

On The Risks of Artificial Intelligence Advancement

Rose Awen Brindle `contactme@rose-brindle.dev`

March 2019

Artificial Intelligence is a quickly developing technology, and it is already becoming a key part of our future. However, it is a highly contentious issue and as it becomes ever more prominent in our society, we can expect to be answering some very hard questions relating to humanity's moral code and our beliefs about the meaning of intelligence and thought. This essay will address current concerns around artificial intelligence and consider whether it will be a net gain for humanity, or possibly bring about its eventual downfall.

What is Artificial Intelligence?

Artificial Intelligence, or AI, is defined as being a system “able to perform tasks normally requiring human intelligence” (Oxford University Press, 2019). In order to complete this, most software using Artificial Intelligence differs from usual software because the program is built to ‘learn’ from data, changing how it responds to inputs, whereas your word processor, for instance, completes pre-determined instructions when required.

Artificially intelligent systems can be divided into two separate groups, narrow and general intelligence. The difference comes from the broadness of this ability to ‘learn’. A narrow Artificial Intelligence can only complete one task or a small selection of similar tasks and improves that single task's execution, whilst Artificial General Intelligence (AGI) is a more accurate representation of human thinking because it can make decisions without prior data on the specific problem; it can reason. As of current writing – and probably for a long time after – we have not created a true Artificial General Intelligence, however Artificial Narrow Intelligence is used in many parts of modern society.

Artificial General Intelligence would be able to learn anything given enough time in the same a human could. This includes programming its own Artificial Intelligence software. The ideas, dangers and possibilities presented by this allow for the creation of Artificial Super-Intelligence, Artificial Intelligence with intelligence far beyond human minds. Experts are not in agreement as to the possibility of this, some 7.5 percent say 20-25 years, although most expect it to be created later, and around a quarter say never (MIT Technology Review, 2016). Although I will be considering Artificial General Intelligence, superintelligence and the singularity in my analysis, my focus will be on the effects of Artificial Intelligence in the less-distant future.

Current Artificial Intelligence (ANI) can be further split into types: Deep Learning, focusing on the replication of the way the human brain works using neurons; and Machine Learning, using graphs. Both types are ‘taught’ what they must know by giving them large data sets in which they must find correlations which they can then apply to new data. For instance, speech recognition software is given a very large collection of audio files (often in the thousands or tens of thousands) in order to find patterns. This training data can come in three forms: the data has the correct answers attached and the software (also called the ‘agent’) works out a correlation linking all the data of one correct answer; the training data is unlabelled, and the agent looks for patterns and sorts them into categories itself or finally the agent learns through interacting with its environment, changing its reaction based on rewards and punishments instead of using specific data sets. Each of these is useful in different use cases, as they each have their own advantages and disadvantages.

The Effects of AI

There are three main topics to consider when looking at the effects of Artificial Intelligence. the advancement of research and the collective knowledge of the human race is one of these. The largest effects on this would be seen after the hypothetical ‘singularity’, whilst currently, it is being used for a variety of very specific scenarios, such as the estimation of damage by possible forest fires (Toujani et al, 2018). Another key aspect is its effects on society, both in terms of dangers and benefits. These are the first things that come to many people's minds, topics

such as mass-unemployment and the collection of personal data (Brindle, 2018). Finally, Artificial Intelligence will begin to present more and more physical dangers and abilities as they are implemented into more devices, including driverless cars. These topics will be covered in the coming paragraphs.

Information and evidence for the points below has been gathered from many sources, mainly scientific papers. During the process of collecting sources, I created and sent out a survey (Brindle, 2018) to first-year college students (16-17 years old) in order to gain an insight into current public expectations and opinion in this field. The survey is, however, not a fully accurate representation of this due to the small sample size, although points can still be and have been, drawn from it.

Uses of AI

As the boundaries of artificial intelligence are being researched and pushed further and further, many specialists use in a wide range of fields are becoming more prominent. As more techniques are developed and implemented, many more applications will become available to businesses and governments. These will often have a profound impact on the productivity of businesses and the streamlining of organisations, as well as possibly decreasing mortality rates and damping the effects of our climate and weather systems on our lives. However, these changes may have a damaging impact on some individuals, forcing a challenging moral debate on which is more important, almost always with a lack of full information on the subjects.

