

Book Review

“AI Assistants.” Roberto Pieraccini, 2021, MIT Press.Marc Cheong ^{1*}¹ Faculty of Engineering and IT, The University of Melbourne, Parkville VIC 3010, Australia.* Correspondence: marc.cheong@unimelb.edu.au

Kudos to Roberto Pieraccini for his compact explainer on *AI Assistants* (2021). Pieraccini has skillfully taken the reader on a journey into the realm of AI assistants – from their history and evolution, to the technology behind these systems, to the current landscape of everyday assistant systems *du jour* – in just over 200 pages. In the same vein as his earlier volume, “*The Voice in the Machine: Building Computers that Understand Speech*” (Pieraccini 2012), Pieraccini has a knack of being able to “convey[...] the essence of modern statistical speech processing without resorting to mathematics”, to borrow from Steve Young’s review (The MIT Press 2022). Combining personal anecdotes, in-depth industry knowledge, and reader-friendly examples, Pieraccini can do so without requiring the reader to have a scientific background, which is a very strong suit for the book.

Compared to *The Voice in the Machine* (Pieraccini 2012), however, this book’s focus is more on “synthesizing specialized subject matter for nonspecialists” (2021, p. ix), as part of MIT Press’s *Essential Knowledge Series*. (As a disclaimer, for this review, I have been kindly provided an advance copy by MIT Press, which might not reflect the final copy in its entirety.) Nonetheless, my review will summarize the book’s main parts, from my interpretation. At a high level, these are: (1) definitions of the core concepts; (2) explorations of four main technical components of AI assistants; and finally (3) the big picture of, and contemporary trends in, AI assistant use. In the spirit of reviewing a book on AI assistants, and for an appreciation for the technology discussed within, I have dictated most of this manuscript’s initial drafts using several dictation engines.

The first part, encompassing Chapters 1 and 2, contain key concepts, terminology, and a brief primer on the technology prerequisites. Chapter 1 introduces the reader to concepts which are frequently invoked in the discussion of virtual assistant technologies. Each key component is introduced with concise descriptions, along with a high-level schematic (2021, p. 8); these set the scene for the second part of the book, where each of these components are discussed in detail. One aspect which I would like to comment on is Pieraccini’s ascribing of *intelligence* to rule-based applications such as a tax-processing software package (2021, pp. 10-11). I agree with Pieraccini’s ascribing of “expertise” to such pieces of software, but not necessarily “intelligence”: though I understand that this decision might have been made due to the scope and space limitations of the book. The example of rule-based systems would probably have been better described as an *expert system* (discussed later in p.24), under the banner of *automated decision-making*, as is the wont of technical reports and policy papers (Administrative Review Council 2004). The ascription of intelligence to a machine has been the subject of many a debate going back to, say, John Searle (Bringsjord and Govindarajulu 2019).

Chapter 2, the introduction to AI and machine learning, is indeed a *tour de force*, to Pieraccini’s credit. With his expert flair, the chapter concisely summarizes the landscape of AI and machine learning without using jargon or complex mathematical formulae, per my introduction to this review (again, quoting Young in The MIT Press, 2022), rendering it extremely reader-friendly to a non-technical audience. Concepts such as classic AI, machine learning (ML), supervised vs. unsupervised learning, neural networks (NNs) and all its

Citation: Cheong, Marc. 2022. “AI Assistants.” Roberto Pieraccini, 2021, MIT Press. *Journal of Ethics and Emerging Technologies* 32: 1. <https://doi.org/10.55613/jeet.v32i1.102>

Received: 23/05/2022

Accepted: 23/05/2022

Published: 31/07/2022

Publisher’s Note: IEET stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

flavors, amongst others, are covered extremely elegantly in a compact 30 pages (2021, pp. 21-51). Pieraccini also provides a convincing example of how a digit recognizer (2021, pp. 39-42) can be implemented using NNs: again, without the use of complex mathematical formulae at all, and clear enough to follow without needing a single diagram! This is reminiscent of – and complements – how Andrei Burkov's *The 100 Page Machine Learning Handbook* (Burkov 2019) summarizes the landscape of ML albeit for those with computing and mathematical backgrounds. The only improvement that I can foresee for Chapter 2 is a potential discussion of the 'AI Winter' of the 1980s and the progress of technological evolution pre-deep learning (Hendler 2008), which may be interesting for the reader to explore.

