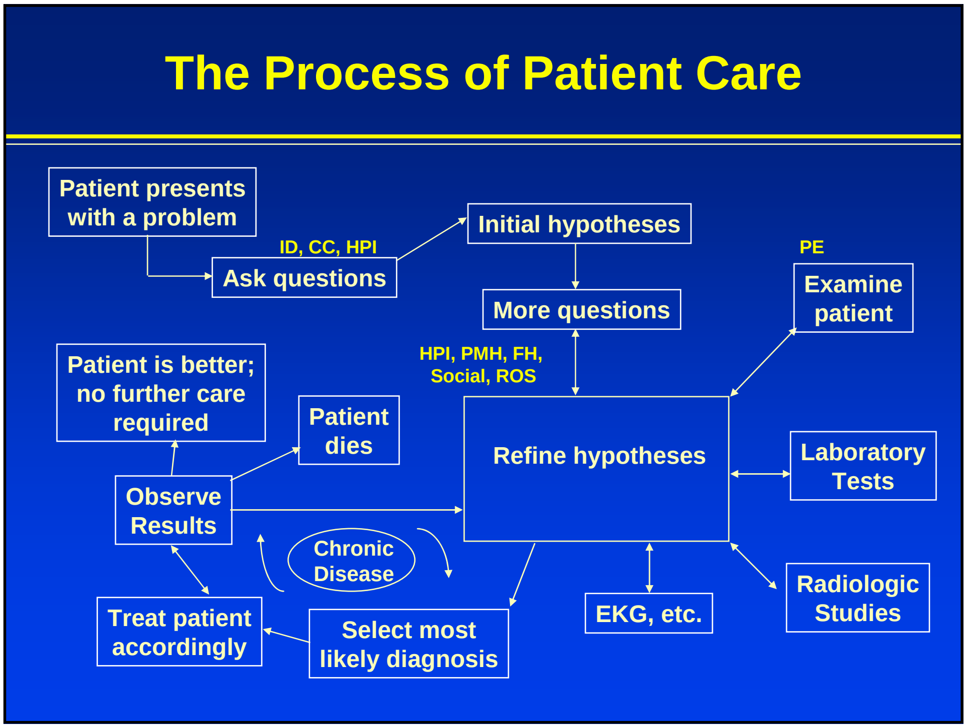
Examples of available methodologies:

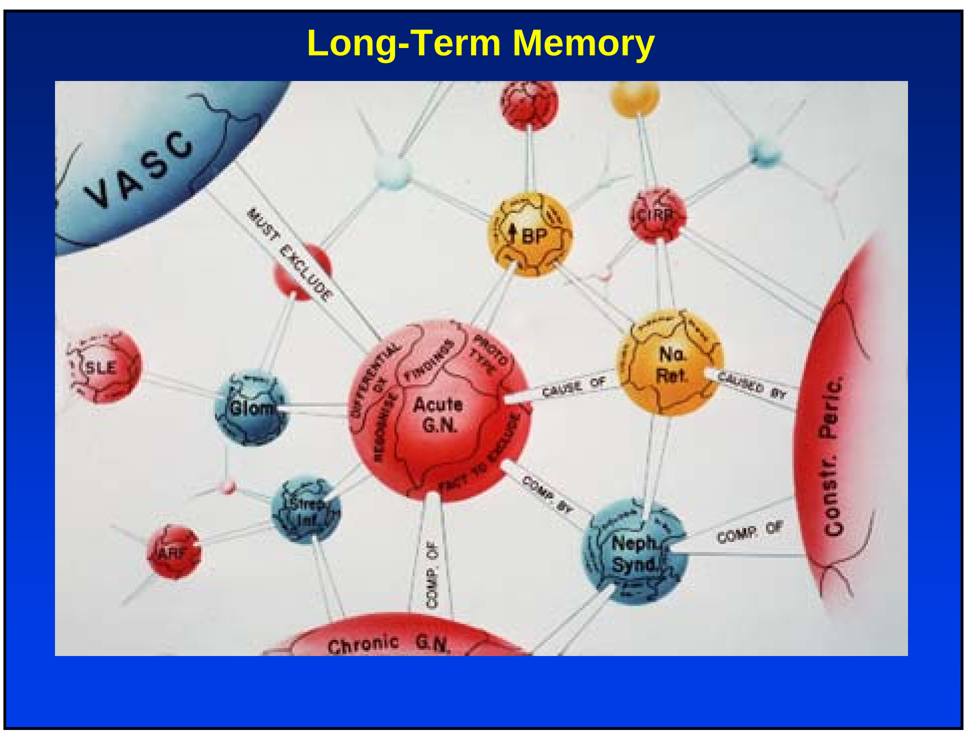
- Protocols and algorithms ("clinical guidelines")
- Clinical databanks
- Mathematical models (often physiologic)
- Statistical pattern recognition and neural networks
- Bayesian statistics and Bayesian networks
- Decision analysis
- Artificial intelligence ("expert systems")
- Syntheses of various techniques