Weather

One possible application of machine learning is in the prediction of weather systems, known as climate informatics, where huge amounts of historical data are used alongside, and to produce, climate models. The massive amount of data collected (around 215 Million data observations per day in the case of the United Kingdom's Met Office (Met Office, 2018)) is exactly what machine learning thrives upon. Although not currently a common part of a climatologists toolbox, Monteleoni et al (Monteleoni et al, 2008), authors of the 2011 paper on the tracking of climate models were able to use artificial intelligence algorithms to find weighted averages of climate models. Another specific case was the use of artificial intelligence to predict and estimate the damage caused by forest fires (Toujani et al, 2018), specifically, a case study in northern Tunisia, where the use of "829 forest fire records" allowed them to successfully predict the size of burned areas.

These uses are very unlikely to result in negative consequences, instead, the improvement of the accuracy of our current climate models would allow for a more accurate version of the weekly forecasts seen on television and websites, but also - and more impactful - it would likely improve the quality of long-distance weather reports. This would allow shipping companies to reduce running costs and fuel consumption by improving weather routing, where, in the case of maritime shipping "a fuel saving of 5% corresponds to a reduction in carbon emissions of almost 150,000 tons per year" (SMHI, 2014).

Education

Artificial Intelligence may also be useful in the field of teaching, where the individualisation of learning is a key priority. An analysis by the Local Government Authority found that (Education & Skills Funding Agency, 2018), by 2023, half of the councils within the united kingdom will not be able to meet the demand for secondary school places. One study (Edwards et al, 2018) in which experts created an artificial intelligence and a robot interface for teaching, three key abilities of the 'teacher' were designed including engaging in social interaction in the classroom and being able to select and employ a pedagogy.

With algorithms able to hone in on the issues each student is facing, as well as being able to take a load off the teacher's backs and free more time for more students, Artificial Intelligence would be able to provide a more personal education style. However, an increase in the use of technology for teaching could take away the teacher-student relationship which many studies e.g. (Kosir; Tement, 2013) have found to be an important part of learning and development. This would be caused by the implementation of the artificially intelligent teacher which had little to no ability for social interaction which may be able to be avoided in the future. Currently, college students would not be fully comfortable with having a 'robot teacher' though, with almost 70% of students indicating they would be uncomfortable in that situation.

Medical

Artificial Intelligence is already gaining a foothold within the medical fields, where it is used to recognise symptoms and support clinicians. In the United Kingdom’s National Health Service, a solution called ‘C the Signs’ works ‘recommending what investigations or referrals the patient may need’ (NHS England, 2017). In this case it is used to support the General Practitioner, however other examples such as Babylon Health’s artificial intelligence chat box, which uses natural language processing to examine a series of answers to a set of dynamic questions to give advice on symptoms (Babylon, 2019).

In both of these use cases, the use of artificial intelligence saves time and can improve the accuracy of medical responses; in such a field, this means saving lives. Babylon’s chatbot has been deployed to Rwanda, where users can access the advice from anywhere via their phones, rather than having to make a possibly many mile trek to reach the nearest doctor. This is one of the key benefits of such a solution, where users with smaller symptoms are able to leave doctors or GPs to focus on other priority medical conditions and the artificial Intelligence can refer the patients if required. However, the accuracy of these situations is often hard to test, and when it is possible, a decision has to be made where the cut off point will be. Do we require them to be better than our average doctor, or as accurate as our best ones?

