Artificial Intelligence—A Modern Approach *A Review*

Martha E. Pollack

■ *Artificial Intelligence: A Modern Approach,* Stuart Russell and Peter Norvig, Prentice Hall, Englewood Cliffs, New Jersey, 1995, 932 pp., $62.90, ISBN 0-13-103805-2.

Within the past three years, several new AI textbooks have been published, all written by well-known researchers (Dean, Allen, and Aloiminos 1995; Russell and Norvig 1995; Shoham 1994; Ginsberg 1993). Thus, it is not surprising that the authors of each of these textbooks have sought a way to distinguish their own book from the rest. The hook for Stuart Russell and Peter Norvig's new textbook, *Artificial Intelligence: A Modern Approach*, is indicated on the book's jacket, which proclaims it to be "the intelligent agent book." Throughout the book, attention is given to the question of how individual AI algorithms might be incorporated into a larger agent that interacts with its environment. I've been teaching from the Russell and Norvig textbook for the past two years,[1] and I think it's a terrific book; however, I think its merit is more or less independent of its identity as the intelligent-agent book.

What makes this textbook so good? First, it is remarkably comprehensive. Its eight main sections cover (almost) everything an AI student needs to know to begin reading the primary literature. Each section comprises several chapters, which provide detailed coverage of the relevant material. Remarkably, the book not only provides sufficient background to begin serious work in AI but also provides just necessary background: There isn't much in it that could readily be omitted by a graduate student in AI. The eight sections are (1) Artificial Intelligence (introductory material); (2) Problem-Solving (search and game playing); (3) Knowledge and Reasoning (propositional and predicate logic, inference techniques, knowledge representation); (4) Acting Logically (planning); (5) Uncertain Knowledge and Reasoning (probabilistic reasoning, Bayesian nets, decision-theoretic techniques); (6) Learning (inductive learning, neural nets, reinforcement learning); (7) Communicating, Perceiving, and Acting (natural language processing, computer vision, robotics); and (8) Conclusions (philosophical foundations and summary).

The inclusiveness of the text comes at a price: At 932 pages, the book is far too long to be completely covered in a single-semester course. In the preface, the authors suggest several alternative paths through the book that could serve as the basis of a one-semester course. At the University of Pittsburgh, my colleagues and I cover roughly the first half of the book (Sections 1–4) in the first-semester introductory graduate AI course, covering most of Sections 5 through 8 in a second-semester course. After shelling out almost $63, our students are relieved to be able to use the same book for two semesters.

What makes this textbook so good? First, it is remarkably comprehensive. … [It] cover[s] (almost) everything an AI student needs to know to begin reading the primary literature.

A second key feature of the book is its use of uniform pseudocode to present numerous AI algorithms. The pseudocode is consistent with what you find in textbooks in other areas of computer science. Too often in the past, AI textbooks have either presented highly informal descriptions of algorithms or have given actual code in Lisp or PROLOG; one suspects that soon enough, we will see "AI in C++" books. Russell and Norvig abstract away from the particulars of any programming language and allow the student to focus on the algorithms themselves. Most instructors will also want to include significant programming exercises; the authors have an extensive, well-indexed code repository, available by anonymous ftp, which contains Lisp programs for algorithms that are discussed in the book and the exercise sections. They also have a textbook home page, http://www.cs. berkeley.edu/~russell /aima.html, which is useful both for students and for instructors.

The organization of the book is generally principled and orderly, and the authors draw numerous connections not only among AI topics but also between AI and other subdisciplines of computer science and between AI and other disciplines. The chapter on philosophical foundations (Chapter 26) is an unusual but welcome inclusion, pointing out the relevance of specific AI techniques described in earlier chapters to debates about the very possibility of AI. Throughout the book, the writing is clear and engaging, and the authors convey an appropriately positive view of the field. To read this book is to get a sense of the intellectual substance of the field—to realize how much good work has been done in AI.

So is the book perfect? No, of course not. At best, any textbook is likely to give rise to a set of minor complaints; here are some of mine in roughly increasing order of importance: First, the section on knowledge and reasoning (Section 3) could be better organized: It includes a chapter on knowledge base design

(Chapter 8) stuck awkwardly between the chapter that introduces the syntax and semantics of first-order logic and the one that discusses inference in first-order logic. Many of the authors' proposed paths through the book involve omitting Chapter 8 anyway, but courses that include it might want to defer it until after the material on logic has been finished.