"Blois Funnel"

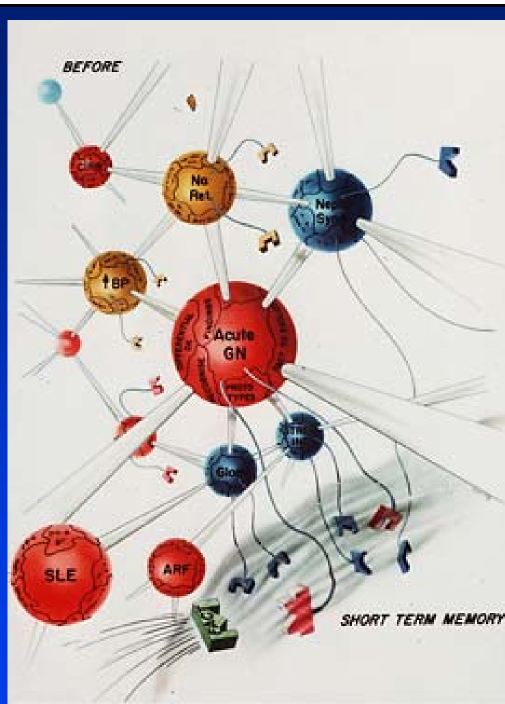


The cognitive span required during diagnosis


The Process of Patient Care
Patient presents with a problem
ID, CC, HPI
Ask questions
Initial hypotheses
More questions
HPI, PMH, FH, Social, ROS
PE
Examine patient
Patient is better; no further care required
Patient dies
Refine hypotheses
Laboratory Tests
Observe Results
Chronic Disease
Treat patient accordingly
Select most likely diagnosis
EKG, etc.
Radiologic Studies

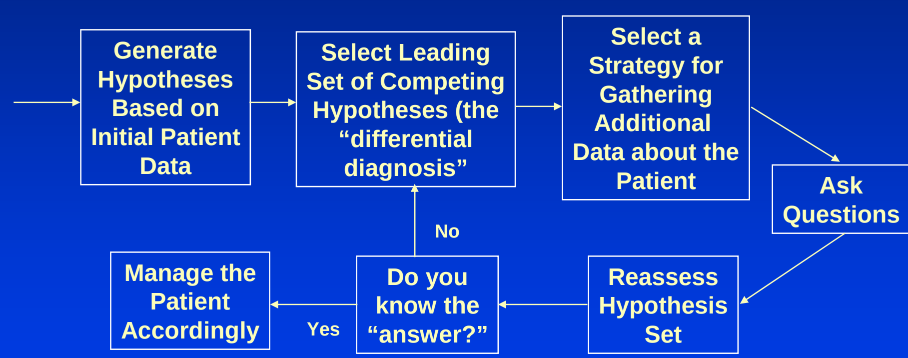

Long-Term Memory
VASC
MUST EXCLUDE
BP
CIRR
SLE
Glom
DIFFERENTIAL DX
FINDINGS
PROTO TYPE
Acute G.N.
CAUSE OF
Na. Ret.
CAUSED BY
Constr. Peric.
COMP. BY
Strep. Inf.
ARF
COMP. OF
Neph. Synd.
COMP. OF
Chronic G.N.

