# A Virtual Archive
# for the History of AI

*Bruce G. Buchanan, Joshua Eckroth, Reid G. Smith*

■ *Publications that have influenced the growth of artificial intelligence are often difficult to obtain. We first collected titles of several thousand publications from many well-known sources and then selected about 2,000 titles considered to be especially influential. We have identified, and in a few cases created, online versions of about half of these "classics in AI." Searchable text of the documents enables additional analysis of trends and influences. Integration into the rest of the AITopics information portal contextualizes the classic publications.*

As a scientific society, AAAI has a stated responsibility to disseminate accurate information about AI and promote accurate views of the field.[1] Because historians will construct their stories from the primary source material they can find, it follows that AAAI has a responsibility to help preserve the publications that have defined the field. For this reason, the strategic plan for AAAI (AAAI 2008) includes the clause, "Preserve the history of the field, including such tasks as archiving AI materials."

The AITopics website[2] was set up in part to meet these responsibilities (Buchanan and Glick 2001). In addition to being a digital archive, the website also aims to be a broad source of introductory readings, news, podcasts, and videos covering the wide range of topics in our field. All of the linked material has been previously published in reputable outlets and is reviewed by editors. This ensures the integrity of the information that is disseminated. However, we found that much of the older material that helped define the field was previously difficult or impossible to find in digital form. Even in hard copy, some materials could be found only in a few research libraries and personal collections.

# One Hundred Thirty Influential Publications in the History of AI Published Through 1950, the Year of Alan Turing's Paper "Computing Machinery and Intelligence"

The history of AI is a complex intellectual fabric of intertwined threads. The landmark publications listed in the shaded boxes on the four pages that follow represent many of the major threads and conceptual advances that professionals in AI regard as significant. Many others are listed under the "Classics" link on AITopics.

Philosophers writing up to the time of Charles Babbage and Ada Lovelace used the concept of mechanized thought as a metaphor to help clarify the nature of the human mind and of rational thought. In this period there was also important work on probability theory, but serious work on post-Aristotelian logic did not start until the mid-19th century. After Gottlob Frege's 1879 publication creating mathematical logic as we have come to know it, logicians were able to define rational thought within formal systems that lend themselves to mechanization. With the development of mechanical devices of all sorts in the 18th century, the metaphor of autonomous machines became the subject matter for Mary Shelly's *Frankenstein* and Karel Čapek's *R.U.R.*

By the early 20th cntury, psychologists were collecting experimental data on human learning and thinking, and by the 1940s biologists were collecting data on the neurological basis of thought. By 1950, scientists from many disciplines, along with mathematicians, linguists, and philosophers, were using the newly developed digital computers to encode and test new theories. In that year Alan Turing brought a new focus to the disparate work when he boldly stated that machines could think.

Much of the historical record of AI is found in text hard copy, videos, and to a small extent, audio tapes and photographs. Demos, interviews, and lectures on videotape supplement the written record. AAAI is not set up to undertake the expensive, long-term commitment to be an archive of the physical materials themselves. Digitizing may play a part in our efforts, and it is worthwhile in itself (especially for pedagogical purposes), but digitizing is not necessarily archiving. There is a need to coordinate with institutional archives that will preserve material in proper environmental conditions, in perpetuity, which is an expensive, long-term commitment. However, the AAAI can maintain a virtual archive so historians can locate original materials easily and researchers can read original materials without traveling to a major library or archive.

Many archivists at university libraries and historical institutions are already preserving the hard-copy publications of individual scientists whose work has been influential in the development of AI. For example, AAAI is shipping hard copies of AAAI Press publications (printed on acid-free paper), including conference, symposia, and workshop proceedings, to the Stanford Archive for long-term preservation. Stanford is also preserving papers from John McCarthy, Edward Feigenbaum, Nils Nilsson, and others. Carnegie Mellon University has an extensive collection of papers in the Simon-Newell collection. The Computer History Museum has papers and artifacts from the history of AI, and the Babbage Institute has many important papers.

AAAI has taken on the role of collecting pointers to digital copies of influential publications, videos (Buchanan, Glick, and Smith 2008) and other material that is physically archived at many different locations. That is, the association is creating a virtual archive for the history of AI.

## Methods

We began linking to online information about AI several years ago from an earlier version of AAAI's information portal, AITopics (Buchanan and Glick 2002). When it became clear that many articles and news stories (Dong, Smith, and Buchanan 2011; Eckroth et al. 2012) pertained to the history of AI, we established a separate section on AITopics for history, which has become one of the more frequently viewed areas of the site.

Many of the links that we found for the history section were pointers to papers that were influential in shaping the field. Several of these had been digitized and stored in diverse online repositories. We also found that some out-of-print books had been digitized and were available online. In short order, an abundance of evidence suggested that the