The second substantive part of the book consists of five chapters (Chapters 3-7), each roughly corresponding to a major component/module of a modern-day AI assistant system. Chapter 3 on speech recognition (SR) discusses the challenges in converting a human speech signal – the sound of our voice – into text that can be understood directly by machines. The history of speech recognition is clearly laid out from the early days of the technology (*circa* the 1970s), and the two paradigms, *statistical models* versus *rule-based approaches* are given an introduction. The case of voice response (IVR) systems (2021, pp. 70-74) is used to introduce the reader to the inner workings, complexity, and motivations for SR technology in public-facing systems. This is indeed effective, as the reader could relate to their own experiences with such systems: to wit, *how many of us have been frustrated with computerized answering machines and would prefer a human operator?* In my opinion, however, instead of expanding on technologies such as *VoiceXML* and *Speech2Text* in detail (2021, pp. 76-80), it would have been preferable if the discussion on the history of SR systems included assistive technology. Assistive technologies leveraging SR have been beneficial to users with disabilities, allowing them to navigate user interfaces, take notes, and learn effectively. SR software (such as those from Dragon) are presented as successes for “niche markets” (2021, p. 69), i.e., catering towards certain industries to reduce the need for typing. In fact, these technologies have vastly helped those with disabilities (Bain, Basson, and Wald 2002; De La Paz 1999). Thus, assistive technologies could have been given a more substantive mention throughout the book, as AI assistants (or its constituent components) play a role in promoting accessibility for modern-day devices. To Pieraccini's credit, this chapter does a very good job in explaining complex concepts, such as one-hot vectors and word embeddings in a practical, applied, manner. The inclusion of more diagrams summarizing the above, for more visual learners, would be beneficial for future editions of the book.

Moving on, Chapter 4 explores how machines understand text which has been transcribed from the ASR system. A very good feature of this chapter is a reflection on how our communication patterns in everyday life (for example, the use of filler words) affect the accuracy of natural language understanding (NLU) methods, which would bring new light to the reader on why NLU implementation is not as straightforward as some might think. Using the examples of understanding phone numbers (2021, pp. 101-104), and understanding booking requests (pp. 106-110), the reader is guided into thinking about the various ways we use language, and how the NLU system needs to consider all these different ways. The concept of grammatical rules is introduced effectively with the examples provided, which is a breath of fresh air compared to, say, how a textbook on parsers and compiler construction will typically approach the topic.

Chapter 5 focuses on natural language and speech generation: in other words, *how does the machine talk back to us?* This chapter is written rather well, again with lots of examples – from the IVR of previous chapters, to examples of conversations with everyday AI assistants (2021, p. 150-152), to an amusing but very informative read-aloud experiment (p. 153) – guiding the reader's train of thought.

Chapter 6 focuses on a more complex component – the dialogue manager – which, in short, is the intermediary between the 'understanding' of what was being mentioned by the user and what is to be communicated to the user. The discussion of finite state machines (FSM), linking to the discussion on IVRs in the prior sections, is one of the strong points of

this chapter. It allows the user to intuit the problem of complexity in building an FSM with realistic examples (2021, p.176), without having to resort to a lengthy discussion on automata and Turing Machines. Chapter 6's conclusion (2021, pp. 191-192), would have been a good opportunity to introduce readers to the concept of the Turing Test *viz.* natural conversation as judged by a human. This could then be segued into a philosophical discussion about, say, Google's Duplex and Mina technologies, and what it means (ethically, phenomenologically) to have a conversation with a purely non-human interlocutor, sometimes purely by chance (as in the case of Google Duplex creating a booking on behalf of a human).

The final section of this book begins with Chapter 7, on how users interact with the AI Assistant. This is a comprehensive overview, reflecting the technical challenges for implementing the current state of the art of consumer AI assistants (Apple's Siri, Google's Assistant, and Amazon's Alexa). Potentially due to space limitations, the discussion on privacy is only provided in passing within Chapter 7 (more on this later). Say, when discussing issues related to "listening for the wake word" (p. 197), more detail on challenges/idiosyncrasies regarding offline speech recognition of the wake word could be added for the curious reader. These range risks of accidental or 'mystery' mis-activations (Gewirtz 2018) and steps to resolve them, to privacy-conscious 'hacks' by users to stop eavesdropping (Wilson 2019), all of which could have benefited from some discussion. The latter part of Section 7 is dedicated to Jibo, "the first social robot", with an interesting and personal anecdote on the development, marketing, production, and user experience of such a robot. This section evokes empathy from the reader as it takes the reader on a journey on how the designers made Jibo 'human', the challenges needed to develop the prototype, all the way up to the unfortunate sunseting of the project.

Finally, Chapter 8 provides a compact summary of the issues covered in the entire book. Pieraccini sums up with a reflection on the long journey AI assistants and its constituent parts have taken, encompassing "more than seven decades of study" (2021, p. 224). The concluding vignette of Apple's futuristic 1987 *Knowledge Navigator* video (2021, p. 227-230) sums up the book, contextualizing the (yet unsolved) challenges identified in the video for the now-enlightened reader.

