

Stuart Armstrong: Predicting AI... or failing to

Extended Abstract: Predictions about the future development of artificial intelligence are as confident as they are diverse. Starting with Turing's initial estimation of a 30% pass rate on Turing test within 50 years (Turing, 1950), computer scientists, philosophers and journalists have never been shy to offer their own definite prognostics, claiming AI to be impossible (Jacquette, 1987) or just around the corner (Darrach, 1975). What are we to make of these predictions? What are they for, and what can we gain from them? Are they to be treated as light entertainment, the equivalent of fact-free editorials about the moral decline of modern living? Or are there some useful truths to be extracted? Can we feel confident that certain categories of experts can be identified, and that their predictions stand out from the rest in terms of reliability?

In this paper, we'll start off by proposing classification schemes for AI predictions: what types of predictions are being made, and what kind of arguments or models are being used to justify them. Armed with this scheme, we'll then analyse some of these approaches from the theoretical perspective, seeing whether there are good reasons to believe or disbelieve their results. The aim is not simply to critique individual methods or individuals, but to construct a toolbox of assessment tools that will both enable us to estimate the reliability of a prediction, and allow predictors to come up with better results themselves.

Those theoretical results will be supplemented with the real meat of the paper: a database of 257 AI predictions (Partially available online at www.neweuropeancentury.org/SIAI-FHI_AI_predictions.xls), made in a period spanning from the 1950s to the present day. This database was assembled by researchers systematically searching through the available online literature, and is a treasure-trove of interesting predictions. Delving into this will enable us to show that there seems to be no such thing as an "AI expert" for timeline predictions: no category of predictors stands out from the crowd.

The final point of interest is the unexpected robustness of some philosophical arguments. Philosophers making very general meta-arguments seem to have higher added value to the reader than computer scientists giving an expected date for the arrival of AI.