| Authors | Title | Year |
|---|---|---|
| Bacon, F. | *The New Organon or True Directions Concerning the Interpretation of Nature* | 1620 |
| Descartes, R. | *Discourse on the Method of Rightly Conducting the Reason, and Seeking the Truth in the Sciences* | 1637 |
| Descartes, R. | *Meditations on First Philosophy* | 1637 |
| Hobbes, T. | *The Leviathan* | 1651 |
| Arnauld, A. | *La Logique, ou L'Art De Penser* | 1662 |
| Cardano, G. | *Liber de Ludo Aleae* | 1663 |
| Locke, J. | *An Essay Concerning Human Understanding* | 1690 |
| Arbuthnot, J. | *Of the Laws of Chance* | 1692 |
| Bernoulli, D. | *Specimen Theoriae Novae de Mensura Sortis* | 1738 |
| La Mettrie, J. O. | *L'Homme Machine* | 1748 |
| Bayes, T. | *An Essay Towards Solving a Problem in the Doctrine of Chances* | 1763 |
| Laplace, P. | *Essai Philosophique sur les Probabilités* | 1816 |
| Hoffmann, E. T. A. | *The Sandman* | 1817 |
| Shelley, M. | *Frankenstein: or, the Modern Prometheus* | 1818 |
| Poe, E. A. | *Maelzel's Chess-Player* | 1836 |
| Cournot, A. (Ed.) | *Recherches sur les Principes Mathématiques de la Théorie des Richesses* | 1838 |
| Lovelace, [Countess] A. A. | *Translator's Notes to an Article on Babbage's Analytical Engine* | 1842 |
| Mill, J. S. | *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence, and Methods of Scientific Investigation* | 1843 |
| Boole, G. | *On a General Method in Analysis* | 1844 |
| Boole, G. | *The Mathematical Analysis of Logic: Being an Essay Towards a Calculus of Deductive Reasoning* | 1847 |
| Boole, G. | *An Investigation of the Laws of Thought on Which Are Founded the Mathematical Theories of Logic and Probabilities* | 1854 |
| Grassmann, H. | *Lehrbuch der Arithmetik* | 1861 |
| Mill, J. S. | *Utilitarianism* | 1863 |
| De Morgan, A. | *On the Syllogism, No. IV, and on the Logic of Relations* | 1864 |
| Butler, S. | *Erewhon* | 1865 |
| Peirce, C. S. | *Description of a Notation for the Logic of Relatives, Resulting from an Amplification of the Conceptions of Boole's Calculus of Logic* | 1870 |
| Jevons, W. S. | *The Principles of Science* | 1874 |
| Gauss, C. F. | *Beiträge zur Theorie der Algebraischen Gleichungen* | 1876 |
| Schröder, E. | *Der Operationskreis des Logikkalküls* | 1877 |
| Johnson, W. W., and Story, W. E. | *Notes on the "15" Puzzle* | 1879 |
| Frege, G. | *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens* | 1879 |
| Tait, P. G. | *Note on the Theory of the 15 Puzzle* | 1880 |
| Peirce, C. S. | *A Theory of Probable Inference* | 1883 |
| Peirce, C. S. | *Logical Machines* | 1887 |
| Peano, G. | *Arithmetices Principia, Nova Methodo Exposita* | 1889 |
| Wells, H. G. | *The War of the Worlds* | 1898 |

*Influential Publications in the History of AI Published 1620 Through 1898.*

vast majority of important works from the history of AI could be found on the web (although many publishers charge a fee for full text access). What was missing, however, was a central and organized virtual archive that provided pointers to the myriad sources.

We began the process of identifying the important historical documents by collecting bibliographies from several well-known authors. To the extensive bibliography that Marvin Minsky assembled when the field was young (Minsky 1961) we added long lists of publications from the bibli-

ographies of more than a dozen publications shown in the Influential Bibliographies section at the end of this article. We used our own judgment to select the most relevant references based on subjective criteria of influence and impact.

From the many thousands of publications residing in the union of these lists, we removed all that were published after 1993 in order to maintain a 20-year perspective. We also removed publications that, in our judgment, were not part of "mainstream" AI although they may have had a large influence on the course of science, or of a particu-

| Authors | Title | Year |
|---|---|---|
| Peirce, C. S. | *Logic As Semiotic: The Theory of Signs* | 1902 |
| Couturat, L. Trans. By Rutherford, D. and Monroe, R. T. | *The Logic of Leibniz (Ca. 1679-1715)* | 1903 |
| Jennings, H. S. | *Behavior of the Lower Organisms* | 1906 |
| Whitehead, A. N. and Russell, B. | *Principia Mathematica* | 1910 |
| Tattersall, C. | *A Thousand End-Games: A Collection of Chess Positions That Can Be Won Or Drawn by the Best Play* | 1911 |
| Whitehead, A. N. | *An Introduction to Mathematics* | 1911 |
| Zermelo, E. | *An Application of Set Theory to the Theory of Chess-Playing* | 1913 |
| Markov, A. A. | *An Example of Statistical Investigation in the Text of "Eugene Onegin" Illustrating Coupling of "Tests" in Chains* | 1913 |
| Löwenheim, L. | *Über Möglichkeiten Im Relativ Kalkül* | 1915 |
| Leśniewski, S. | *Podstawy Ogólnej Teorii Mnogošci* | 1916 |
| Kohler, W. | *The Mentality of Apes* | 1917 |
| Čapek, K. | *R.U.R. (Rossum's Universal Robots)* | 1920 |
| Post, E. L. | *Introduction to a General Theory of Elementary Propositions* | 1921 |
| Wright, S. | *Correlation and Causation* | 1921 |
| Keynes, J. M. | *A Treatise on Probability* | 1921 |
| Fisher, R. A. | *On the Mathematical Foundations of Theoretical Statistics* | 1922 |
| Wittgenstein, L. | *Tractatus Logico-Philosophicus* | 1922 |
| Piaget, J. | *The Language and Thought of the Child* | 1926 |
| Skolem, T. | *Über Die Mathematische Logik* | 1928 |
| Carnap, R. | *Der Logische Aufbau Der Welt* | 1928 |
| Von Mises, R. | *Wahrscheinlichkeit, Statistik Und Wahrheit* | 1928 |
| Gödel, K. | *Über Die Vollständigkeit Des Logikkalkuls* | 1929 |
| Herbrand, J. | *Recherches Sur La Théorie De La Démonstration* | 1930 |
| Empson, W. | *Seven Types of Ambiguity* | 1930 |
| Ramsey, F. P. | *Truth and Probability* | 1931 |
| Gödel, K. | *On Formally Undecidable Propositions of Principia Mathematica and Related Systems I* | 1931 |
| Bartlett, F. C. | *Remembering: A Study in Experimental and Social Psychology* | 1932 |
| Thorndyke, E. | *The Fundamentals of Learning* | 1932 |
| Tarski, A. | *Pojęcie Prawdy W Językach Nauk Dedukcyjnych* | 1933 |
| Wright, S. | *The Method of Path Coefficients* | 1934 |
| Bliss, C. I. | *The Method of Probits* | 1934 |
| Tarski, A. | *Die Wahrheitsbegriff in Den Formalisierten Sprachen* | 1935 |
| Kleene, S. C. | *General Recursive Functions of Natural Numbers* | 1935 |
| Church, A., | *An Unsolvable Problem of Elementary Number Theory* | 1936 |
| Turing, A. | *On Computable Numbers, With An Application to the Entscheidungsproblem* | 1936 |
| Rashevsky, N. | *Physico-Mathematical Aspects of Excitation and Conduction in Nerves* | 1936 |
| Church, A. | *A Note on the Entscheidungsproblem* | 1936 |
| Wells, H. G. | *World Brain: The Idea of a Permanent World Encyclopedia* | 1937 |
| De Finetti, B. | *Le Prévision: Ses Lois Logiques, Ses Sources Subjectives* | 1937 |
| Carnap, R. | *The Logical Syntax of Language* | 1937 |
| Rashevsky, N. | *Mathematical Biophysics: Physico-Mathematical Foundations of Biology* | 1938 |
| Kantorovich, L. V. | *Mathematical Methods of Organizing and Planning Production* | 1939 |

