Note: This text is automatically translated (Google translate) from my original Danish article “AI – og den forunderlige ting kaldet menneskelig sund fornuft”.

Ingeniøren, November 2021.

Automatically Google translated text:

AI and that marvelous thing called common sense.

It rings true when you hear someone tell you, that AI still does not have anything resembling ordinary, human-like common sense. And that AI therefore in reality can’t quite do as much as some people think. And certainly only within narrow, clearly demarcated areas. But what is common sense really, and why don't we just make some AI with built-in common sense?

AI has by now survived several so-called AI winters, periods of low activity and cuts. Periods where every has been tired of the wild predictions of over-optimistic AI researchers that could not be realized. The first AI winter started in the 1970s, where after enormous progress from the early days of computers in the 1950s, it was difficult to move on, out into the real world, from the small "toy problems" one had so far had success with. The real world had simply turned out to be too complex for the fragile AI techniques of the day and the limited computing power of computers back then.
The 1980s were again a heyday period for AI, though not under the now somewhat disgraced name AI, but under terms like "expert systems" or "knowledge-based systems". Within limited, narrow areas, one could create systems that could solve tasks on an equal footing with human experts. One of the better-known systems from the period was MYCIN, that could help doctors with advice on blood infections. As one began to monetize such systems, AI slowly began to emerge from the darkness of winter.
But even the "expert systems" were too fragile, too expensive to maintain, and gave strange answers if asked about something outside their narrow area of expertise. Which is why we, from the late 1980s, again moved into an AI winter. Where for the next many years it was better to say that one worked with things like data mining, data analysis or speech recognition than to talk about AI, which again had proved not to quite to be able to live up to expectations.

Later, better computer hardware and huge amounts of data collected on the Internet again changed the situation. A neural net, Alexnet, designed by Alex Krizhevsky in collaboration with Ilya Sutskever and George Hinton (all from the University of Toronto), won in 2012 quite superiorly a competition to classify images into 1000 categories. Something of a comeback for neural networks, where Marvin Minsky and Seymour Papert, in 1969, had done their part to puncture most dreams about what neural networks could (By showing that a neural network in a layer, a perceptron, could not even solve very simple problems, such as the so-called XOR problem). And when it was finally understood that neural networks with several layers could easily solve these simple problems, and much more, well then there was still a big leap forward to our "deep learning" time, where neural networks with many layers, trained on large amounts of data can suddenly solve problems that have hitherto been completely unsolvable.

We are now in a time where deep learning does not just give us image recognition and self-driving cars. But can also be used to find irregularities in financial transactions, detect problems in health data, to translate texts from one language to another, text classification, and much more. The last AI winter is now only a distant memory.
And we are not getting any new AI winters just yet. Everyone in the IT industry will have plenty to do, for many years, taking advantage of deep learning techniques in the many niches where there are opportunities, but which you just have not yet explored. The question is more "how far can this go". Are we moving towards human-like intelligence, or will deep learning also soon reach a plateau where one faces obvious problems and from which it is difficult to make new progress? With less resonance for over-optimistic bids for the future, less interest, and in the long run perhaps fewer investments and public funding. Which, in the end, could be the beginning of a new AI winter.

Indeed, what are the interesting AI problems right now? Sorting numbers has been going well since 1959, when an algorithm, Quicksort, was invented that could do it quite quickly. Playing different board games also goes quite well. IBM's Chess Computer Deep Blue won over the then world champion in chess, Garry Kasparov, in 1996, and since then computers have been better at playing chess than humans. The game Go was a somewhat more difficult task for the computers, but still in 2016 the company DeepMind, with the program AlphaGo, managed to beat Lee Sedol, one of the world's absolute best Go players. Recognizing faces in pictures has been quite easy for the last 10 years. Relatively sensible systems for translations between different languages have been owned by everyone for the last 10 years. Just as good progress has also been made with the driverless cars etc. It is only when you ask the computers to interpret what is actually going on in a picture, or read a text, and answer questions about what they have read, that the problems begin to arise.

