data-ppf.github.io mar 12
lecture 8 of 14: birth and death of AI

chris wiggins + matt jones, Columbia

In the very brief “history of AI” given in the first ten minutes of AI course, they depict this time period in AI research as being obsessed with the idea of recreating human intelligence and laments how misguided this instinct was. To an extent, the McCorduck paper seems to support this depiction? I was actually kind of lost on this point. On the one hand, the researchers are very intent to draw out a general theory of human intelligence - the literal processes of human thought. On the other hand, Minsky doubts that human and artificial intelligence need to resemble each other. To me, this objection to the Logic Theorist seems to express - or at least leave room for - the approach that would dominate AI after its winter: studying the structure of info and problem solving processes independent of it is accomplished in humans. In which case, the approaches to AI weren’t as homogenous at this time as AI textbooks claim? But then on the next page (and many years after the conference), he admits he is much more interested in human psychology than AI. So where did he stand? Despite his preference for psych, does he think AI should be more automata studies than psychology? Am I misinterpreting his distinction btwn human and artificial intelligence?

Artificial intelligence includes:

- discovering what human intelligence *is* by attempting to create machine intelligence
- getting machines to perform activities usually thought to require intelligence

Why AI?

"AI can have two purposes. One is to use the power of computers to augment human thinking, just as we use motors to augment human or horse power. Robotics and expert systems are major branches of that. The other is to use a computer's artificial intelligence to understand how humans think. In a humanoid way. If you test your programs not merely by what they can accomplish, but how they accomplish it, they you're really doing cognitive science; you're using AI to understand the human mind." - Herb Simon

Histories of AI

Whence AI?

Overall, I found this week's readings about the history of AI to be fascinating, but McCorduck's was by far the most entertaining, memorable paper for me. McCorduck mentions that science is both a social and an epistemological issue, which I find very interesting. She argues that we find some level of comfort in finding people who share our beliefs and who accept the same evidence as us; we like to feel like we belong and thus organize groups and conferences with those with whom we share beliefs. The more people we find who share those beliefs, the more like we are to accept those beliefs as truth and as knowledge. This was how, she argues, the Dartmouth Conference was born. She also mentions the social pattern of clannishness born out of the conference. There appears to have been a lineage of AI researchers that came out of it, and the conference seems to have only accepted researchers from a certain set of institutions, ones that were and still are considered to be very prestigious, to participate in their panels and give presentations. This created an in-group within the field of AI research and a sort of "cliquishness," making it very difficult for researchers from other institutions to have their work funded or recognized.

✓ 1

Whence AI?

[redacted]

As professor Wiggins pointed out, there are many ways to read Turing's piece (a prediction vs a gameplan). It may be rather teleological to say he predicted what would happen over the coming years, but nonetheless, it is remarkable the foresight he had. His imitation test was particularly interesting to me, and the re-evaluation/broadening of definitions of intelligence and thinking that it necessitated. Turing articulates that we must be able to separate our understanding of things like thinking and intelligence from how these concepts function and appear in humans, because seeing if one could replicate a human is not the line of inquiry here. "Thinking" for a machine does not have to look like "thinking" for a human in order to be classified as thinking, which is something I don't think a lot of people understand about AI (in Turning's words, "may not machines carry out something which ought to be described as thinking but which is very different from what a man does?").

In McCorduck's article, it was interesting to see how interdisciplinary of a process developing and studying AI was (although that is not to say that it was diverse). It reminded me a bit of what we spoke about last week with Bayes and trading zones, and creating spaces in which people from different backgrounds and modes of thinking could still find a common plane of communication. (edited)

Whence AI?