*Influential Publications in the History of AI Published 1902 Through 1939.*

lar science. Copernicus's treatise on the heliocentric theory of the universe, for example, is an all-time "classic" that changed the course of science but has no "direct-line" descendents in contemporary AI. Darwin's *On the Origin of Species*, on the other hand, is another classic in the history of science that quite directly laid the foundation for investigations into evolutionary learning systems.

We frequently found that one or more early documents predated a book or major publication that defined new concepts, issues, or methods. Rather than include all of this work, we rather arbitrarily

| Authors | Title | Year |
|---|---|---|
| Russell, B. | *History of Western Philosophy* | 1940 |
| Ashby, W. R. | *Adaptiveness and Equilibrium* | 1940 |
| Leonard, H. S. and Goodman, N. | *The Calculus of Individuals and Its Uses* | 1940 |
| Church, A. | *Calculi of Lambda Conversion* | 1941 |
| Kolmogorov, A. N. | *Interpolation Und Extrapolation Von Stationaren Zufalligen Folgen* | 1941 |
| Tarski, A. | *Introduction to Logic and to the Methodology of Deductive Sciences* | 1941 |
| Asimov, I. | *Runaround* | 1942 |
| Wiener, N. | *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series* | 1942 |
| Craik, K. J. | *The Nature of Explanation* | 1943 |
| Rosenblueth, A., Wiener, N., and Bigelow, J. | *Behavior, Purpose, and Teleology* | 1943 |
| Pitts, W. | *A General Theory of Learning and Conditioning* | 1943 |
| Hull, C. L. | *Principles of Behavior* | 1943 |
| Mcculloch, W. S. and Pitts, W. | *A Logical Calculus of the Ideas Immanent in Nervous Activity* | 1943 |
| Post, E. | *Formal Reductions of the General Combinatorial Problem* | 1943 |
| Von Neumann, J. and Morgenstern, O. | *Theory of Games and Economic Behavior* | 1944 |
| Wertheimer, M. | *Productive Thinking* | 1945 |
| Von Neumann, J. | *First Draft of a Report on the Edvac* | 1945 |
| Ashby, W. R. | *The Physical Origin of Adaptation by Trial and Error* | 1945 |
| Wiener, N., and Rosenblueth, A. | *Conduction of Impulses in Cardiac Muscle* | 1946 |
| Cox, R. T. | *Probability, Frequency, and Reasonable Expectation* | 1946 |
| Rashevsky, N. | *The Neural Mechanism of Logical Thinking* | 1946 |
| Finney, D. J. | *Probit Analysis: A Statistical Treatment of the Sigmoid Response Curve* | 1947 |
| Simon, H. A. | *Administrative Behavior* | 1947 |
| Ashby, W. R. | *Principles of the Self-Organizing Dynamic System* | 1947 |
| Shannon, C. | *A Mathematical Theory of Communication* | 1948 |
| Turing, a. | *Intelligent Machinery* | 1948 |
| Jeffreys, H. | *Theory of Probability* | 1948 |
| Mccarthy, J. | *The Brain As a Computing Machine* | 1949 |
| Ryle, G. | *The Concept of Mind* | 1949 |
| Berkeley, E. C. | *Giant Brains, Or Machines That Think* | 1949 |
| Weaver, W. | *Translation* | 1949 |
| Hebb, D.O. | *The Organization of Behavior* | 1949 |
| Reichenbach, H. | *The Theory of Probability: An Inquiry Into the Logical and Mathematical Foundations of the Calculus of Probability* | 1949 |
| Dantzig, G. B. | *Programming of Interdependent Activities: Ii* | 1949 |
| Orwell, G. | *Nineteen Eight-Four* | 1949 |

*Influential Publications in the History of AI Published 1940 Through 1949.*

selected one, or a small number, to represent the influential ideas. We used the 18 main topics from the AITopics website topic hierarchy as a first approximation of the main themes in AI. It was subsequently expanded into nearly 180 topics and subtopics, making up a two-level hierarchy, which we employ as an operational taxonomy of the field. The hierarchy is shown in figure 1 and serves as the site map for AITopics.[3] As we collect and organize the documents and pointers that make up the virtual archive, we categorize each resource into one or more topics and subtopics.