Second, the discussion of inference in first-order logic could be improved. The authors begin by presenting a subset of a natural deduction system, then show how it can be replaced with a single rule of generalized modus ponens, which reduces the problem of profligate branching as a result of universal elimination. Finally, they note that the resulting system is incomplete and, so, present resolution as a way of achieving logical completeness. So far, so good, except that the explanation of why the modus ponens–based inference system is incomplete is relegated to a single brief paragraph. Although the book mentions that generalized modus ponens should only be applied to theories comprising Horn sentences ("otherwise, [the non-Horn sentences] could never be used" [p. 270]), the connection between incompleteness and non-Horn sentences is never drawn directly.

Third, Section 4 (Acting Logically) reads as if there were no alternative to partial-order planning, or what is more commonly called plan-space planning. The authors' decision not to include a discussion of total-order planning, or "state-space planning," probably springs from their justified, and by and large successful, decision not to include merely historical information in the main body of each chapter; after all, most current planning research involves plan-space algorithms. However, total-order planning has not completely been discredited as an effective strategy, and students ought at least to be introduced to the issues surrounding the decision to use one approach or the other. In my class, I supplement the material in this section with Weld's (1994) survey article.

Finally, despite the general accolades the book gets for comprehensiveness, it does omit one important topic—truth maintenance systems—which is only mentioned briefly in a 2½ –page subsection, without benefit of any algorithms. There isn't an accessible, short tutorial paper on this topic that can be used as a supplement, at least as far as I know, so this oversight is particularly problematic.

In their preface, Russell and Norvig assert that the book is intended primarily for use by undergraduates but could also be used in graduate courses if augmented with primary sources. In fact, I have found it to be quite well suited for use in graduate courses; the augmentation with primary materials happens in my classes automatically as the students prepare a required term paper. I haven't used the text in an undergraduate class but think that many undergraduates would find some of the material to be too advanced, in particular in the sections on uncertain reasoning and learning, both of which require a fair bit of mathematical sophistication.

My minor complaints aside, I've found it a pleasure to teach from this book, and I have also used it frequently as a reference source. I have to confess, though, that although I agree with the authors' claim that "the problem of AI is to describe and build agents that receive percepts from the environment and perform [intelligent and/or rational and/or human-like] actions" (p. vii), in my class lectures, I give little air time to the intelligent-agent theme. In some cases, this is because the connection seems forced, for example, in the discussion of game playing. In other cases, it seems somewhat irrelevant, for example, in much of the discussion on search algorithms, which are important even outside the context of agency. In yet other cases, the connection between particular algorithms and a larger agent architecture seems too obvious to belabor in class, although I suspect that many of my students are glad to have such connections made explicit in the textbook. Mostly, I bypass the material on agents because lecture time is a highly limited resource, and there's such a wealth of material in Russell and Norvig's book; it takes a full semester of classes to present the AI algorithms and their analyses even for half the book, without spending time on the details of how these algorithms can be connected in a more complete agent.

If you want to teach an AI course around an agent theme—and I don't necessarily think this idea a bad one—this book will make it easy for you to do so and to do so well. However, even if you think that *agents* is just the latest buzzword, don't let the fact that this text is billed as "the intelligent-agent" book dissuade you from adopting it for your class or buying it as a reference book. *Artificial Intelligence: A Modern Approach* will provide a first-rate education in AI even to the reader who skips all the specially agent-oriented material.

Notes

1. I have been using preprints; the book appeared in print late last fall.

References

Dean, T.; Allen, J.; and Aloimonos, Y. 1995. *Artificial Intelligence: Theory and Practice*. Redwood City, Calif.: Benjamin/Cummings.

Ginsberg, M. 1993. *Essentials of Artificial Intelligence*. San Francisco, Calif.: Morgan Kaufmann.

Shoham, Y. 1994. *Artificial Intelligence Techniques in PROLOG*. San Francisco, Calif.: Morgan Kaufmann.

Weld, D. 1994. An Introduction to Least Commitment Planning. *AI Magazine* 15(4): 27–61.

Martha E. Pollack is associate professor of computer science and intelligent systems at the University of Pittsburgh. She received the Computers and Thought Award in 1991 and a National Science Foundation Young Investigators Award in 1992. Her current research interests include computational models of rationality, plan generation and recognition, natural language processing, and AI methodology.