Government and War

Of all the examples given here, this is the most controversial, and for good reason. An improvement in a countries ability to create killing weapons, without risking as many of its own people’s lives comes hand in hand with many technological advancements. Artificial Intelligence is no different, however, there are also many other use cases within a government, such as improving the response time of emergency services and assisting the interaction between government and the public. In these cases Artificial Intelligence is used within a limited interface, with no way of it interacting with the world outside of it, however, there is a growing concern over the use of Artificial Intelligence within robots as a military resource. Cummings (Cummings, 2017) theorises that the rapid advancement of commercial autonomous systems could normalize the acceptance of such systems in warfare. This possibility is already becoming a reality with the increasing use of UAVs (Unmanned Aerial Vehicles) by militaries such as the US Army (U.S Army, 2018), which could be easily moved over to Artificial Intelligence in the future.

The advancement of such systems will have a number of effects on both warfare and the public. In war, the use of UAVs would decrease the cost of lives, a key incentive for their use by governments. However, if artificial intelligence is applied to this, there is a possibility that targets may be mistaken by the Artificial Intelligence causing a loss of civilian lives. This is due to the way most machine learning algorithms make decisions, they calculate and find the action with the option most likely to reach the expected outcome, comparing its current situation to past examples. This means that it is possible for such a system to find incorrect patterns in the data it has been trained on. In many cases, patterns which seem obviously to be obviously not a causal pair to a human, without context can seem to be this. For example, Figure 1, although it is very unlikely that one causes the other, seem to be have a strong correlation.

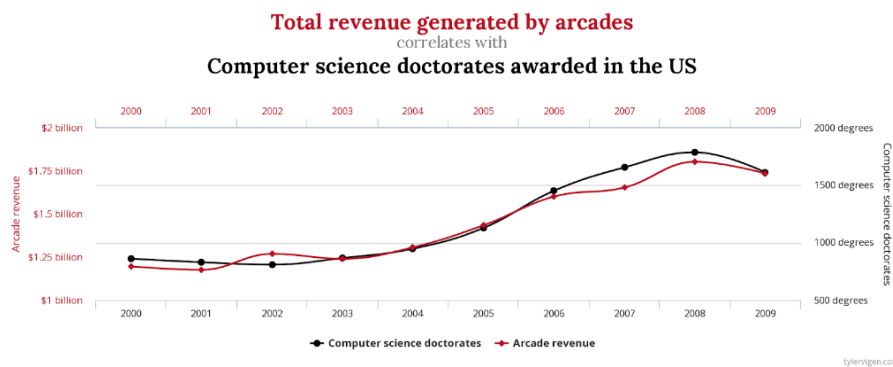


Figure 1: A graph from showing a seemingly high correlation between Total Revenue Generated by Arcades and the number of US Computer Science Doctorates Awarded (Vigen, 2018).

Another possibility is that once implemented they may become available for other, public, uses, such as search

and rescue, where they would be able to reduce search times (Eyerman et al, 2018) and cover impassable terrain. However UAVs are often expensive, and with almost all Search and Rescue teams, especially mountain rescue, working with limited funding directly from the government or local police, it may be a challenge to fit this new cost in.

Artificial General Intelligence

I have given this type of artificial intelligence its own section as many of the possibilities it presents for our future are uncertain, however they all have widespread effects, and therefore I felt they should be included in this analysis.

The Singularity

The creation of an Artificial Super-Intelligence through an Artificial General Intelligence would bring about a situation known as the Technological Singularity. This is the hypothesis that these super intelligent Artificial Intelligence would cause a massive exponential increase in human knowledge and events after that are therefore unpredictable. This is caused by the Artificial General Intelligence learning to improve itself and therefore, in time, it is possible the exponential growth of its intelligence (see diagram) would allow it to do and understand anything.

With this comes worries about how this will affect the way humanity lives: humans will no longer be needed to work jobs, causing a massive change in the way our society works or the Artificial General Intelligence may decide that humans are a part of the problem.

Repetition

One of the key use of both Artificial Intelligence and Artificial General Intelligence is to remove repetitive tasks, with more advanced Artificial Intelligence being able to complete tasks with more possibilities or checks. This simplification would cause a major shift in the spread of job types, with the removal of many low-pay, repetitive roles and an increase in the need for high-skill, specific roles.