So it was not that surprising that it gave rise to quite a commotion, when 1 year ago in the English newspaper The Guardian one could read an article with the headline "A robot wrote this entire article. Are you scared yet, human?". Apparently, the american AI company OpenAI had succeeded in getting their deep learning program GPT-3 (Generative Pre-trained Transformer) to produce articles on a human-like level. Something that at least the Guardian thinks should have a headline that could be communicated directly to our Amygdala fear center. Without completely giving time for the more rational centers of the brain, in our prefrontal cortex, to be able to figure out what was up and down in the story.
Of course, we have had chatbots that we could talk to talk to in natural language for a long time. At MIT, Joseph Weizenbaum created the ELIZA program in the 1960s who, using simple rules and searching for keywords in sentences, often had the good fortune to change a little in the user's own sentence so that it could be sent back in the form of a question. All in all, a kind of parody of a visit to an extremely annoying psychotherapist (Where questions like "Can you X?" Are answered in template form with answers like "Do not you think I can X?", Etc.). Weizenbaum himself was amazed at how few simple rules were actually needed to make, at least some, people believe that the program actually occasionally understood a bit of what was being talked about.

More than 50 years later, of course, computers can do much more. The largest GPT-3 (neural network) model has 175 billion parameters trained to make "next word prediction" . Where an input sequence of words is given, a new sequence of words is produced[13] . Based on what the program thinks is most likely after it has been trained on several large datasets, including "The Common Crawls" Many petabytes of data, collected from billions of web pages on the Internet. The cost of

training the GPT-3 model is believed to have been in the order of millions of dollars and even larger systems have already arrived.

Many of the users who have had the opportunity to try out the GPT-3 system, has been overly enthusiastic. And have not talked about "human-like text generation" on the basis of "statistical correlations between word sequences"[19] , but about exciting conversations about "creativity" , "the geopolitical implications of AI" , "the meaning of life" and more.
More problematic has been that GPT-3 has also had an annoying ability to come up with more or less racist and sexist opinions. It all seem to be convinced that Muslims in general are violent and that problems in the healthcare sector can be solved by patients taking their own lives etc. In addition, the creators of GPT-3 themselves have stated that it is difficult for GPT-3 to stay focused in a conversation, that GPT-3 often contradicts itself, and does not always seem to know how the physical world actually works.
All in all, it does of course sound kind of bad. But hardly completely surprising if you think a little about what kind of result you would get if you trained a system on data from social media sites and alike.

Nor can one completely deny that GPT-3 is sort of human-like intelligence? Haven't we all bumped into people repeating dubious things they've read on the internet. People that seem to have difficulties keeping focus in a conversation, and certainly has difficulty understanding complex issues? Nor can we fully claim that we have never before heard any of the politically incorrect things GPT-3 can say.
But if we enter a future, where GPT-3-like systems themselves begin to post on social media, respond to our private emails, help dystopian political parties with inquiries from citizens, interfere in our Google searches and more. Then you can hardly call it "Responsible AI", if you have not thought a bit about what kind of data you have used to train the systems with.
And, by the way, of course, here that the story, in 2020, about Timnit Gebru's exit from Google's "Ethical AI Team" becomes particularly worrying. Did she, a black woman born in Ethiopia, find it easier to see, and understand, the many potential problems that can lurk in large datasets than the many men working on the systems that use this data? With consequent controversy and conflict over what the way forward should be. Or was it more a conflict between an ideal, but impractical, approach, versus a more pragmatic approach to solving problems in the best possible way here and now. Or something completely different?
If nothing else, with Gebru's exit from Google, we again got attention to the problems that can lie in waiting when training on large data sets. If one had forgotten that Google Photos AI in 2015 could come up with classifying blacks as "gorillas", repeated this year, where Facebook AI classified blacks in videos as "primates".

Of course, one could imagine rules to help the system out of the problems again. Eg. by training the system to recognize unfortunate things in a conversation. So the system would immediately change topic if a conversation is moving in an unfortunate direction[31] . But without any real understanding of the system, you can easily get the idea that every time you get around a problem, the system probably just steers itself into the next.
Responsible parents would not be in doubt if it was about children. They would be sent to school. A good school. Where facts are checked, and where teachers who comment on things like gender, age, ethnicity etc. do so in a way that parents can agree with. Responsible parents would probably then demand that the children be taught something about critical thinking, and be given clear routines

and wisdom, about history, mathematics, the legal community, morals and much more, which can help them to govern themselves and the society they live in.

But it's starting to get complicated. Where people suddenly know much more than what is being said directly in a conversation.
In addition to ploughing through social media, should out poor computers also be forced to watch all the soap operas ever shown on TV? To be able to figure out that if a man says "I'm leaving you", and a woman then answers by asking "who is she", then this woman is probably quite angry?
The american cognition-researcher Gary Marcus, is skeptical, and does not think that a "simple trick", or two, plus a lot of data, is enough to provide human-like intelligence. The brain can think abstractly, which is not the same as feeling something, which in turn is not the same as remembering something. The brain can understand how things are connected, and can keep track of what it knows about the world, so we can move around the world, and notice if things change.