At the time of the AAAI conference in August 2012, we had compiled a list of 1600 publications that seemed to have had a significant influence on AI research and practice. We solicited comments from conference attendees and asked the AAAI Fellows to send us links to their own publications that they, themselves, considered representative of their most influential work. Our collection subse-

| Authors | Title | Year |
|---|---|---|
| Shannon, C. E. | *A Chess-Playing Machine* | 1950 |
| Wiener, N. | *The Human Use of Human Beings* | 1950 |
| Shannon, C. E. | *Programming a Computer for Playing Chess* | 1950 |
| Davies, D. W. | *A Theory of Chess and Noughts and Crosses* | 1950 |
| Estes, W. X. | *Toward a Statistical Theory of Learning* | 1950 |
| Slater, E. | *Statistics for the Chess Computer and the Factor of Mobility* | 1950 |
| Piaget, J. | *The Psychology of Intelligence* | 1950 |
| Churchman, C. W., and Ackoff, R. L. | *Methods of Inquiry* | 1950 |
| Kolmogorov, A. N. | *Foundations of the Theory of Probability* | 1950 |
| Walter, W. G. | *An Imitation of Life* | 1950 |
| Nash, J. | *Equilibrium Points in N-Person Games* | 1950 |
| Turing, A. M. | *Computing Machinery and Intelligence* | 1950 |
| Thorpe, W. H. | *The Concepts of Learning and Their Relation to Those of Instinct* | 1950 |
| Gibson, J. J. | *The Perception of the Visual World* | 1950 |
| Carnap, R. | *Logical Foundations of Probability* | 1950 |

*Influential Publications in the History of AI Published in 1950.*

quently has grown to about 2,000 publications.

Next, we began correlating publications in our list with online full-text resources. Nilsson's recent book (Nilsson 2010) was helpful as it contains many pointers to online versions. Several other links came from AAAI Fellows who pointed to their home pages. Some of the classic papers are included in AAAI or IJCAI conference proceedings, which are online. However, the majority of the 1800 publications were linked to online resources by way of a custom automated web crawler. We automated Google Scholar[4] searches for authors and titles and gathered the resulting links to ScienceDirect, Springer, ACM Digital Library, JSTOR, and other repositories. From this automated effort and the manual collection of links, a full 90 percent of the publications have pointers to an online resource. About half the publications also have pointers to full-text PDFs, although not all of these are freely available. We hope to increase the count of full-text pointers, but most digitized articles cannot be accessed without cost.

Some older (pre-1970) publications have never been digitized. Because they hold special distinction in our history, we have begun digitizing several out-of-print collections, shown in table 1, and have requested permission from publishers to digitize several more.

## Implementation

Resource limitations dictated using free, off-the-shelf software. The website is built with Drupal,[5] an open-source content management system. Each publication is stored as a distinct record with asso-

Blake, D. V., and Uttley, A. M., eds. 1959. *Mechanisation of Thought Processes 1–2.* London: Her Majesty's Stationary Office.

*Proceedings of the Machine Intelligence Workshops,* 1–14.

Anthony, R. N., ed. 1956. *Proceedings Automatic Data Processing Conference.* Boston: Alpine Press. Conference held at Harvard University, Sept. 8–9, 1955.

Webber, B., and Nilsson, N. 1981. *Readings in Artificial Intelligence.* Palo Alto, CA: Tioga Press.

*Table 1. Collections and Papers Digitized Specifically for the AITopics Virtual Archive.*

ciated metadata such as authors, title, publication year and venue, abstract, and in some cases, a representative image such as the cover of a book. Additionally, each publication is tagged with one or more topics and subtopics from the taxonomy shown in figure 1

Using a variety of custom software,[6] we were able to obtain full-text digital versions of 50 percent of the collection. About half of these digital versions were scans of the original hard copies. They have no text representation, and therefore required optical character recognition (OCR) in order to extract the text. Except for the publications we had scanned and OCR'd professionally, OCR was performed using an open-source tool, Tesseract.[7] Although the Tesseract software did not

**AI Overview**

Computer Science • Goals and Visions • Grand Challenge Problems • Roadblocks • The Future

**Applications**

Agriculture and Natural Resources • Archaeology • Architecture and Design • Art • Artificial Noses • Assistive Technologies • Astronomy and Space Exploration • Automatic Programming • Automotive Industry • Autonomous Vehicles • Aviation • Banking and Finance • Bioinformatics • Biometrics • Business and Manufacturing • Chatbots • Chemistry • Decision Support Systems • Earth and Atmospheric Science • Engineering Design • Fraud Detection • Hazards and Disasters • Knowledge Management • Law • Law Enforcement and Public Safety • Machine Storytelling • Marketing • Medicine • Military • Music • Networks • Oil and Gas • Politics and Foreign Relations • Recommender Systems • Robots in the Home • Robots in the Workplace • Science and Mathematics • Smart Houses and Appliances • Social Science • Software Engineering • Spam Filtering • Surveillance • Telecommunications • Transportation and Shipping

**Cognitive Science**

Childhood Development • Cognitive Architectures • Creativity and Intelligence • Emotion • Nature of Expertise • Nature of Intelligence • Neuroscience • Physiology of Perception and Motion • Problem Solving • Simulation

**Education**

Automated Grading • Instructional Theory • Intelligent Tutoring

**Ethics and Social Issues**

Robot Ethics • Security and Privacy • Social Issues

**Games and Puzzles**

15 Puzzle Arimaa • Backgammon • Bridge • Checkers • Chess • Chinese Checkers • Crossword Puzzles • Dolls and Toys • General Game Playing • Go Hearts • Hex • Jeopardy! • Othello • Poker • RoShamBo (rock, paper, scissors) • Rubik's Cube • Scrabble • Spades • Sports • Sudoku • Video Games

**History**

Brief Histories and Timelines • Histories of Related Topics • Influential Software and Systems • People, Programs, Institutions • Tributes

**Interfaces**

Human Computer Interaction • Human Robot Interaction • Interface Technologies and Devices