Extinction

In media such as film and television, one of Artificial Intelligence's common castings is as the bringer of the downfall of the human race. In reality, it is unlikely that a narrow Artificial Intelligence could cause this, as many steps of the way it "thinks" are encoded by the creator, and so worst-case situations can be avoided. However, in the case of artificial general intelligence, it is possible that it may choose targets or actions which purposefully or inadvertently destroy us without our knowledge or control. The reasons for a Artificial Intelligence ending humanity are many, both as an accident or on purpose: for example, a diagram by Alexey Turchin (2015) shows 55 different possibilities, sorted in order of when they would happen in the timeline of Artificial Intelligence creation.

Many of the problems and dangers to humanity presented by researchers should, however should be solvable if possible eventualities are planned for, and systems put in place to avoid them. However to do this we must agree what the Artificial Intelligence may not do, a code of ethics. This is an extremely controversial problem, for example Boddington (pg 89, 2017) questions whether mass-unemployment caused by artificial intelligence is freedom or a catastrophe.

Economic Effects

Mass Unemployment

Of the 16 respondents to my survey, 12 of them (75%) cited mass-unemployment as a worrying possibility in the face of advancing Artificial Intelligence technologies (Brindle, 2019). This is likely because, in comparison to other options such as "Artificial Intelligence destroying / severely crippling the human race" and "Artificial Intelligence being hacked for personal data", it's effects would be much more widespread, whilst also seemingly being a likely possibility. Examples from the past can be seen clearly, such as the shift from a mainly agriculture based economy, with 50% of the US population working in the industry (Daly, 1981), to one based around manufacturing, leaving only 1.5 percent of the population employed within agriculture (U.S. Bureau of Labor,

2016), in the United States of America. However, speculation is rampant and some reports instead expect an overall increase in employment, such as PricewaterhouseCoopers' (2017) estimation that 7,176,000 jobs will be created in the United Kingdom versus 7,008,000 lost between 2017 and 2037.

The effects of mass-unemployment if it were to happen would be widespread, with industries from transportation to public administration expected to see more than 20 percent of jobs within them displaced (PwC, 2017) within the next 18 years. In areas of high unemployment, social issues and political instability often become more prominent, whilst individual well-being and health are also affected. With the possibility of these effects becoming national issues, whether or not more jobs will be created precautions will need to be put in place to ensure the effects of unemployment do not come crashing down on us.

Societal Effects

The Scramble for Data

With Artificial Intelligence becoming more prevalent, companies developing them are going to require more and more data to “teach” them. This presents a major issue. Where does this data come from? In many current uses, such as marketing, AI requires data on people or their actions in order to make predictions about other people. With similar uses for AI becoming more and more extensive, businesses will require an ever increasing amount of information on their target demographics, often collected through data-mining techniques such as classification and association (Brown, 2012). This forces organisations training AI systems to move closer and closer to the line between researching and privacy breaching. This already becoming a worry for many people, with more than half of the respondents to my survey citing the amount of data stored for AI as a concern.

Proposed Strategies for Development

From my research, I have found that Artificial Intelligence will have very few, if any, massively damaging impacts. Where concerns are valid, proper planning and coordination should allow us to continue using Artificial Intelligence as the extremely powerful, but safe technology it has been to date. Artificial General Intelligence will have little to no impact on our world in the foreseeable future, the technology will stay in the realms of science fiction for a long time and so its effects are hard to predict, but again, planning will mitigate any negative ones that could appear. The topics considered below are three areas that should be focused on in order to implement the needed planning to keep Artificial Intelligence a viable technology.

Understanding

One of the key things I have found from my research is that information, including webpages and news articles, regarding Artificial Intelligence is often very technical, to the point of leaving very few avenues for public accessibility. As mentioned earlier, in my survey, more than 80% of people responded saying they believe they had an average to low understanding of Artificial Intelligence. Whilst in another, wider survey by Bristows (2018, n=2103), 50 percent of respondents believed they had limited or worse knowledge of Artificial Intelligence, with only 10% claiming they “know alot” or “were experts in” Artificial Intelligence. This highlights an issue that, if left unfixed could cause a massive divide in opinion and understanding between Artificial Intelligence researchers and the general public and governments, making it highly probable that possible needs for legislation and oversight could simply go unnoticed.

