JORDAN BOYD-GRABER

QUESTIONING ARTIFICIAL INTELLIGENCE

2022

2

Question Answering: The Foundation and Savior of Civilization

Despite its central importance to AI, asking and answering questions is not new—it stretches back into our collective unconcious. Riddles and trivia have persisted for millenia not just because they're entertaining. Indeed, being able to ask and answer questions is close to godliness. From the Sphynx to Gestumblindi, this chapter examines how myths from several cultures connect answering questions to forming identity, unlocking the secrets of the universe, and gaining intelligence. And connecting this to nascent "superhuman" AI. Later, beyond proving worthiness in myth, asking questions helped combat corruption and nepotism, the downfall of multiple civilizations. However, the administration of carefully formed questions saved the China and the United States via the reforms of Wu Zetian to the Pendelton Act. Exams created meritocracy and social mobility while also building an understanding of what it means to write a "fair" and "useful" question.

With many tasks in computer science, asking the question "where did this task come from" has a simple answer usually in the form of an acronym . Character recognition was sparked by the MNIST[1] dataset (Wilkinson et al., 1992; LeCun et al., 1998); entailment was sparked by SNLI (MacCartney and Manning, 2008; Bowman et al., 2015); sentiment classification was defined by Pang and Lee (2008).

And while there are particular *flavors* of question answering that habe been defined to make life easier or more discriminative for computers, the basic task of "question in", "answer out" is natural to any inquisitive three year old. Of course, question answering is not the only "natural" task like this: translation likewise is recognizable to any stranger in a strange land trying to cope with an unfamiliar language: dozens of books have been written on that subject, my favorite is *Le Ton Beau de Marot* (Hofstadter, 1997).

Question answering's roots likewise run deep. In this chapter, I want to argue that question answering is not just a useful task but one that is woven deep into our Jungian subconsious. Our collective civilization is defined by how question are answered; it forms an identity, reveals fundamental truths

[1] Throughout, I will do my best to cite the origin of ideas, but MNIST—like other early works in AI—is hard to pin down to a single citation. The OCR group at the United States National Institutes of Standards and Technology (NIST) was charged with helping the post office sort letters automatically by "reading" envelopes, so they organized a conference to try to get people to give it their best shot and released data to let people train models. One of the approaches from LeCunn et al. years later defined what people mean when they actually say MNIST, which then became a part of countless tutorials for people learning machine learning.

of the universe, and creates a shared reality. And beyond these ancient components, it has historically created a historical basis for a shared expectation of competence that created the modern beuroracy.

2.1   The Sphynx

In the Greek of *Œdipus* (Sophocles, 1959), the Sphynx asked everyone who entered the city a riddle. Here's PDQ Bach's interpretation of the riddle (Schickele, 1990):[2]

> What starts out on four legs then goes round on two Then finishes on three before it's through

And the answer is:

> Well a baby then a man, that's plain, then finally an old geezer on a cane

Now I won't spoil the whole story, but by answering the question correctly, Oedipus revealed that he was smarter than the average bear and also unlocked secrets from his past. In a commencement address to Whitman college, Dana Burgess Burgess explained the difference between this and Google's Natural Questions (covered in detail in Chapter 8) and how that relates to the mission of professors (like me):

> Nobody told Oedipus who he was; he figured that out for himself. ... You can do a Google search to find out the capital of Arkansas (Little Rock?) but you can't do a Google search to find out who you are, what you're good at, what makes you happy, what matters for your life. Information transferal isn't any help with that stuff.

When folks say that a college professor is like a sphinx, they don't usually mean that in a nice way. It suggests that we know the answers students need, but we keep them secret for the sake of our own power. ... That's part of the riddle of [The Sphynx's] hybrid nature: the curse of being forced to solve a riddle is also the gift of the ability to solve riddles.

2.2   The Socratic Method

This aspect of question answering also brings us to another Greek connection to asking questions: teaching through the Socratic method (Trepanier, 2017). Through asking the right questions, a teacher guides the answerer to understanding. While Cranfield questioners are seeking information from a more knowledgeable answerer, Manchester questioners often test less knowledgeable answerers. Similarly, perhaps by asking the right questions, the QA community can coax computers to understand more than they do now (Dunietz et al., 2020; Perez et al., 2020).

Because I'm a professor, I am somewhat biased in my appreciation of this use of questions: both in the classrooms and on exams. If you haven't heard of

[2] I use this fairly unserious version of the riddle because it is fun and Renger (2013) reports hexameter versions, and the origin (Pindar fragments are the first mention) is unclear. So if one has to arbitrarily pick, a version localized into American vernacular is perhaps the most accessible.

this before, you might have experienced it (especially if you're in my classes). Let's hear how John Houseman explain what it is in the 1971 film The Paper Chase (pap, 1973):

> We use the Socratic Method here. I call on you, ask you a question…and you answer it. Why don't I just give you a lecture? Because through my questions, you learn to teach yourselves.

> Through this method of questioning, answering…questioning, answering…we seek to develop in you the ability to analyze that vast complex of facts that constitute the relationships of members within a given society. Questioning and answering.

> At times you may feel that you have found the correct answer. I assure you that this is a total delusion on your part. You will never find the correct, absolute, and final answer. In my classroom, there is always another question…another question to follow your answer.

> Yes, you're on a treadmill. My little questions spin the tumblers of your mind. You're on an operating table. My little questions are the fingers probing your brain.

> We do brain surgery here. You teach yourselves the law but I train your mind. You come in here with a skull full of mush…and you leave thinking like a lawyer.