[redacted]

Although the focus of these readings were on the Histories of AI , I am actually more intrigued on how gender disparities found in development of research in AI specifically in that of the Dartmouth conference, and how this disparity lends itself to the problem that Pamela Lays out in her paper, that of Professional Nepotism, and of individualism and clannish like mentalities of people sticking to "their people" when discussing work in their field, that of which McCarthy mentions as being one of the main problems of the conference. That people were stubborn about pursuing the ideas that "he had come with",..leading to a lack of real exchange of ideas. In response to this, I do think woman are more naturally attuned to sharing, giving care and nurturing other people then men are. I think this is more because of how they are socialized then how they naturally are. If there is one pattern I do see in the history of ideas in scientific computational thought is that men are obsessed with power and the hoarding of ideas to maintain their power, thus more competitive and less likely to speak to others to gain insight on opposite points of view. I wonder if the Dartmouth conference would have turned out differently had there been woman researchers invited to the field, not only to participate in the exchanging of ideas but to help create an organized structure* on how the conference members where to interact with each other and share their work throughout the conference. Like for example a mandated peer -review session after the end of every two weeks, between all members of the conference. Or a rule that mandated that people from different fields had to work in labs together so "psychologist would not just stick to speaking to psychologist " and so on. (edited)

✔ 1

Whence AI?

In McCorduck's article, I thought it was interesting to see the ways in which funding networks shape both access to Artificial Intelligence as a field, and the reception of new ideas, especially those proposed by "outsiders." Since ARPA seems to have different protocols for reviewing and supporting academic work than do other journals, for instance, especially the lack of a peer-review system, it seems reasonable that the direction of the field is determined more by ARPA as a DoD agency, then by what I think of as more "organic" intellectual development. But it's interesting to see that this funding structure of military-associated agencies supporting early-stage work in computing continues the trend that we saw in WWII and in Bletchley Park, where epistemological developments are motivated by the martial aims of the state, even if McCorduck doesn't discuss ARPA's interest in AI. I also thought it seems reasonable that this institutional structure, especially in its lack of systems like peer-review, could be responsible for the lack of AI practitioners "talking to each other" about their research, as McCarthy diagnoses the state of the field (133). Also on this structure of the field, I thought it was interesting to see how interdisciplinary AI appears as a field from these papers. While these contacts between disciplinary backgrounds seem helpful for challenging conventional ideas or introducing new models (like childhood learning models from developmental psychology), McCorduck's description of the Dartmouth conference also makes it clear that this disciplinary mixing introduces new challenges of translation and communication. This was particularly apparent in Minsky's assessment of Newell and Simon presenting their Logic Theorist and General Problem Solver "to psychologists," obscuring the potential cross-disciplinary applications of the underlying theories in those tools (126). On an unrelated note, I'm quite curious to know what Turing means by his "overwhelming statistical evidence" for ESP (Turing 453).

Whence AI?

A main idea that I was most interested in throughout the readings was the early representations of how scientists were hoping to design computers to have human emotions and capabilities. For example, in the reading about the Dartmouth conference "A Proposal For The Dartmouth Summer Research Project on Artificial Intelligence" it describes the expected result and says "Because of this preliminary internal study, these external experiments would appear to be rather clever, and the behavior would have to be regarded as rather 'imaginative.'" This description with words including imaginative, and later, where it describes how the individual "will have to do something which is unreasonable or unexpected as judged by the heritage of wisdom accumulated by the culture." This made me think about how these scientists' particular understanding of culture was informing the way that artificial intelligence was designed, and their culture is embedded in the creation of artificial intelligence. Throughout these readings, I thought about how any cultural bias seeps into any given scientific experiment based on its funding sources, the culture of the team, the culture of the environment where it is being designed, etc. I think the Dartmouth Conference is an example to me of the beginnings of both the revolutionary power of artificial intelligence, but also the beginnings of important questions about bias and who computers were being designed for in its earliest stages.

Turing, “Computing Machinery and Intelligence”

In *MIND*

Imitation Game

The new form of the problem can be described in terms of a game which we call the 'imitation game'. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'. The interrogator is allowed to put questions to A and B thus:

C: Will X please tell me the length of his or her hair?

Now suppose X is actually A, then A must answer. It is A's

28 433

object in the game to try and cause C to make the wrong identification. His answer might therefore be

'My hair is shingled, and the longest strands are about nine inches long.'

