Kieran Finnegan
Chapter 13

1. TRUE/FALSE - The computer on the starship Enterprise – with its vast store of knowledge and seamless understanding of the questions put to it – has long been a lodestar for human-computer interaction, envied by Star Trek fans and AI researchers alike (and the intersection of these groups is, shall we say, not insignificant).

True

2. While computers can accurately transcribe our requests, they remain far from the “final frontier” of question answering systems. What, according to Melanie Mitchell, is that final frontier?

Understanding the question being asked.

3. What is the name of the natural-language researcher who let the Watson development team at IBM?

David Ferrucci

4. In just two or three sentences, describe Watson.

Watson is a machine designed to win at jeopardy. It was trained on the specific format of Jeopardy questions as well as a massive trove of general knowledge.

5. In conceding Watson’s inevitable victory, he added the following pop-culture reference to his answer card for the final Jeopardy clue: “I for one welcome our new computer overlords.” Later, he quipped, “To my surprise, losing to an evil quiz-show-playing computer turned out to be a canny career move. Everyone wanted to know What It All Meant, and Watson was a terrible interview, so suddenly I was the one writing think pieces and givein TED Talks ... Like Kasparov before me, I now make a reasonable living as a professional human loser.” Who is he?

Ken Jennings

6. TRUE/FALSE - Following Watson’s win, the AI community was divided as to whether Watson was a true advance in AI or a “publicity stunt” or “parlor trick,” as some called it. While most people agreed that Watson’s performance on Jeopardy! was extraordinary, the question remained: Was Watson actually solving a genuinely hard problem – responding to sophisticated questions posed in colloquial language? Or is the task of responding to Jeopardy! clues, with their very particular linguistic format and fact-driven answers, actually not so hard for a computer with a builtin access to Wikipedia, among other huge data repositories? Not to mention that the computer has been trained on a hundred thousand Jeopardy! clues with formats very similar to the ones it was faced with.

True

7. TRUE/FALSE - Beyond sharing the same name, the same planet-with-swirling-lights logo, and the well-known pleasant robotic voice, the “Watson” that IBM’s marketing department is pitching today has very little in common with the “Watson” that beat Ken Jennings and Brad Ruttler at “Jeopardy!” in 2011. Moreover, today the name Watson refers ont to one coherent AI system but rather to a suite of services that IBM offers to its customers – mainly businesses – under the Watson brand. In short, Watson essentially refers to whatever IBM does in the space of AI while bestowing on these services the valuable halo of the Jeopardy! winner.

True

8. TRUE/FALSE - For a variety of reasons, IBM Watson Group, as advanced and useful as its products might be, has seemed to struggle more than other tech companies. Some of the company’s high-profile contracts with customers (for example, Houston’s MD Anderson Cancer Center) have been canceled. A raft of negative articles about Watson have been published, often quoting disgruntled former employees arguing that some executives and marketers at IBM have far overpromised what the technology can deliver. Overpromising and under-delivering are, of course, an all-too-common story in AI; IBM is far from being the only guilty party. Only the future can tell what IBM’s contribution will be in AI’s spread to health care, law, and other areas in which automated question-answering systems could have a huge impact. But for now, in addition to its Jeopardy! win, Watson may be a contender for the “most notorious hype” award, a dubious achievement in the history of AI.

True

9. What is SQuAD? Squad is short for Stanford Question Answering Dataset, and is a large dataset of questions to test AI’s answer extraction skills.

10. TRUE/FALSE - In 2018, two groups – one from Microsoft’s research lab and the other from the Chinese company Alibaba – produced programs that exceeded Stanford’s measure of human accuracy on the task of reading comprehension. Soon after, both Microsoft and Alibaba issued statements suggesting that their programs perform reading comprehension tasks at least as well as humans do.

True

11. Please recount MM’s sarcastic description of a recurring recipe for AI research. AI is trained on very specific subset of a task, manages to outperform the small human sample size at said task, claims that “AI has passed human ability at “ the general skill abound.

12. What is a “Winograd Schema?” These are language understanding questions, named for the NLP researcher Terry Winograd.

13. TRUE/FALSE - In 2011, three AI researchers – Hector Levesque, Ernest Davis, and Leora Morgenstern – proposed using a large set of Winograd schemas as an alternative to the Turing test. The authors argued that, unlike the Turing test, a test that consists of Winograd schemas forestalls the possibility of a machine giving the correct answer without actually understanding anything about the sentence. The three researchers hypothesized (in notably cautions language) that “with a very hig probability, anything that answers correc tly is engaging in behaviour that we would say show thinking in people.” The researchers continued, “Our [Winograd schema] challenge does not allow a subject to hid behind a smokescreen of verbal tricks, playfulness, or canned responses. . . . What we have proposed here is certainly less demanding than an intelligent conversation about sonnets (say), as imagined by Turing; it does, however, offer a test challenge that is less subject to abuse.”

True

14. How do computer programs compare with humans at answering Winograd schema questions?

Humans perform better than AI at answering Winograd schema questions.

15. TRUE/FALSE - Similar to computer-vision programs, NLP systems can be vulnerable to “adversarial examples.”

True

16. TRUE/FALSE - MM believes it to be extremely unlikely that machines will ever reach the level of humans on translation, reading comprehension, and the like by learning exclusively from online data, with essentially no real understanding of the language they process, which relies on commonsense knowledge and understanding the world.

True