

How Long Until Human-Level AI? Results from an Expert Assessment

Seth D. Baum, Pennsylvania State University
Ben Goertzel, Novamente LLC
Ted G. Goertzel, Rutgers University

Published in: *Technological Forecasting & Social Change*, 2011, 78(1): 185-195
This file version: 20 April 2011

Abstract

The development of human-level AI has been a core goal of the AI field since its inception, though at present it occupies only a fraction of the field's efforts. To help understand the viability of this goal, this article presents an assessment of expert opinions regarding human-level AI research conducted at AGI-09, a conference for this AI specialty. We found that various experts strongly disagree with each other on certain matters, such as timing and ordering of key milestones. However, we did find that most experts expect human-level AI to be reached within upcoming decades, and all experts give at least some chance that some milestones will be reached within this time. Furthermore, a majority of experts surveyed favor an integrative approach to human-level AI rather than an approach centered on a particular technique. Finally, experts are skeptical about the impact of massive research funding, especially if it is concentrated in relatively few approaches. These results suggest that the possibility of achieving human-level AI in the near term should be given serious consideration.

Keywords: Expert elicitation; artificial intelligence; artificial general intelligence; technological forecasting; expert judgment

1. Introduction

The field of Artificial Intelligence was founded in the mid 1950's with the aim of constructing "thinking machines" – computer systems with human-like general intelligence, and capability equaling and eventually exceeding that of human beings. After a few initial successes, early AI researchers predicted that human-level AI would soon be achieved. In 1965, Herbert Simon predicted that "machines will be capable, within twenty years, of doing any work that a man can do" [1, p.108]. Two years later, Marvin Minsky predicted that "within a generation... few compartments of intellect will remain outside the machine's realm" [1, p.109]; see also [2]. But the task proved recalcitrant and the failure of their optimistic timing projections was discouraging.

Subsequent timing predictions became far more muted and AI research shifted focus towards producing AI systems demonstrating intelligence on specific tasks in relatively narrow domains. Much of this "narrow AI" research has been dramatically successful, with applications in many areas of science and industry. But one lesson learned through this work is that the magnitude of

the difference between narrow, task-specific AI capability and more general, human-like intelligence (referred to throughout this paper as *artificial general intelligence*, or AGI) including the ability to generalize across divergent contexts and to maintain a sense of self. For most recent and contemporary AI researchers, building AI with human-like general intelligence is simply not a concrete research goal.

Over the past few years, however, there has been a resurgence of research interest in the original goals of AI [3-7]. This resurgence is based on the assessment that advances in computer hardware, computer science, cognitive psychology, neuroscience, and domain-specific AI put contemporary researchers in a far better position to approach these goals than were the founders of AI. Recent years have seen a growing number of special sessions, workshops, and conferences devoted specifically to topics such as human-level AI and AGI. This includes the annual Biologically Inspired Cognitive Architectures AAAI Symposium, a series of workshops on integrative cognitive architectures; the most recent is the AAAI-10: Special Track on Integrated Intelligence (II) [8]. Another major AGI event is the international AGI conference series (AGI-08, AGI-09, AGI-10, etc.) [9]. Some recent convocations of leading AI researchers, such as the AI@50 conference [10], the 50th Anniversary Summit of Artificial Intelligence [11], and the annual Singularity Summits at Stanford University, San Francisco, and New York City [12-13] have also focused largely on the AGI problem. A report from AI@50 notes that these days “much of the original optimism is back, driven by rapid progress in artificial intelligence technologies” [14]. Meanwhile, MIT has recently launched a 5-year initiative aimed at exploring novel approaches to creating advanced thinking machines [15]. And noted futurist and AI inventor Ray Kurzweil has widely publicized his view that human-level AGI will be achieved within the next few decades [16-17] a view recently supported by Intel Chief Technology Officer Justin Rattner [18].

We summarize these recent trends as a resurgence in “AI optimism”, whereby “optimism” in this context we mean specifically “timing optimism”. Timing optimism is of course not the same as optimism about the beneficial effects of human-level AGI. One can be optimistic that AGI will be achieved soon, but pessimistic that the consequences of AGI will be harmful. Likewise, one can be pessimistic that AGI will not be achieved soon, but optimistic that the consequences of AGI will be beneficial, whenever AGI is achieved.