For this reason, more articles and information should be written with the intent to inform the general public on the possibilities of our future in relation to Artificial Intelligence. In these cases, attempts should be made to avoid technical terms and jargon as well as detailing the effects the Artificial Intelligence would have on its intended field as a priority over how the Artificial Intelligence system works.

In the same vein, politicians from all countries should be informed on these effects, ready to respond as appropriate to any calls for Artificial Intelligence legislation as it becomes increasingly prominent in our society. The United Kingdom is already taking steps to achieve this, having already considered a report by its Select Committee for Artificial Intelligence (UK Government, 2018) considering whether the United Kingdom is “ready, willing and able?” to cope with the changes Artificial Intelligence presents, taking into account a wide range of professional opinion and knowledge.

Legislation

With the many applications of Artificial Intelligence, including the ones considered earlier, specific legislation will likely be required in order to combat the risks presented by this technology.

For example, the use of Artificial Intelligence in warfare would be extremely risky at our the current level of success in Artificial Intelligence creation, with the likelihood that civilians would be paying the price. The United States of America, the main user of UAVs currently, has a somewhat lax view of the use of Autonomous Killing machines. Policy currently requires target selecting without the human intervention to be certified as compliant within appropriate levels.

In many of the situations revolving around Artificial Intelligence, a decision needs to be made on who is responsible for its actions. As Artificial Intelligence is currently only able to make decisions based on programming by its creator, the creator has the ability to set limits or remove options and so is to blame for the decisions made by the Artificial Intelligence. However if Artificial Intelligence becomes able to choose its own path without its options being implemented by the owner, then it may be that one day the Artificial Intelligence will be to blame.

With these and many other situations in mind, Artificial Intelligence needs to become a much higher priority for legislation implementation than it currently is. Laws should, as always, take into account the valid fears of the public and the opinions and expectations of professionals.

Ethical Code

Many of the most worrying possibilities surrounding Artificial Intelligence, such as the dangers of Artificial General Intelligence and the effects of mass-unemployment, could be avoided by setting out rules all Artificial Intelligence must follow, a code of ethics. Boddington (2017) considers the procedures required for the organisation and enforcing of one. She lists 4 key requirements:

- **Transparency:** the decisions could affect millions if not billions of people, so openness concerning the decisions made is paramount
- **Communication:** participation in public discussion and with government bodies to address concerns that will inevitably arise
- **Revision and Critique:** Where technology improves or opinions change, procedures should be in place to facilitate revisions to the code following the other points
- **Timing:** careful thinking requires time, although in some cases changes must be made in response to errors and issues

With these in mind, it becomes a requirement that the code must be created by an organisation with the weight behind itself to enforce it, or instated by a business for its own use. In all these cases, the possibility for over-complication is extremely high. However, the simpler the requirements and the shorter the list, the less controversial the ideas presented are likely to be. This would also improve the ease at which programmers could implement the framework.

Conclusion

There is not the leeway to make major errors in relation to Artificial Intelligence as it becomes more powerful and more ingrained within our society. Discussions currently underway must become more prominent so as to take in to consideration the responses of a wide variety of people and cultures, whilst more discussions focussing on individual fields need to take place so that requirements are put in place before they are needed, not as a response. Actions need to be taken now, but if they are, we can expect to see Artificial Intelligence surpass many other technologies in its usefulness to humanity.

BIBLIOGRAPHY

Met Office (2018) Big data - more usable, more useful - Met Office. Available at: <https://www.metoffice.gov.uk/services/data-provision/big-data-drive> (Accessed: 28 February 2019).

Oxford University Press (2018) artificial intelligence | Definition of artificial intelligence in English by Oxford Dictionaries. Available at: https://en.oxforddictionaries.com/definition/artificial_intelligence (Accessed: 28 February 2019).

Certes Computing Ltd (2018) Types of Artificial Intelligence: A Detailed Guide. Available at: <https://certes.co.uk/types-of-artificial-intelligence-a-detailed-guide/> (Accessed: 28 February 2019).