Short-Term Memory



Hypothesis Evocation

The Hypothetico-Deductive Approach



Internist-1 / QMR

Task: Diagnosis in internal medicine and neurology

- Scope: The entire field!
- Began in early 1970s as collaboration between Dr. Jack Myers (physician) and Prof. Harry Pople (computer scientist) at University of Pittsburgh
- Dr. Randy Miller (physician) worked on project throughout the 1970s and became project leader in 1980s
- Internist-1 was mainframe (Lisp) version of program, used to develop methods and extensive clinical knowledge base
- QMR was PC version developed by Dr. Miller and collaborators during 1980s and made commercially available

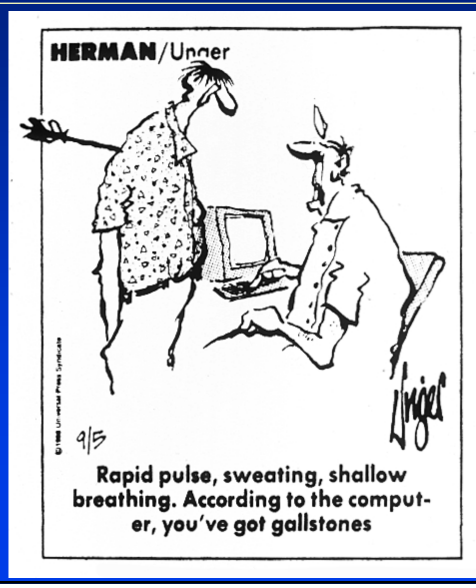
Internist-1 → QMR

Myths Regarding Decision-Support Systems

Myth:

Diagnosis is the dominant decision-making issue in medicine

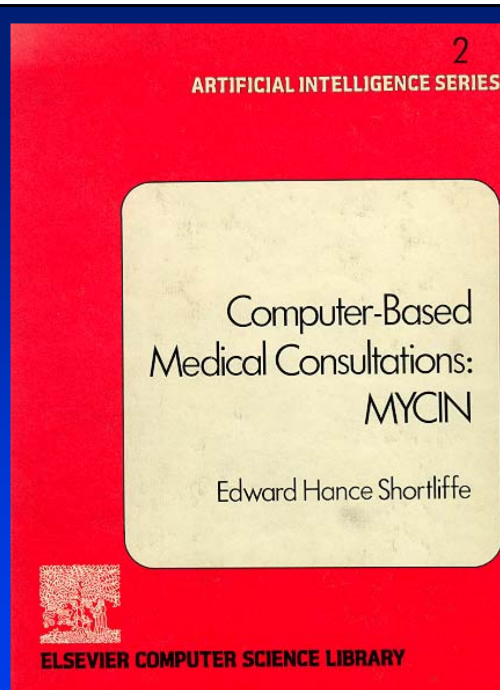
Limitations of Computer-Based Diagnosis: How Much Can The Computer Know?



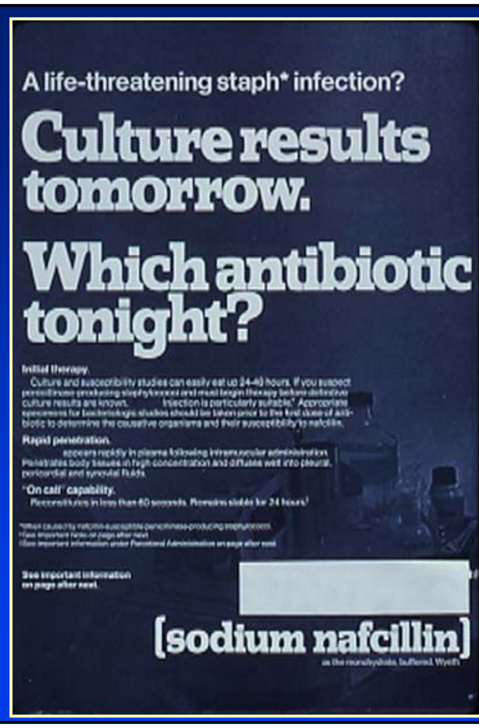
Computer-Based Medical Consultations: MYCIN