**Machine Learning**

Data Mining • Decision Tree Learning • Evolutionary Algorithms • Inductive Learning • Learning Theory • Memory-Based Learning • Neural Networks • Pattern Recognition • Reinforcement Learning • Scientific Discovery • Statistical Learning Methods

**Natural Language**

Discourse and Dialogue • Explanation and Argumentation • Grammars and Parsing • Information Retrieval • Machine Translation • Natural Language Generation • Natural Language Understanding • Question Answering • Text Classification • Text Summarization

**Philosophy**

Arguments Against AI • Computational Philosophy • Consciousness • Turing Test

**Representation and Reasoning**

Abductive Reasoning • Agents • Analogical Reasoning • Automated Theorem Proving • Bayesian Inference • Belief Revision • Case-Based Reasoning • Commonsense Reasoning • Constraint-Based Reasoning • Decision Trees • Description Logic • Design • Diagrammatic Reasoning • Diagrams and Models • Expert Systems • Fuzzy Logic • Heuristic Search • Logic and Formal Reasoning • Metareasoning • Multiagent Systems • Nonmonotonic Logic • Ontologies • Planning and Scheduling • Qualitative Reasoning • Rule-Based Reasoning • Scripts and Frames • Semantic Networks • Situation Calculus • Spatial Reasoning • Uncertainty

**Robots**

Biologically-Inspired Robots • Humanoid Robots • Manipulation and Locomotion • Robot Cognition • Robot Perception • Robot Planning and Action

**Science Fiction**

SciFi Books and Authors • SciFi Movies

**Speech**

Acoustic Processing • Speech Recognition • Speech Synthesis

**Systems and Languages**

Problem-Independent Architectures • Problem-Specific Architectures • Programming Languages

**Vision**

Face Recognition • Handwriting Recognition • Image Understanding • Optical Character Recognition

**Web and AI**

Semantic Web • Social Networks • Web Search

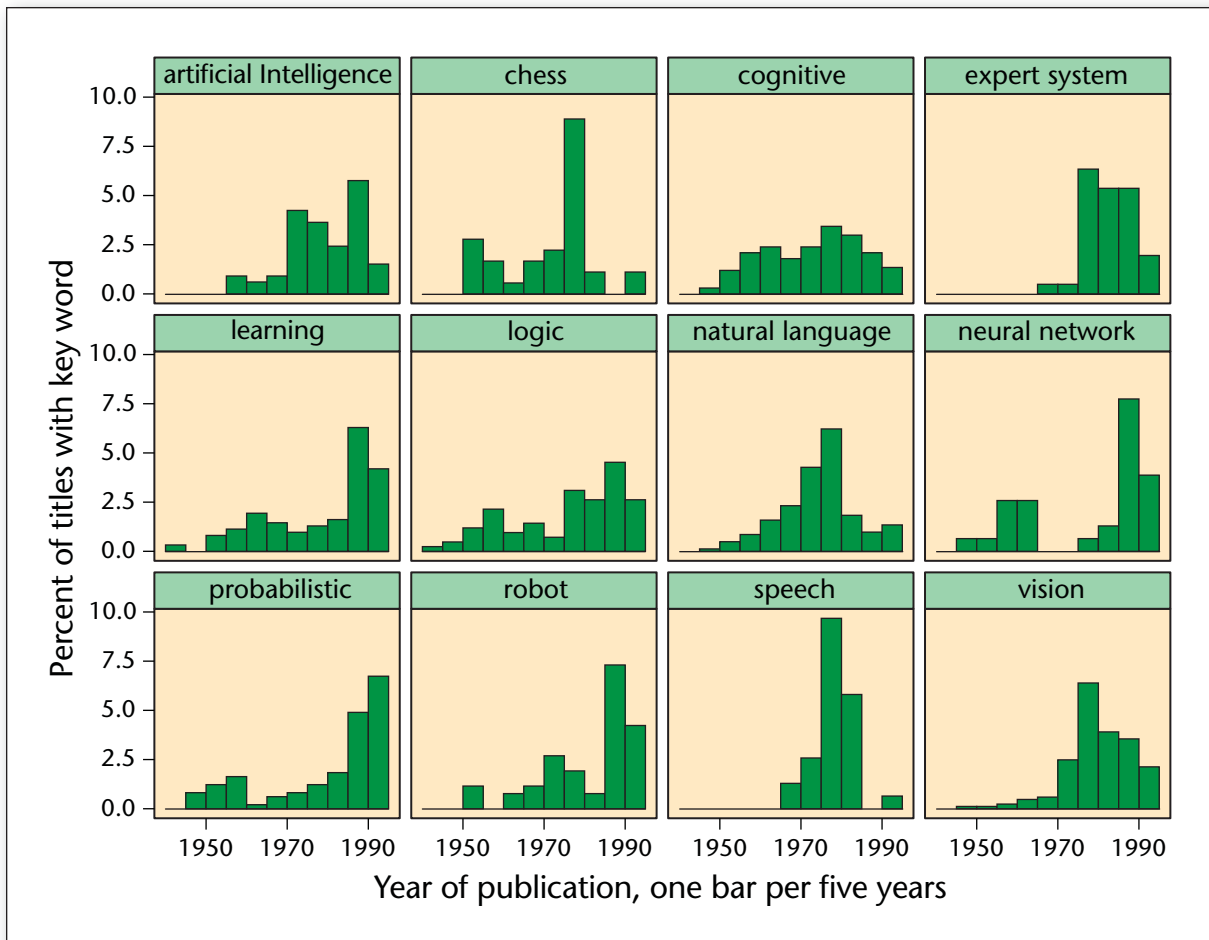*Figure 1. Two-Level Hierarchy of Topics Serving as a Taxonomy of AI Concepts.*

*Figure 2. Relative Frequency of Various AI-Related Key Words Found in
Titles of Classic Publications, as They Appear over the Years.*

The relative frequency of a key word equals the count of publications, published in a particular year, with that key word in the title divided by the number of publications from that year. Each key word is a composite of similar phrases. For example, "probabilistic" includes the key words "probability," "probabilistic," "statistical," "Bayesian," and so on.

always perform as well as other commercial tools, low-quality OCR was quite satisfactory for allowing us to find interesting relationships among the publications, as described in the next section.