Robots with human-like intelligence must similarly have internal models of what they know about the world. They must be able to access knowledge about the real world so that they can figure out that people in Poland mostly speak Polish, not Chinese. They must be able to reason. So they understand that when milk is not toxic, then people do not die from drinking milk. And they must certainly also have an understanding of how the physical world works here on Earth. When we board a train, we humans are surprised if a wormhole suddenly opens up in the space-time structure, which instantly brings us, and the train, on the other side of the Earth.
According to Marcus, correlations between word sequences are therefore far from enough to give us human-like intelligence. At least as long as these systems do not also have built-in modules that can reason. As well as modules that have knowledge of our world, our values and how the physical world works. If one does not want to go so far as to say that GPT-3 is completely brain dead, then one must at least admit that such systems still have major shortcomings, and often make rather foolish mistakes. Issues of cardiac medicine and counseling of suicidal people are still not something we are quite ready to let the chat systems take care of.
On the other hand, it is also clear that the systems have become much better lately. So in an accurate evaluation of how good such systems are, one must understand what type of model failed, and under what circumstances. How far is the tested model from a " *state of the art* " model? Were they tested to demonstrate how intelligent the model is, or to find as many foolish mistakes as possible?

And when talking about systems with human-like *common sense*, of course, one must also be precise with what one really means. In chess, Deepminds has AlphaZero[1] system has since 2017 been the best in the world. If you now asked AlphaZero if humans playing chess show "common sense", the answer would be no, as AlphaZero has only experienced that people's chess moves lead directly to loss.
Computers are also at the forefront when it comes to not believing in anything you hear. In any case, computers are better at detecting "fake news" on news sites than humans. And the computers are also becoming quite sensible when it comes to driving a car etc. On the whole, computers have an advantage when it comes to training on many examples, within a limited area.
Israeli psychologist Daniel Kahneman talks about two ways of thinking, "fast and slow" . Where computers now using deep learning, in many areas, are becoming quite good for the "thinking fast" part, with quick automatic answers. While things are not going so well for the computers with the "thinking slow" part, where a general overview and logical thinking is required. GPT-3 is already in difficulty when requesting a summary of a text (See DataTech, 1 Oct. 2021). And computers do not

(yet) have the necessary overview to be able to follow a story, and answer questions about the story. And there is not at all the integration between the two systems, fast and slow, that you see in humans.

So what's the next step? Facebook has said that they are working on systems that understand the world in the same way as humans. Ie. systems that are trained using video, recorded as seen through the eyes of a person . Which of course will be useful in Facebook's new Metaverse, where all (Facebook) users are equipped with Ray-Ban smart glasses that can record everything you see and hear. Besides, the glasses, in this future universe, act as a computer screen that can show users everything they previously could only see on their phone. Including telling someone with big arrows, in your field of vision, where to go when you are out walking. If you have forgotten your keys, work on the research project Ego4D (See DataTech, 28 Oct 2021) on a system that can go back in all that is recorded and find out where the keys are. Yes, so far so good. At least as long as you do not have concerns about storing private pictures on Facebook's servers. But even more futuristic, this project also suggests that they intend to work with sequences of video, in the same way as others work with sequences of text, for forecasting. So that the system can begin to guess what usually follows after a given sequence. E.g. when you cook potatoes, most people come salt in the pot after the water has come in. So, if Facebook AI ever manages to train systems based on such video sequences. Yes, then the Facebook glasses can draw attention to problems, such as if one e.g. have forgotten the salt, and future robots know what a good "next move" is in similar situations.

But the problems of lack of common sense have also become easier to spot along the way.
If you imagine e.g. that all shopping centers must be equipped with Pepper-like robots to give directions and make the customers feel at home. Yes, then the robots must use computer vision to identify what is happening in front of them? Which can then be given as input to GPT-3, who can talk to the customer. Where, from a "Responsible AI" consideration, one must hope that the vision system does not have the same unfortunate tendency as the Facebook AI system, and begin to classify some customers as "primates". Not to mention what can happen if that kind of information is passed on to GPT-3.
Hurt feelings are bad enough, of course. But if one now imagines that the robot has to move around in the shopping center, based on training with 1-person video recorded by millions of Metaverse users. Then at least some of us start to think that Gary Marcus probably has a point that sequences are not always enough to make common sense. Although many others have sat on a bench, it is not a good idea if there is a sign saying “recently painted” on it. You have to know something about, among other things, logic, morality and physics before you can act with common sense. And wisdom, btw. Is action you will be happy with, also in the long run.
Indeed, yes, common sense is still something we can only wish that we had more of. Even the robots might agree.