Of course, many in the field are less optimistic about the timing of human-level AGI. Not all experts believe that the time is ripe for a return to the original goals. Craig Silverstein of Google (a company that carries out a huge amount of narrow-AI research under the supervision of AI guru Peter Norvig) recently told a reporter that such computers are “hundreds of years away” [19]. Mark Andreessen, the founder of Netscape, said that “we are no closer to a computer that thinks like a human than we were fifty years ago” [19]. Some AI researchers, such as Selmer Bringsjord (chair of RPI's Department of Cognitive Science) even doubt that computers will ever display humanlike intelligence [20]. And the standard university AI curriculum continues to focus almost entirely on narrow AI, with only passing reference to the field's original grand goals.

While AI experts with less optimistic views on timing rarely take the time to elaborate the reasons for their views, several relevant perspectives are well articulated in the literature. Hubert

Dreyfus's classic critique of AI [21-22] and John Searle's perspective of "biological naturalism" [23] both argue that achieving human-like intelligence would require human-like physical embodiment and social context. Philosopher William Hurlbut [24] argues that (very roughly speaking) human mind emerges from human brain according to patterns and processes too complex for humans to fully understand anytime in the foreseeable future. Roger Penrose [25] and Stuart Hameroff [26] believe that human intelligence relies on quantum or quantum gravity phenomena that don't exist in ordinary digital computers. Finally, some experts may believe that there is no fundamental barrier to creating human-level AGI on digital computers, but that it's simply a very hard problem which would require one or more radical breakthroughs so far beyond current understanding that they will probably take the AI community many centuries to solve. Anecdotally, based on conversations with survey participants and others at AGI-09, our impression is that many AI researchers who are timing pessimists don't have any highly specific objection to AGI optimism in mind, but rather feel that there is a large amount of uncertainty involved so that there are probably many large "hidden rocks" along the path to human-level AGI, which might or might not prove insurmountable. If the path to AGI is laden with unforeseen obstacles, then there may be no way to predict when or if the requisite scientific breakthrough might occur. On the other hand, AGI timing optimists generally believe that human-level intelligence is a straightforward although complex "mechanical" process that can be understood by extending today's scientific ideas. If this is correct, then it is reasonable to consider detailed projections regarding how much time will pass until human-level AGI is built.

One issue on which there is minimal disagreement is that achievement of AGI would have a radically transformative effect on multiple aspects of science, industry, commerce, government, and arts. In the most extreme case, an intelligence explosion would ensue in which initial AGIs would design and build ever-smarter AGIs, leaving humans in the dust. In this eventuality, super-intelligent AGIs might use their vast intelligence to either solve a great many of humanity's problems or to destroy (perhaps inadvertently) their human creators [27]. But whether the results are good or bad, the achievement of human-level AGI would be among the most important events in human history.

It is thus of considerable importance and interest to make the best estimates possible of when and how AGI milestones may occur. Unfortunately, the database for making such estimates is quite limited. We can extrapolate hardware trends such as Moore's Law into the future, but there is no guarantee that these trends will persist, or that the requisite software will accompany them. Nor is it certain that hardware equivalent to the wetware in the human brain will be necessary or sufficient for human-level AI. There may be software paradigms that will permit human-level AGI on sub-human-level hardware, or perhaps more advanced hardware will be required for some reason.

Given the limitations of hardware and software data, expert opinion can be an important information source for our understanding of AGI. Expert opinion refers simply to the views and estimates of people who have exceptional knowledge about some topic, in this case AGI. Expert opinion, if carefully elicited, can yield meaningful insights on what is and isn't known about likely developments in a range of phenomena, including AGI. Expert assessment is a particularly insightful methodology when – as is the case with AGI – experts lack consensus on important parameters.