New York: American Elsevier

1976



The Clinical Problem



Evaluating Performance

- Need for a gold standard
  - What would the "correct" decision have been?
  - Lack of agreement among experts
  - Challenge of assessing patients without seeing them
- Reasonable time demands for evaluators
- Blinded study design
- Control for regional variability
- Define reasonable performance goals

The Ten Patients

Randomly selected cases of meningitis, screened only to insure that at least one bacterial, one viral, one fungal, and one tuberculosis case was included:

- Case 1: Viral
- Case 2: Viral
- Case 3: Group B Streptococcus
- Case 4: Neisseria Meningitidis
- Case 5: Staphylococcus Aureus
- Case 6: Listeria Monocytogenes
- Case 7: Streptococcus Pneumoniae
- Case 8: Listeria Monocytogenes
- Case 9: Mycobacterium Tuberculosis
- Case 10: Cryptococcus Neoformans

The Ten Prescribers

Ten prescribers reviewed the histories of the 10 patients with meningitis, and for each case they indicated their preferred therapy, based on the available information.

Prescribers 1-5: Infectious disease experts on the faculty at Stanford Medical School

Prescriber 6: A fellow in infectious diseases at Stanford

Prescribers 7 and 8: A medical resident and medical student on the Stanford infectious disease rotation

Prescriber 9: MYCIN

Prescriber 10: The actual therapy given by the ward team (often after infectious disease consultation)

The Eight Evaluators

Eight infectious disease experts from academic centers other than Stanford participated in the study. Each had published papers on meningitis diagnosis and treatment.

For each of the 10 cases, the evaluator was asked to indicate how the patient should be managed. Each evaluator was also asked to review the 10 therapies suggested by the prescribers and to classify each of the 10 as either acceptable or unacceptable treatment.

Summary

Eight evaluators rated each of the 10 therapies acceptable or unacceptable for each of the 10 patients —> total of 100 judgments per evaluator

Ten prescribers were rated by 8 evaluators regarding their therapy for each of 10 patients —> total of 80 judgments per prescriber

Thus, each prescriber could have as many as 80 “acceptable” ratings by the evaluators

Evaluation of Ten Prescribers by Eight Meningitis Experts (N=80)

Prescriber	Score (Max=80)	Percent Acceptable
MYCIN	55	69
Faculty-5	54	68
Fellow	53	66
Faculty-3	51	64
Faculty-2	49	61
Faculty-4	47	59
Actual Treatment	47	59
Faculty-1	45	56
Resident	39	49
Student	28	35

No. of Cases in Which Therapy Failed to Cover a Treatable Pathogen (N=10)

Prescriber	Number
MYCIN	0
Faculty-5	1
Fellow	1
Faculty-3	1
Faculty-2	0
Faculty-4	0
Actual Treatment	0
Faculty-1	0
Resident	1
Student	3

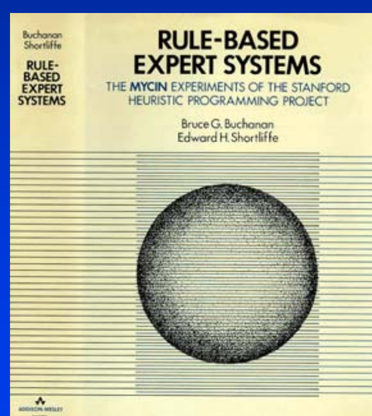
Evaluation of Each of Eight Experts By The Other Seven

Evaluator	Percent Acceptable
8	81
1	76
5	74
3	71
6	69
4	62
2	57
7	56

Published Evaluation

Yu VL, et al. Antimicrobial selection by a computer - a blinded evaluation by infectious disease experts. Journal of the American Medical Association 242:1279-1282, 1979.