October 2021.
Simon Laub

Previous article, see:
https://www.version2.dk/blog/roede-linjer-big-data-internettets-aendringer-vores-samfund-oest-vest-1092736

---

[1] https://www.forbes.com/sites/cognitiveworld/2019/10/20/are-we-heading-for-another-ai-winter-soon/?sh=28913f5556d6

[2] https://www.britica.com/technology/MYCIN
[3] https://spectrum.ieee.org/history-of-ai
[4] https://papers.nips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
[5] https://www.bbc.com/news/technology-51064369
[6] “A brief history of AI”, Michael Wooldridge, Flatiron Books, 2020. https://www.amazon.com/Brief-History-Artificial-Intelligence-Where/dp/1250770742
[7] https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3
[8] https://theconversation.com/the-emotion-centre-is-the-oldest-part-of-the-human-brain-why-is-mood-so-important-63324
[9] https://web.njit.edu/~ronkowit/eliza.html
[10] https://news.harvard.edu/gazette/story/2012/09/alan-turing-at-100/
[11] https://arxiv.org/pdf/2005.14165.pdf
[12] https://openai.com/blog/better-language-models/
[13] https://www.nature.com/articles/s42256-021-00395-y
[14] https://commoncrawl.org/
[15] https://commoncrawl.org/2018/07/3-25-billion-pages-crawled-in-july-2018/
[16] https://news.mit.edu/2020/shrinking-deep-learning-carbon-footprint-0807
[17] https://www.microsoft.com/en-us/research/blog/using-deepspeed-and-megatron-to-train-megatron-turing-nlg-530b-the-worlds-largest-and-most -powerful-generative-language-model /
[18] https://openai.com/blog/openai-api/
[19] https://www.cnbc.com/2020/07/23/openai-gpt3-explainer.html
[20] https://www.nytimes.com/2020/11/24/science/artificial-intelligence-ai-gpt3.html
[21] https://twitter.com/solarperimeter/status/1442466911913054215/photo/1
[22] https://twitter.com/Siddharth87/status/1283920116007092224/photo/1
[23] https://thenextweb.com/news/gpt-3s-bigotry-is-exactly-why-devs-shouldnt-use-the-internet-to-train-ai
[24] https://twitter.com/an_open_mind/status/1284487376312709120
[25] https://www.nature.com/articles/s42256-021-00359-2.epdf
[26] https://www.nabla.com/blog/gpt-3/
[27] Limitations, https://arxiv.org/pdf/2005.14165.pdf
[28] https://www.wired.com/story/google-timnit-gebru-ai-what-really-happened/
[29] https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai
[30] https://eu.usatoday.com/story/tech/2021/09/03/facebook-video-black-men-primates-apology/5721948001/
[31] https://www.technologyreview.com/2020/10/23/1011116/chatbot-gpt3-openai-facebook-google-safety-fix-racist-sexist-language-ai/
[32] https://cacm.acm.org/magazines/2021/1/249452-insights-for-ai-from-the-human-mind/fulltext?mobile=false
[33] https://thegradient.pub/has-ai-found-a-new-foundation/
[34] https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp/
[35] https://cims.nyu.edu/~sbowman/bowman2021hype.pdf
[36] https://deepmind.com/blog/article/alphazero-shedding-new-light-grand-games-chess-shogi-and-go
[37] https://news.umich.edu/fake-news-detector-algorithm-works-better-than-a-human/
[38] "Thinking, Fast and Slow", Daniel Kahneman. Farrar, Straus and Giroux, 2011. https://www.amazon.com/Thinking-Fast-Slow-Daniel-Kahneman/dp/0374275637/
[39] https://pro.ing.dk/datatech/artikel/gpt-3-skriver-tvivlsomme-bogreferater-og-amazon-chauffoerer-klager-over-daarlig-ai
[40] https://about.fb.com/news/2021/10/teaching-ai-to-view-the-world-through-your-eyes/
[41] https://www.theguardian.com/technology/2021/sep/15/techscape-smart-glasses-facebook
[42] https://pro.ing.dk/datatech/artikel/ego4d-facebook-wants-ai-find-your-keys-and-understand-your-conversations-15430
[43] https://www.bristol.ac.uk/news/2021/october/ego4d.html