Throughout my reading of the book, I cannot help but notice several overall improvements that could have been made. Discussions on the human factors, the societal factors, and the philosophical and ethical factors concerning AI Assistants (and each of its constituent parts) could have been made more in-depth. Several examples follow. Firstly, the book is targeted towards an US-centric English-speaking, audience. This lends to framing of issues such as a foreign "accent" (2021, pp. 3, 54) as an obstacle for voice recognition systems, rather than positing it as an issue with the diversity of data (analogous to the case of, say, modern-day ML applications with a class imbalance in its training data). Another example (2021, p. 157) is the discussion about how different languages have different "difficulties" with speech generation, which was abruptly cut short: the reader is left guessing as to why these issues exist in the first place (e.g., *Is it due to the incorrect encoding of the character data for other languages?*).

Secondly, another suggestion is that the book could have focused on more 'humanistic' impacts/considerations of SR systems – accessibility (as seen earlier), privacy, user diversity, psychological aspects of AI use, and so forth. Take privacy for example. The issue of privacy is brought up in the concluding chapter (2021, pp. 226-227), amongst other places. In some parts, the discussion posits private data as "value" (2021, p. 227), to be traded for accuracy and technological improvement, which I might not fully agree with. This discussion could be more nuanced – sometimes, users have no choice but to agree to this trade off without an ability to opt out; in other cases, users do not know what is being done with their data – these subtleties lend themselves for further discussion. All these could have been phrased as tensions between the humanistic aspects of AI assistant use, versus the technological challenges in building such systems, to give readers as sense of perspective from both sides of the coin.

Another minor issue, which is potentially due to my reviewing of an advance copy, is that there are certain parts in the book in which key definitions are not provided to the reader until much later. (For example: “HMMs” were mentioned in p. 35; but not defined as “Hidden Markov Models” until p. 42, and the “mel” scale was used in p. 46 albeit only introduced on p. 84). Cross-referencing of these concepts to other chapters where they were introduced could be more effectively made. Hopefully these issues will not find themselves in the final print edition of this book.

All-in-all, I commend Pieraccini for this excellent book which could quickly upskill the non-technical reader over the course of a day. Pieraccini skillfully teaches key concepts from a technologist’s perspective – through a selection of well-thought-out examples and everyday applications – without having to burden the reader with superfluous technical jargon (as is the case with many textbooks!).

Funding: This work received no external funding.

Acknowledgements: The author thanks Fabio Tollon, the journal’s Book Reviews Editor, and MIT Press, for providing the advance copy of the book for review.

Conflicts of Interest: The author declares no conflict of interest.

References

- Administrative Review Council. 2004. “Automated Assistance in Administrative Decision-Making: Report to the Attorney-General.” Attorney-General’s Department, Australian Government.
<https://www.ag.gov.au/LegalSystem/AdministrativeLaw/Pages/practice-guides/automated-assistance-in-administrative-decision-making.aspx>.
- Bain, Keith, Sara H. Basson, and Mike Wald. 2002. “Speech Recognition in University Classrooms.” In *Proceedings of the Fifth International ACM Conference on Assistive Technologies - Assets '02*. New York, New York, USA: ACM Press.
<https://doi.org/10.1145/638249.638284>.
- Bringsjord, Selmer, and Naveen Sundar Govindarajulu. 2019. “Artificial Intelligence.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2019. Metaphysics Research Lab, Stanford University.
<https://plato.stanford.edu/archives/win2019/entries/artificial-intelligence/>.
- Burkov, Andriy. 2019. *The Hundred-Page Machine Learning Book*. Quebec City, Canada: Andriy Burkov.
- De La Paz, Susan. 1999. “Composing via Dictation and Speech Recognition Systems: Compensatory Technology for Students with Learning Disabilities.” *Learning Disability Quarterly: Journal of the Division for Children with Learning Disabilities* 22 (3): 173–82. <https://doi.org/10.2307/1511284>.
- Gewirtz, David. 2018. “Has Alexa Snapped? Why Your Echo Sometimes Does Creepy Things.” ZDNet. March 9, 2018.
<https://www.zdnet.com/article/has-alexa-snapped-why-alexa-sometimes-laughs-or-does-other-creepy-things/>.
- Hendler, James. 2008. “Avoiding Another AI Winter.” *IEEE Intelligent Systems* 23 (2): 2–4. <https://doi.org/10.1109/mis.2008.20>.
- Pieraccini, Roberto. 2012. *The Voice in the Machine: Building Computers That Understand Speech*. MIT Press.
- — —. 2021. *AI Assistants*. The MIT Press Essential Knowledge Series. London, England: MIT Press.
- The MIT Press. 2022. “The Voice in the Machine | The MIT Press.” The MIT Press. 2022. <https://mitpress.mit.edu/books/voice-machine>.
- Wilson, Mark. 2019. “This Is the First Truly Great Amazon Alexa and Google Home Hack.” Fast Company. January 14, 2019.
<https://www.fastcompany.com/90290703/this-is-the-first-truly-great-amazon-alexa-and-google-home-hack>.