Buchanan BG, Shortliffe, EH. Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. Reading, MA: Addison-Wesley 1984



http://www.u.arizona.edu/~shortlif

Myths Regarding Decision-Support Systems

Myth:

Clinicians will use knowledge-based systems if the programs can be shown to function at the level of experts

Challenges in Biomedical Settings

Principal barriers lie less with the technology and more with socio-cultural factors that prevent effective implementation of what is already possible:

- Culture
- Integration (e.g., time-sharing vs networked systems that communicate and share data)
- Workflow
- Privacy
- Finance
- Evaluation / Verification
- Safety / Liability

1980's

- "Overselling" of artificial intelligence
- Resurgence of interest in Bayesian approaches
  - Belief networks and influence diagrams
- Neural networks
- Major changes due to new hardware and software technologies
  - Macs and PCs: viable delivery model
  - Graphical interfaces: rethinking the nature of user interactions with computers
  - Networking: new options for integrating advice systems with their environment
- "Greek oracle" model fell into disfavor

Myths Regarding Decision-Support Systems

Myth:

Clinicians will use stand-alone decision-support tools

AMIE 1991 - Maastricht

Artificial Intelligence in Medicine 5 (1993) 93–106
Elsevier

93

ARTMED 199

The adolescence of AI in Medicine: Will the field come of age in the '90s?*

Edward H. Shortliffe

Section on Medical Informatics, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, CA 94305, USA

Received September 1992
Revised November 1992

*** This paper is based on a keynote address presented by the author at the *Artificial Intelligence in Medicine - Europe (AIME)* meeting, Maastricht, The Netherlands, June 24- 27, 1991.**

Key Points of "Adolescence" Article

- AIM is not a field that can be set off from the rest of biomedical informatics or the world of health planning and policy
- Impact is dependent on the development of integrated environments that allow merging of knowledge-based tools with other applications
- We need vision and resources from leaders who understand that medical practice is inherently an information-management task

Three Challenges Identified

- Need to develop national and international biomedical-networking infrastructures for communication, data exchange, and information retrieval
- Need for credible international standards for communications, data, and knowledge exchange
- Education: We need more professionals who are broadly educated regarding the interdisciplinary nature of biomedical informatics (including AIM)

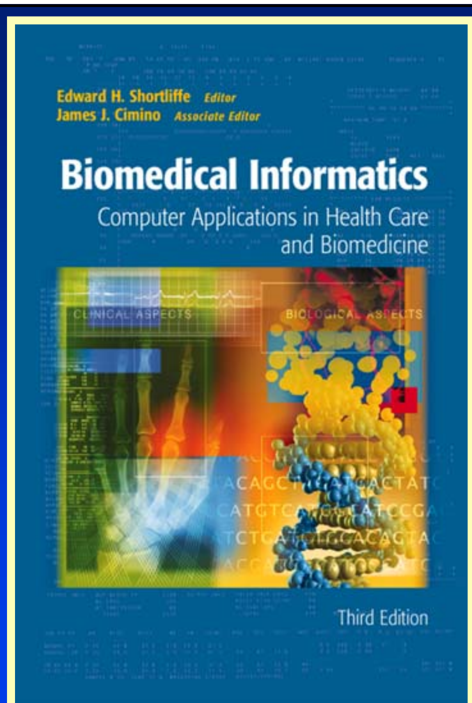
Biomedical Informatics Textbook

1st edition 1990
(Addison Wesley)

2nd edition 2000

3rd edition 2006

Springer New York



1990's

- Integration and networking become central issues
  - World Wide Web revolutionizes our thinking about distributed information access
- Knowledge-representation research matures
  - Ontology development and tools
  - Challenges of temporal representations and reasoning finally begin to yield to researchers
- Integration of decision-support features with databases arrives in some commercial products
- Standards emerge as a major issue
  - terminology, representation of decision logic, data models
  - crucial to promote sharing and collaboration