MIT Technology Review (2016) No, the Experts Don't Think Superintelligent AI is a Threat to Humanity. Available at: <https://www.technologyreview.com/s/602410/no-the-experts-dont-think-superintelligent-ai-is-a-threat-to-humanity/> (Accessed: 28 February 2019).

Toujani, A., Achour, H. and Faiz, S. (2018) 'Estimating Forest Fire Losses Using Stochastic Approach: Case Study of the Kroumiria Mountains (Northwestern Tunisia)', *Applied Artificial Intelligence*. Taylor & Francis, 32(9-10), pp. 882-906. doi: 10.1080/08839514.2018.1514808

Brindle, R. A. (2018) 'Artificial Intelligence: Now and the Future - Survey'.

Monteleoni, C., Saroha, S. and Schmidt, G. (2008) 'Tracking Climate Models', pp. 2-3. doi: 10.1002/sam.10126.

SMHI (2009, Updated 2014) Weather Routing reduces carbon dioxide emissions. Available at: <https://www.smhi.se/en/news-archive/smhi-weather-routing-reduces-carbon-dioxide-emissions-1.8308> (Accessed: 28 February 2019).

Education & Skills Funding Agency (2018) 'School Capacity : academic year 2016 to 2017'.

Edwards, B. I. and Cheok, A. D. (2018) 'Why Not Robot Teachers: Artificial Intelligence for Addressing Teacher Shortage', *Applied Artificial Intelligence*. Taylor & Francis, 32(4), pp. 345-360. doi: 10.1080/08839514.2018.1464286.

Košir, K. and Tement, S. (2013) 'Teacher-student relationship and academic achievement: a cross-lagged longitudinal study on three different age groups', *European Journal of Psychology of Education*, 29(3), pp. 409-428. doi: 10.1007/s10212-013-0205-2.

NHS England (2017) C the Signs - How artificial intelligence (AI) is supporting referrals. Available at: <https://www.england.nhs.uk/cancer/case-studies/c-the-signs-how-artificial-intelligence-ai-is-supporting-referrals/> (Accessed: 28 February 2019).

Babylon (2019) Babylon Health AI. Available at: <https://www.babylonhealth.com/ai> (Accessed: 28 February 2019).

Cummings, M. L. (2017) 'Artificial Intelligence and the Future of Warfare'.

U.S Army (2018) Aircraft | goarmy.com. Available at: <https://www.goarmy.com/about/army-vehicles-and-equipment/army-helicopters-and-uavs.html> (Accessed: 28 February 2019).

Vigen, T. (2019) Spurious Correlations. Available at: <http://www.tylervigen.com/spurious-correlations> (Accessed: 28 February 2019).

Eyerman, J. et al. (2018) 'Drone Efficacy Study: Evaluating the Impact of Drones for Locating Lost Persons in Search'.

Turchin, A. (2015) Human Extinction Risks due to Artificial Intelligence Development - 55 ways we can be obliterated. Available at: <https://ieet.org/index.php/IEET2/more/turchin20150610> (Accessed: 4 March 2019).

Boddington, P. (2017) Towards a Code of Ethics for Artificial Intelligence. doi: 10.1007/978-3-319-60648-4.

U.S. Bureau of Labor (2016) Employment by major industry sector. Available at: <https://www.bls.gov/emp/tables/employment-by-major-industry-sector.htm> (Accessed: 5 March 2019).

Daly, P. A. (1981) 'Agricultural employment: has the decline ended?', *The Monthly Labor Review*.

PwC (2017) 'UK Economic outlook: What will be the next impact of AI and related technologies on jobs in the UK?', (July). Available at: <https://www.pwc.co.uk/economic-services/ukeyo/ukeyo-july18-full-report.pdf>.

Boddington, P. (2017) Towards a Code of Ethics for Artificial Intelligence. doi: 10.1007/978-3-319-60648-4.

Brown, M. (2012) Data mining techniques – IBM Developer. Available at: <https://developer.ibm.com/articles/ba-data-mining-techniques/> (Accessed: 28 March 2019).