I want to particularly hone in on the assertion that it "trains your brain". While not quite mythological, the term "Socratic Dialog" goes back to ancient Greece. Plato featured Socrates in his dialogues Euthyphro and Ion (Plato, 1949), giving the perspective of the student the student, the person answering the questions, struggling with multiple interpretations and "showing their work" as they emerge on the other side, enlightened by the exchange.

And indeed, one of the best moments for me as a professor is to see the aha moment in students eyes after they answer a question correctly. And this is still how we train lawyers, who have to use argument and discussion to keep our society ordered! But as we will see later in this chapter, that is not the only way the questions and answers create order in our civilization.

2.3 *Gestumblindi*

In Norse myth, Gestumblindi got into some trouble with King Heidrek and asked the Gods for help (Tolkien, 1960).[3] In response, Odin came to Gestumblindi—looking exactly like him—and stood in his place in front of Heidrek. What took place was an interrogation that went in two directions: King Heidrek attempted to determine who this Gestumblindi-looking person was in front of him and Odin attempted to see how smart King Heidrek was.

What I like about this story is that it presages many of the themes that we will see with AI: ordinary people thinking that AI will help them, an interrogation contest to see who is human or not, and an impersonation. It also presages many of our fears about AI: Odin's final question to Heidrek is:

[3] I had originally wanted to use Auden's translation (since Auden is more highbrow), but Tolkein's translation also evokes (and reminds the reader) of Gollum's battle of riddles with Bilbo that won him the ring, an example of riddles deciding the fate of a (fictional) civilization.

What said Ódin
in the ear of Baldur,
before he was borne to the fire?

We talk in this book frequently about bad questions (Chapter 11), but this a particularly bad question. Not only is it something that only Odin knows, it probably has not happened yet (depending on what you believe about the circularity of time), and when it does happen it will end the world by bringing about Ragnarök.

Once Heidrek hears this question, he knows that the figure before him is not Gestumblindi but rather Odin. And like a frustrated customer who thought they were talking to a human for ten minutes only to figure out that they're talking to a bot, Heidrek takes out his sword and attacks, only for Odin to turn into a bird and fly away.

This verbal jousting is common for the Norse Gods (Mitchell, 2020); being a God requires Omniscience, so the only way to tell if you are in the company of God is through riddles. The only problem is that—at least in Norse Myth— challenging a God is hazardous to your health: those who challenge Odin to a battle of wits often end up dying. My goal here is not to say that we should not ask probing questions; subsequent chapters will argue that we should ask these questions often and advise on how to ask the questions well. Rather, the lesson of Gestumblindi is that the are of making a riddle is prized. Just as some Christians argued that "cleanliness is next to godliness" (Jackson, 1872), in the Norse tradition, questioning is synonymous with being a God, which will return when we talk about the apocalyptic visions of AI (Chapter (?)).

2.4   The Foundation of Civilization

Less exciting than the mythic foundations of civilization are the humdrum matters of administration. But even if it's boring, a pillar of civilization is making sure that those who work for society (in other words, civil servants) are competent. As governments became larger and had to rule an entire continent, countries had to develop civil service examinations. Many countries did this, including the US, whose Pendelton act ended an era of political patronage (Johnson and Libecap, 2007).

But I'd rather talk about the granddaddy of civil service exams, kējǔ, which for 1300 years determined who became part of the intelligentsia of Imperial China (Qian and Totten, 1982). It encouraged social mobility by getting smart people good jobs and it encouraged good governance by making sure that important jobs were done by smart people. And this is important when you have a diverse, massive, empire that is run by a centralized bureaucracy. Again, this shows that civilization is not possible without effective systems of question answering.

This system worked well for so long because it was built on a foundation of fairness. The organizers thought about the potential for bias: the exams were

transcribed so that bad handwriting wouldn't be judged against an applicant, children of current members of the imperial court had to submit their exams in their home province, etc.

Now, it wasn't perfect, Most notably its quotas led to imbalances between regions. But it was important enough that recent scholarship has suggested that its abolition after the Russo-Japanese war and a turn to "Western-style" education on the Prussian model that turns out good soldiers, helped hasten the end of the Qing dynasty (Bai and Jia, 2016).

As Chu (2015) argues, it's not just that you have a test to make sure that the smartest people are doing important jobs in society. Part of the process is also to explain why people got the questions wrong! This ensures that people trust how the questions are being graded. If it's just a black box, it is not an improvement on the capricious systems of patronage that it's supposed to replace. As evidence of this, during the Dà Míng dynasty, there was furor over the punctuation of graded exams.

So why am I, a computer science professor, talking all of this nonsense about history? We use questions to test the intelligence of both humans and machines. I talk about the Turing Test (Chapter 3), the classic test of computer intelligence, and modern leaderboards (Chapter 10), but in testing computer intelligence we shouldn't forget about the lessons we've learned about human intelligence.

One of those lessons that I think that we've forgotten is that questions are not just an evaluation but also for instruction: in the lingo of artificial intelligence, they're training data (Chapter 8) . And because these training data build intelligence, they should be as high quality as possible so that the AI that results is as high quality as possible.

But even if you just take question answering as an evaluation for how smart an AI is, society should be able to trust those evaluations. Just as the civil service exams made people trust their interactions with the government, as AI becomes more tightly integrated into our economy and our society, our vetting of AI will become more important (Chapter 13).

Question answering is just one of many ways to secure trust, but it's analogous to how to ensure lawyers, doctors, and pilots are qualified for their jobs. So these tests should be unbiased, reliable, and the feedback from the tests should be transparent and understandable.

And we should do these things not just because we're trying to make society work like a well-oiled machine or because we want to be confident in our estimates of statistics about artificial intelligence. Just like Socrates, the questions we ask are trying to get to the truth, and scientific inquiry requires openness and a willingness to question the outcomes of a test.