In order that tones of voice may not help the interrogator the answers should be written, or better still, typewritten. The ideal arrangement is to have a teleprinter communicating between the two rooms. Alternatively the question and answers can be repeated by an intermediary. The object of the game for the third player (B) is to help the interrogator. The best strategy for her is probably to give truthful answers. She can add such things as "I am the woman, don't listen to him!" to her answers, but it will avail nothing as the man can make similar remarks.

We now ask the question, 'What will happen when a machine takes the part of A in this game?' Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, 'Can machines think?'

Turing 433-4

More clearly

puters are in a sense equivalent.

We may now consider again the point raised at the end of §3. It was suggested tentatively that the question, 'Can machines think?' should be replaced by 'Are there imaginable digital computers which would do well in the imitation game?' If we wish we can make this superficially more general and ask

Shifting the question

[redacted]

Following up a bit on what @Sagar Lal posted, I also find the shift of the fundamental question of AI from “Can machines think?” to the question of whether a machine can imitate any aspect of learning/intelligence to the point that it would be called intelligent fascinating. I’m not very convinced by Turing’s claim that the question “Can machines think?” is too undefined to actually ask - he develops a perfectly usable working definition of machines just a bit later in the paper anyway, and I’m sure a psychologist could eventually come up with a good operational definition of what “thinking” is. In a way, the shifting of this question makes AI’s task harder: a machine must not only be able to appear as if it is thinking, but to appear to have all aspects of human mental ability, essentially having to be able to replace a human to be considered a thinking machine. I use the term “replace” here because Turing’s imitation game makes pretty clear that the ultimate goal of AI is to be able to replace a human with equivalent machines: Lighthill picks up on this by pointing out that Category A of AI is basically improving the ability of machines to do human tasks. Even the traditional test of AI (playing chess) puts as its goal the replacement of a human with a computer at a human game. This doesn’t necessarily mean we should be terrified of AI, but it does mean we should be a bit skeptical about AI proponents’ claims that AI are made to work “together” with people: as we can see, from the start thinking machines were designed with the ideal of being able to take a human’s spot.

✓ 1

Objection!

a very poor showing. He would be given away at once by slowness and inaccuracy in arithmetic. May not machines carry out something which ought to be described as thinking but which is very different from what a man does ? This objection is a very strong one, but at least we can say that if, nevertheless, a machine can be constructed to play the imitation game satisfactorily, we need not be troubled by this objection.

Presuppositions on intelligence

[redacted]

What I found most amusing is that in looking into what makes a computer “thinking”, Turing asks what it means to be alive and to be human, which is a pretty old question. I really like that he says critiques of machine learning claim that computers can’t “fall in love, enjoy strawberries and cream” (448). Do we need machines to do this? His answer is also rather interesting, that people just claim this because they have never seen a machine do these things before. Which is true, but it sounds like the reverse of what we have been talking about in class about certainty and the sun coming up. We have always seen the sun come up in the past, and we have never seen a machine enjoy strawberries and cream. But the claim about the sun coming up seems more valid than a claim that no machine ever could enjoy strawberries and cream. Except for that the idea of enjoyment ties into our “expert knowledge” and our preconceived ideas about the probability of a machine enjoying something, which comes not only from what we know about machines but also from what we know about enjoying things, namely that we see it as a very human activity. While his particular example seems far-fetched, it suggests that things that are less far fetched but still seem impossible for machines to do actually come from our experience and notions about what machines are likely to do. (edited)

✓ 1

Presuppositions on intelligence

The question and answer method seems to be suitable for introducing almost any one of the fields of human endeavour that we wish to include. We do not wish to penalise the machine for its inability to shine in beauty competitions, nor to penalise a man for losing in a race against an aeroplane. The conditions of our game make these disabilities irrelevant. The 'witnesses' can brag, if they consider it advisable, as much as they please about their charms, strength or heroism, but the interrogator cannot demand practical demonstrations.