## Preliminary Results

We now have a collection of nearly 1850 classic publications, most of which have links to external online repositories or to PDFs stored on AAAI servers, and all of which have tags from the broad topic taxonomy. Influential publications through 1950, the year of Alan Turing's seminal paper in *Mind*, are shown in the sidebar to this article to illustrate the kinds of publications contained in the archive.

We have also explored various analyses of the collection. As shown in figure 2, we plotted the rel-

ative frequency of various AI-related key words found in titles of classic publications, as they appear over the years. We see, for example, that classics about probabilistic reasoning were increasing into the 1990s and classics about expert systems were decreasing, after first appearing around 1970.

We caution the reader against drawing unwarranted conclusions from the key-word frequency plots. The list of classics is a small subset of the greater universe of publications about AI, of which there many hundreds of thousands.[8] Also, after new lines of research have been established with some classic publications, we expect to see more "follow-up" publications than classics. For example, papers about expert systems continue to be published although the number of classics has declined. The plots show, rather, the nature of the
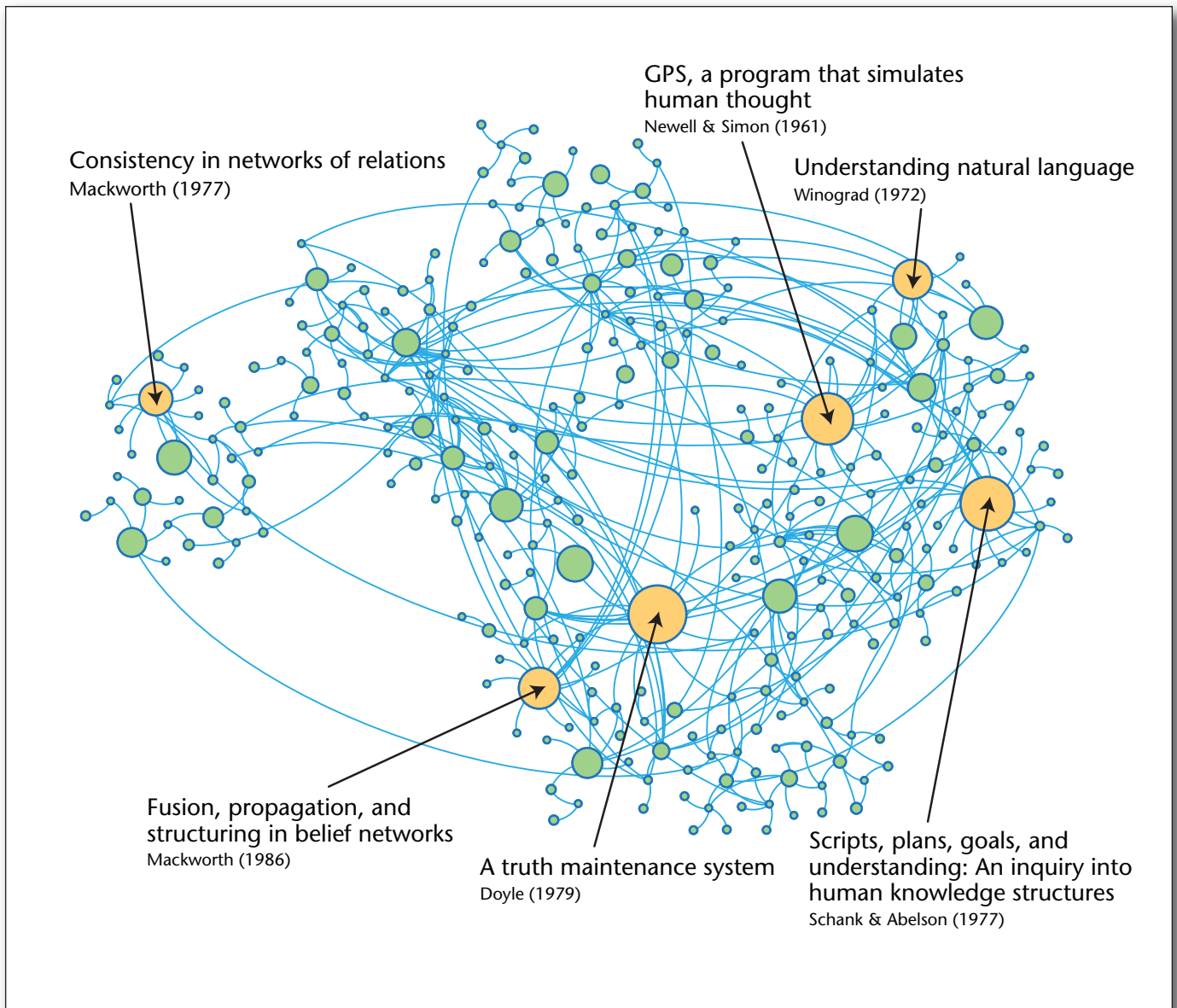
*Figure 3. Network of Citations Among the Classics.*

Each edge is a citation from one classic in the collection to another classic. Each node represents a single publication. Its size represents its PageRank score, which can be generally interpreted as greater importance or significance. A handful of the highest-ranked publications are identified in captions. Only a subset of the collection is represented here as we do not have the full text of every publication.

collection of classic publications and approximate times at which the most influential work was being published.