2000-2010

- Integration of decision support with workflow continues to be viewed as a central requirement
- Patient safety and error reduction become major motivators
- We see increasing incorporation of decision-support functionalities in commercial products
- New issues arise regarding relationships between vendors and hospital IT staffs, especially in the incorporation of decision-support and knowledge-management tools that are supported by the institution's clinical staff

Assessing the AIM Field

AIME 1991

- In what ways has the field advanced?
  - New topics/themes?
  - New methods?
  - Growth?
  - Recognition?
- In what ways, and to what extent, has the field had a direct influence on clinical medicine or other biomedical fields?
- How well is the field being supported?
  - Funding agencies?
  - Academic institutions?
  - Research organizations?
  - Colleagues?

Topics and Themes at AIME '07

- Clinical Data Mining
- Data and Knowledge Representation
- Computer-based Knowledge Generation
- Knowledge-Based Health Care
- Probabilistic and Bayesian Analysis
- Feature Selection / Reduction
- Visualization
- Classification and Filtering
- Information Retrieval
- Agent-Based Systems
- Temporal Data Mining
- Machine Learning
- Knowledge Discovery in Databases
- Text Processing
- Natural Language Processing
- Ontologies
- Decision Support Systems
- Image Processing
- Pattern Recognition
- Guidelines
- Workflow

Observations About AIM in 2007

- AI issues are ubiquitous in biomedicine, but most people do not call what they are doing AI
- When someone refers to their commercial product as an “AI System”, it probably isn’t
- AI research has diffused throughout biomedical informatics, and AIME and the AIM Journal stand out as the principal remaining forces for defining and recognizing AI in Medicine as a subfield in biomedical informatics and computer science
- Increased emphasis on guidelines and their encoding has been at the expense of classical patient-specific decision support for diagnosis and therapy planning

Observations About AIM in 2007

- Tremendous progress in several areas: knowledge representation (and associated tools), machine learning / knowledge discovery, temporal representation and reasoning, and others
- Slow progress in adoption of key standards for integration and knowledge sharing (controlled terminologies, semantic structuring of terminologies, standards for representing clinical decision logic – e.g., Arden, Proforma, GLIF, and others)
- Academic CS has begun to embrace biomedical applications as valid areas of emphasis for CS faculty (especially in bioinformatics)

AMIE 2007 - Amsterdam

* This paper is based on a panel presented by the authors at the *Artificial Intelligence in Medicine - Europe (AIME)* meeting, Amsterdam, The Netherlands, June 11, 2007.

ELSEVIER

ARTIFICIAL INTELLIGENCE IN MEDICINE

http://www.intl.elsevierhealth.com/journals/aiim

POSITION PAPER

The coming of age of artificial intelligence in medicine☆

Vimla L. Patel[a,*], Edward H. Shortliffe[a,b], Mario Stefanelli[c], Peter Szolovits[d], Michael R. Berthold[e], Riccardo Bellazzi[c], Ameen Abu-Hanna[f]

[a] Department of Biomedical Informatics, Arizona State University, ABC1, 425 North Fifth Street, Phoenix, AZ 85004, USA
[b] Department of Basic Medical Sciences, University of Arizona College of Medicine, ABC1, Room 227, 425 North Fifth Street, Phoenix, AZ 85004-2157, USA
[c] Dipartmento di Informatica e Sistemistica, University of Pavia, Via Ferrata 1, 27100 Pavia, Italy
[d] Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, The Stata Center, 32-254, 32 Vassar Street, Cambridge, MA 02139, USA
[e] Department of Computer and Information Science, University of Konstanz, BOX M712, 78457 Konstanz, Germany
[f] Department of Medical Informatics, Academic Medical Centre, University of Amsterdam, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands

Conclusions: Decision Support

- Integration with routine workflow is the key
- Transparency helps to assure acceptance
- The Web is a great facilitator of integration
  - Does not avoid the need for standardized terminologies and data-sharing protocols
- Implementation of vendor-supplied clinical information systems can present new challenges when attempting to integrate locally-produced decision-support functionalities