Turing 435

Arguments against

(5) *Arguments from Various Disabilities.* These arguments take the form, "I grant you that you can make machines do all the things you have mentioned but you will never be able to make one to do X". Numerous features X are suggested in this connexion. I offer a selection:

Be kind, resourceful, beautiful, friendly (p. 448), have initiative, have a sense of humour, tell right from wrong, make mistakes (p. 448), fall in love, enjoy strawberries and cream (p. 448), make some one fall in love with it, learn from experience (pp. 456 f.), use words properly, be the subject of its own thought (p. 449), have as much diversity of behaviour as a man, do something really new (p. 450). (Some of these disabilities are given special consideration as indicated by the page numbers.)

Turing 447

What needed?

that moment. In this sort of sense a machine undoubtedly can be its own subject matter. It may be used to help in making up its own programmes, or to predict the effect of alterations in its own structure. By observing the results of its own behaviour it can modify its own programmes so as to achieve some purpose more effectively. These are possibilities of the near future, rather than Utopian dreams.

The criticism that a machine cannot have much diversity of behaviour is just a way of saying that it cannot have much storage capacity. Until fairly recently a storage capacity of even a thousand digits was very rare.

Turing 449

Your turn

- Pick one of the critiques
- Reconstruct the positions
  - Turing attributes
  - Turing's reply

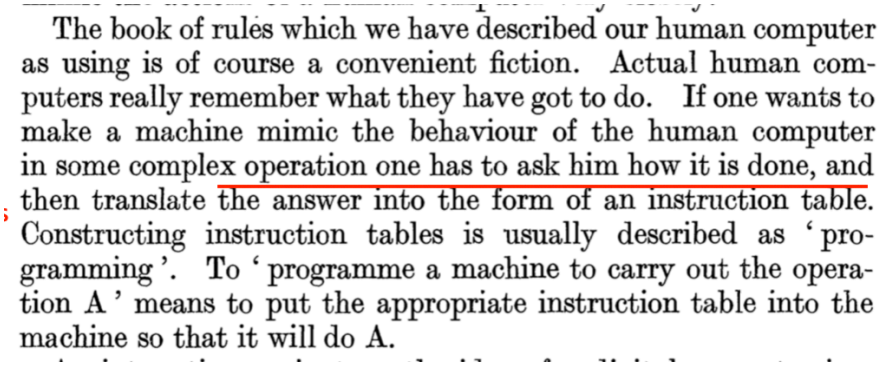
- What features might machines have to overcome the objection?
- How does Turing refute conception of the machines?

Early Strategies

Programming

J. McCarthy, M. L. Minsky, N. Rochester, and C.E. Shannon. August 31, 1955.

"We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it."



The book of rules which we have described our human computer as using is of course a convenient fiction. Actual human computers really remember what they have got to do. If one wants to make a machine mimic the behaviour of the human computer in some complex operation one has to ask him how it is done, and then translate the answer into the form of an instruction table. Constructing instruction tables is usually described as 'programming'. To 'programme a machine to carry out the operation A' means to put the appropriate instruction table into the machine so that it will do A.

How Turing go beyond this model?

Explicit rules from computers

Critiques

Lighthill report

- The Category B research work on problem solving in these abstract play situations has produced many ingenious and interesting programs. A fair description of the success of these programs seems to be that they are effective when and only when the programming has taken into **account a really substantial quantity of human knowledge about the particular problem domain**. Just as in category A, the pure mathematical logic methods suffer defeat at the hands of the combinatorial explosion, and have to be replaced by heuristic methods. Some very interesting researches have been carried out to develop general problem-solving programs, and such work can be of research interest to psychologists, but the performance of these programs on actual problems has always been disappointing. Students of all this work have generally concluded that it is unrealistic to expect highly generalised systems that can handle a large knowledge base effectively in a learning or self-organising mode to be developed in the 20th century. (13 of our pdf)

Aaron Plasek

Threat of Automation