Can Computers Think?
Wading Knee-deep into the Sea of Artificial Intelligence
David McKinley
OLLI Presentation – Fall 2024

On-line resources

There is an ocean of material available on-line about artificial intelligence, and trying to wade in just knee-deep is a challenge, as any one resource might suddenly drop you off an underwater cliff. My goal has been to understand the technology well enough that I felt like I could at least see how it was possible that it could perform the amazing feats I can observe, without necessarily being able to build my own AI application from scratch. Sort of like having a magician explain the "trick" to some amazing illusion, while still realizing that I do not have the skill to pull off such an illusion myself.

So, with that goal in mind, here are some of the resources I found most useful for gaining that level of understanding. Most of them are YouTube videos, because I found most written descriptions to be either too basic or "hand-wavy" to answer the questions I had, or too full of dense math to be enjoyable to study. Many of the videos I found online fall into one or the other of those camps, as well, but here are some that I found that were particularly helpful for reaching my goal.

3Blue1Brown YouTube series on Neural Networks
https://www.youtube.com/playlist?list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi

A series of videos by Grant Sanderson that soars above all the others I've come across, hitting right in the sweet spot for me, at least, explaining the nitty-gritty of "how it works" in a way that makes sense intuitively, without sacrificing accuracy or thoroughness. The first four of these videos cover the basics of Neural Networks, and I stole liberally from them in putting together that part of the first session for this course. The last three currently available cover some of the material from the second session – in particular, about Transformers. The seventh video ends with a teaser for an upcoming chapter 8, but there were 5 months between the appearance of chapters 6 and 7, and in that teaser he warns that he is busy with other projects, so it might be a while.

I strongly recommend watching these if you were intrigued by my presentation, but still just didn't quite feel like you "got it." These videos do a better job with all the animation used than I think I was able to do in the confines of a live lecture, plus you can watch them two or three times until it really does start to make sense, if you want. I've watched them more times than I can count, by now, and I still pick up a bit more from time to time.

How Neural Networks Learned to Talk – ChatGPT: A 30 year history
https://www.youtube.com/watch?v=OFS90-FX6pg

An interesting survey of the step-by-step research that led to where we are today with large language models, and including a glimpse into the philosophical debates about what it all means, which I tried to hint at in the second session.

What is Word2Vec? A Simple Explanation
https://www.youtube.com/watch?v=hQwFeIupNP0
A pretty good high-level description about word embedding, and Word2Vec. About 18 minutes.

Lecture 2 | Word Vector Representations: word2vec
https://www.youtube.com/watch?v=ERibwqs9p38
For the real enthusiast, this is the real deal. An hour+ lecture by a Stanford Engineering professor on Word2Vec, going through setting up a Skip-Gram example neural network in detail, working out all the calculus along the way. The hand-written diagrams used in the second session describing Word2Vec were screen-shots from this lecture. Even if you cannot follow the math, it is a bit interesting to get a feel for the sort of math that comes up in the real world of neural network models.

How Stable Diffusion Works
https://www.youtube.com/watch?v=sFztPP9qPRc
In the class, I showed some fun examples of image generation from text descriptions, but did not really get into much detail on how that really works, opting instead to focus just on language processing. In part, this is because I believe that, as discussed in the previously linked video about the history of ChatGPT, this is ultimately a more significant development on the path toward general artificial intelligence. But, it is also because I don't believe I really understand the technology of image generation well enough to be comfortable trying to explain it.

This video, by a young Australian who goes by "Gonkee" on YouTube, is the best I have found to date in its attempt to explain some of the big breakthroughs in image generation, called diffusion models, though I have not been able to quite "get it" yet. (e.g., how an image is generated starting from from random noise, if that's really what is happening.) I don't actually know how much of an authority he is, but it sounds like he knows what he's talking about, even if I haven't been able to follow his explanations all the way through. If someone else watches it, and thinks you can explain it to me, please contact me! I'd love to understand this particular technology better.

Toward the end of the video, he delves a bit into language models, and describes self-attention in a slightly different way from Grant Sanderson, which might make help make the "query/key/value" concepts in transformer models a bit clearer.

Alan Turing's AI Paper – Computing Machinery and Intelligence
https://courses.cs.umbc.edu/471/papers/turing.pdf
Alan Turing's 1950 paper describing the Turing Test and building computing machines with the capability to learn. Quite readable.

Attention is All You Need
https://arxiv.org/pdf/1706.03762
The groundbreaking 2017 Google paper that introduced the Transformer Model. Quite inscrutable.

Towards a Standard for Identifying and Managing Bias in Artificial Intelligence
https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf
A government paper from the National Institute of Standards and Technology (part of the US Department of Commerce) which I personally find troubling, though I don't have any specific arguments against its content. On the other hand, it is 86 pages long, and I'll admit to only having scanned some sections of it.

Arguments about Free Will
https://www.bethinking.org/human-life/the-libet-experiment-and-its-implications-for-conscious-will
https://mindmatters.ai/2019/09/was-famous-old-evidence-against-free-will-just-debunked/
I really only scratched the surface of philosophical arguments for or against the existence of free will in the second session. Reading only a little into the subject, you will almost certainly come across references to the "Libet Experiments" which supposedly settled the question in the 1980s. The first link above is to an article that describes the Libet Experiments in some detail, and discusses critiques of the experiments and conclusions that have been drawn from them. The second link is a lighter article on the same subject, and includes a number of links to other articles available on the topic.

Beyond these online sources (and many more easily found via search engines), I found two short books approachable and interesting to read. Both are available from Amazon, or if you'd like to borrow my copy, let me know.

Free Will, by Mark Balaguer, published in 2014 by MIT Press.
A book, mainly countering arguments against the existence of free will (i.e., the author believes that arguments for the proposition that free will is only an illusion are not convincing), it nonetheless provides a good overview of the nature of the philosophical and neurological arguments.

The Illusion of Determinism, by Edwin A. Locke, published in 2017 by the University of Maryland
A book in which the author believes he is making a strong argument against determinism, in favor of the existence of free will. (I remained unconvinced.)

The future of AI – What is an AI anyway?
https://www.youtube.com/watch?v=KKNCiRWd_j0
If you didn't like my speculations on the future of AI (or if you did, and want to hear another, potentially more informed opinion), this recent TED talk by Mustafa Suleyman, the AI CEO at Microsoft, is somewhat compelling.

"You Don't Understand AI Until you Watch This"
https://www.youtube.com/watch?v=1aM1KYvl4Dw
This is certainly not my favorite video, and I put his title in quotes, because I'm not sure I agree with his claim, or the implication that you will understand AI (very well) after you watch it, either. But, for a 37 minute video, it covers a lot of interesting ground, and is not egregiously bad, though a lot is quite over-simplified and hand-wavy in my opinion. I provided the link, with a tepid endorsement, because I do think he makes some interesting arguments toward the end about some of the big issues percolating around AI these days, though I think he is a bit overly definitive in stating his opinions on them as fact.