As mentioned earlier, about 50 percent of the collection was downloaded in full-text form for our own use, converting scanned documents into text as mentioned above. This full text cannot always be made available to the public for legal reasons, but we are working on extracting information of various kinds from the texts.

Thus far, our sole use of the full-text subset is

citation extraction. The full-text articles are searched for references to other publications in the collection. Many citations were missed, some due to unexpected citation formats or string matching limitations, but mostly because we only had the full text of only half of the collection. For example, more often than not, we do not possess the full text of classic books. False positive citations are minimal, however, because each reported citation was manually verified.

Using this "who cites whom" information

*Figure 4. Screenshot of Information About One of the Publications in the Classics List.*

In addition to publication metadata such as authors, source, abstract, and related topics from the topic hierarchy, some publications have citation links to other classics in the collection, for example aitopics.org/publication/truth-mainte-nance-system.

among the classics, we built a citation network. This network, shown graphically in figure 3, is able to highlight especially important classic publications. Some publications are cited much more often than others; conversely some cite very popular publications. Because the network of citations is similar to the World Wide Web — some websites have many incoming links, and some link to popular sites — we use the PageRank algorithm (Page et al. 1999) to analyze citations. This is a technique for evaluating quantitatively the importance of a website, or publication, based on the incoming and outgoing links or citations. Essentially, publications with greater PageRank scores are more significant because they are commonly cited or cited

by significant works and themselves cite other significant works.

The graph in figure 3 shows each node scaled to its relative PageRank score, so that a larger node represents a more significant publication. The six highest-ranked publications are identified with captions. We expect that readers familiar with the history of AI readily recognize these highly-ranked publications and consider them important to the field.

By integrating the citation network, topic categorization, and other metadata, we have made a collection that is more than just a list. Within AITopics, the historical, the recent, and the entertaining coincide by integrating classics, recent

*Figure 5. Screenshot of the Robots Topic Overview Page.*

Related publications from the collection of classics are shown in a table. Although not shown, related news, videos, podcasts, and other resources are available from the same screen (aitopics.org/topic/robots).

news, videos and podcasts under the same organizational structure. The collection is integrated into the AITopics website in several ways. The classic publications can be viewed, sorted, and searched, just as any other publication, podcast, or video. Citations from a classic publication to others, and vice versa, can also be found (see figure 4). Additionally, when viewing a topic or subtopic, a reader can discover related news, videos, podcasts, and other online resources such as tutorials and overview articles (see figure 5). In essence, the integration of the classics and other information on the AITopics website *contextualizes* the classic publications.

## Conclusions

Collecting a virtual archive of publications that have influenced the growth of AI is an open-ended task. Although many worthy publications have undoubtedly been overlooked, we hope to have included most of the main threads of AI research and a large number of the publications that are widely identified with the introduction of those threads. Linking titles to online copies allows students and other researchers to read these publications without seeking out physical copies, some of which are housed in just a few locations. Contextualizing the publications by showcasing related

news, online resources, videos, and podcasts also helps students and others make sense of the larger AI field.

Much remains to be done. Most importantly, there will always be justifiable additions and corrections to the list of influential publications, and we continue to search for digitized and freely accessible versions of every title. Part of this process is obtaining permission to digitize and link to full-text versions of hard-to-find material, which we continue to pursue.

Further analysis of the collection seems to be a useful enterprise. More sophisticated textual analysis may expose how new terms, such as "knowledge-based systems" and "semantic networks" have come into usage over time. More extensive citation analysis may show which publications served as the building blocks for significant threads of research.

We also hope to enhance the integration of the classic collection into the larger AITopics website. Due to our present inability to obtain or expose the full text in some cases, we cannot yet provide a public search facility that can look deep into the text of every publication. However, when we can link the full text of classics to contemporary work, we believe readers will find it quite useful to be able to find early explications of important ideas just by performing a simple search.

Assembling and maintaining the virtual archive of classics in AI is very much a work in progress. Suggestions for other work to include will be welcome. We also welcome collaboration to analyze the data we have already collected.

## Acknowledgements

## Notes

1. The mission statement for the organization published on www.aaai.org states: "Founded in 1979, the Association for the Advancement of Artificial Intelligence (AAAI) (formerly the American Association for Artificial Intelligence) is a nonprofit scientific society devoted to advancing the scientific understanding of the mechanisms underlying thought and intelligent behavior and their embodiment in machines. AAAI aims to promote research in, and responsible use of, artificial intelligence. AAAI also aims to increase public understanding of artificial intelligence, improve the teaching and training of AI practitioners, and provide guidance for research planners and funders concerning the importance and potential of current AI developments and future directions."

2. aitopics.org.

3. aitopics.org/sitemap.

4. scholar.google.com.

5. drupal.org.

6. Available at github.com/AAAI/article-analysis.

7. code.google.com/p/tesseract-ocr.

8. In just one publisher's list of journals, a search for publications with "artificial intelligence" in their titles returns more than 300,000 references.

## References

AAAI Strategic Planning Working Group [Ron Brachman, Bruce Buchanan, Carol Hamilton, Eric Horvitz, Alan Mackworth (Chair), Martha Pollack, Francesca Rossi, Manuela Veloso, and David Waltz]. 2008. AAAI Strategic Plan — 2008, unpublished document dated June 24, 2008, Palo Alto, CA: AAAI.

Buchanan, B. G., and Glick, J. 2001. AI Topics — Organizing Online Knowledge Sources About AI for the Lay Public. *Intelligence* 12(3): 25–30.

Buchanan, B. G., and Glick, J. 2002. AI Topics: A Responsibility to Celebrate AI Responsibly. *AI Magazine* 23(1): 87–94.

Buchanan, B. G.; Glick, J.; and Smith, R. G. 2008. The AAAI Video Archive. *AI Magazine* 29(1): 91–94.

