AI and Society: Predicting scientific progress

James Maclaurin

Department of Philosophy • University of Otago

Paradigms and normal science

- Philosophers and historians of science see scientific progress as punctuated by scientific revolutions (Copernicus, Lavoisier, Darwin…). These changes in the way we do science are known as **paradigm changes**. What happens in between is **normal science**.
- This large scale pattern also appears at smaller scales if we want to describe the progress of e.g. Darwinian biology.
- The history of predicting science is full of examples of massive over and under-estimation in the progress and commercialisation of science. This is partly due to the fact that paradigm change is a very creative process and partly due to highly multifactorial and stochastic nature of scientific and technological change.
- The exponential growth often employed by futurologists, is more likely to be correlated with rates of progress in normal science that with the probability of an unexpected paradigm change.
- Scientific discovery, technological progress and commercial markets are often out of step with one another.
- Therefore, both positive and negative estimations of medium to long-term progress in science and technology are inherently unreliable.

Social effects on progress

- Rate of progress in normal science is strongly influenced by effort and investment (Google has created a self-driving car very quickly, where generations of backyard mechanics have failed to create a flying car).
- Major technological change often depends on a lineage of commercially successful pre-existing technologies. So smart phones are the offspring of typewriters, radios, telephones, adding machines, positioning systems etc.
- Effort and investment is proportional to risk (e.g. danger to health) and reward (e.g. the size of a potential commercial markets).
  - We should keep in mind that Risk = (cost - benefit) x probability
  - We are right to put a lot of time and effort into very unlikely but extremely harmful threats (e.g. asteroid detection). So it might still be appropriate to put significant effort into understanding the dangers and benefits of strong AI even if there is a low probability of it being developed.

Problems specific to AI

- Some of the constraints on the development of strong AI seem to be philosophical as we don't yet have a clear idea about the nature of qualia (the subjective, internal component of sense perceptions) and consciousness.
- However, if we are concerned about social, perhaps even existential harms from strong AI, qualia and consciousness seem to be a red herring as philosophical detectable in behaviour.
- If we can make a strong AI, we can make something that is much smarter than us because we can incorporate in it weak AI that already far surpasses human capacities.
- So when thinking about the risks of strong AI, we should perhaps assume that any AI we build will likely be amalgamated with existing technologies to produce systems capable of performing many human functions.