Dong, L.; Smith, R. G.; and Buchanan, B. G. 2011. NewsFinder: Automating an Artificial Intelligence News Service. In *Proceedings of the Twenty-Third IAAI Conference on Innovative Applications of Artificial Intelligence,* 4–12. Menlo Park, CA: AAAI Press.

Eckroth, J.; Dong, L.; Smith, R. G.; and Buchanan, B. G. 2012. NewsFinder: Automating an AI News Service. *AI Magazine* 33(2): 43–54.

Minsky, M. 1961. A Selected Descriptor Indexed Bibliography to the Literature on Artificial Intelligence. *IRE Transactions on Human Factors in Electronics,* HFE-2: 39–55. Reprinted in *Computers and Thought*, ed. E. A. Feigenbaum and J. Feldman. New York: McGraw-Hill, 1963.

Nilsson, N. 2010. *The Quest for Artificial Intelligence*. Cambridge: Cambridge University Press.

Page, L.; Brin, S.; Motwani, R.; and Winograd, T. 1999. The PageRank Citation Ranking: Bringing Order to the Web. Technical Report, Stanford InfoLab, Stanford University, Stanford, CA.

## Influential Bibliographies in Publications

Publications whose bibliographies provided the initial set of several thousand influential publications in the history of AI. Authors included these titles for many reasons, so we used our own judgment to select approximately 2,000 that represent major conceptual advances, research breakthroughs, and experimental results. Also included are selected references from the papers in *Machine Intelligence 1–6*; 1967–1971, selected references from *Machine*

## Announcing the AAAI Chapter Program

AAAI is delighted to announce the establishment of the AAAI Chapter Program. AAAI members are invited to establish regional chapters that promote the mission of AAAI, including increasing knowledge of, and greater interest in, artificial intelligence (AI); promoting greater awareness of AI and its potential among the chapter's local community at large; fostering greater interactions between the chapter AI community and the international AI community; and promoting greater participation of and membership in AAAI by chapter members.

To kick off this program, members have recently established chapters in Beijing, China and at the University of Southern California in Los Angeles. Chapters operate as independent organizations, but are affiliated with AAAI. Complete AAAI Chapter Guidelines are available at www.aaai.org/ Membership/membership-chapters.php.

If you are interested in starting a AAAI Chapter, please contact chapters13@aaai.org, or fill out the application form on the AAAI website.

*Learning* 1(1); 1986, Science Citation Index of most frequently cited papers (compiled by Colleen McCarthy); and the AAAI list of papers receiving the "Distinguished Paper" award.

Barr, A.; Cohen, P. R.; and Feigenbaum, E. A., eds. 1981. *Handbook of Artificial Intelligence, vols. I–IV*. Reading, MA: Addison-Wesley.

Blake, D. V.; and Uttley, A. M., eds. 1959. *Mechanisation of Thought Processes vols. 1 and 2*. London: Her Majesty's Stationary Office.

Laird, J. 2012. *The SOAR Cognitive Architecture*. Cambridge, MA: The MIT Press.

McCarthy, J., and Hayes, P. 1969. Some Philosophical Problems from the Standpoint of Artificial Intelligence. In *Machine Intelligence* 4, ed. D. Michie and B. Meltzer, 463–502. Edinburgh: Edinburgh University Press.

Michie, D. 1974. *On Machine Intelligence*. Edinburgh: Edinburgh University Press.

Minsky, M. 1961. A Selected Descriptor Indexed Bibliography to the Literature on Artificial Intelligence. *IRE Transactions on Human Factors in Electronics, HFE-2*: 39–55. Reprinted in *Computers and Thought,* ed. E. Feigenbaum and J. Feldman. 1963. New York: McGraw Hill.

Nilsson, N. 2010. *The Quest for Artificial Intelligence*. Cambridge: Cambridge University Press.

Russell, S.; and Norvig, P. 2010. *Artificial Intelligence: A Modern Approach*, 3rd ed. Engelwood Cliffs, NJ: Prentice Hall.

Sloman, A. 1978. *The Computer Revolution in Philosophy*. Atlantic Highlands, NJ: Humanities Press.

Smith, C. H. 1981. An Inductive Inference Bibliography. Department of Computer Sciences, Purdue University. CSD TR 323. March 1961 Revision.

Subramanian, D., and Buchanan, B. G. 1985. *A General Reading List for Artificial Intelligence*. Stanford University Computer Science Dept. Report No. STAN-CS-86-1093, Stanford, CA.

Webber, B., and Nilsson, N. 1982. *Readings in Artificial Intelligence*. Menlo Park, CA: Tioga Press.

*Stanford Ph.D. Qualifying Exam in AI*. 1988. Unpublished study list.

**Bruce G. Buchanan** is University Professor of Computer Science Emeritus at the University of Pittsburgh. Before joining the Pitt faculty as a professor of computer science, medicine, and philosophy he was a professor of computer science (research) at Stanford. He has supervised more than three dozen Ph.D. dissertations, many on interdisciplinary topics. He holds a Ph.D. in philosophy from Michigan State University; he is a Fellow of AAAI and the American College of Medical Informatics and an elected member of the National Academy of Science Institute of Medicine; and he has served as treasurer and president of the Association for the Advancement of Artificial Intelligence.

**Joshua Eckroth** is a Ph.D. candidate at the Ohio State University. He is studying artificial intelligence, cognitive science, and mathematical logic. His research focuses on how an abductive reasoning system may employ abuctive metareasoning to improve its estimates of a changing world.

**Reid G. Smith** is the enterprise content management director and IT upstream services manager of Marathon Oil Corporation. Prior to joining Marathon, he was responsible for information solutions at Medstory. Earlier, he led the knowledge management program at Schlumberger and managed a number of the company's research laboratories in the United States and the United Kingdom. Smith holds a Ph.D. in electrical engineering from Stanford University and is a Fellow of the Association for the Advancement of Artificial Intelligence.