We are aware of only two previous studies explicitly exploring expert opinion on the future of AGI. In 2006, a seven-question poll was taken of participants at the AI@50 conference [28]. Four of the seven questions are particularly relevant. Asked “when will computers be able to simulate every aspect of human intelligence?”, 41% said “More than 50 years” and 41% said “Never”. Thus approximately 80% of the participants at the conference were AI “timing pessimists” and 20% were “timing optimists”. Seventy-nine percent said that an accurate model of thinking is impossible without further discoveries about brain functioning. Sixty percent of the participants strongly agreed that “AI should take a multidisciplinary approach, incorporating stats, machine learning, linguistics, computer science, cognitive psychology, philosophy, and biology”. Finally, 71% said that “statistical/probabilistic methods are most accurate in representing how the brain works”.

The other survey, taken in 2007 by futurist entrepreneur Bruce Klein [29], was an online survey that garnered 888 responses, asking one question: “When will AI surpass human-level intelligence?” Most of those who chose to respond to this survey were AI “optimists” who believed that human-level artificial intelligence would be achieved during the next half century. The distribution of responses is reproduced in Fig. 1. While the results show that a substantial number of optimists exist, Klein’s study has several important limitations. First, the sample population is broad, containing many non-experts. Second, the study only asks for point estimates for when superhuman AI will occur, neglecting any uncertainty in study participants’ beliefs. Finally, the study only asked one question, despite the plethora of other important aspects of AGI.

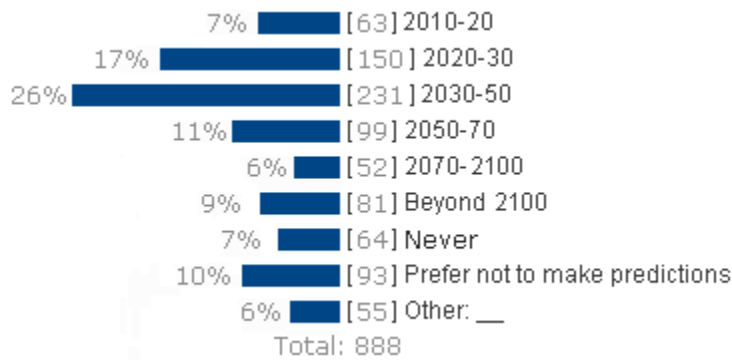


Fig. 1. Results from the survey taken by Klein [29] in 2007 asking the question “When will AI surpass human-level intelligence?”

The AI@50 and Klein studies are interesting because they show that significant numbers of experts and interested persons believe that AGI with intelligence at least equaling the human level will exist within upcoming decades. The current study probes more deeply into the thinking of people with substantial AGI expertise. Many (though not all) of these experts can be classified as timing optimists. While their views are not representative of the larger population of AI experts, they are nonetheless worth considering. In many cases, these experts are working seriously on AGI because they believe there is a strong likelihood of achieving success with the

fundamental theoretical and hardware tools at hand or likely to be developed in the next few decades.

This study is based on an expert assessment survey distributed to attendees of the Artificial General Intelligence 2009 (AGI-09) conference [30]. This was a gathering of AI researchers and other individuals interested specifically in pursuing theoretical or practical work on AGI. The survey covered the timing of AGI milestones, the effectiveness of various technical approaches to achieving AI, the potential for AGI to help or harm humanity, and the possibility of an AI being conscious.

2. Methods

Expert assessments have been used for several decades to provide timely information and aid decision making. Expert assessments have been used in the highest levels of study, including the United Nations-sponsored Intergovernmental Panel on Climate Change [31]. A range of assessment methods exist [32-34]. There is no single optimal methodology for assessing a group of experts. Each assessment should be customized to meet the questions of the research, the abilities of the experts, and any other relevant factors. Our study draws heavily on recent expert elicitation methodology that emphasizes highlighting the diversity of expert opinion instead of reaching consensus [34, p.146-154]. While this expert elicitation has been used primarily for estimation of static parameter values, it has also been used for several technological forecasting study [35-38].

Our study used a two-part survey. The first part is a standardized survey instrument distributed at the AGI-09 conference. This survey is adapted from previous expert elicitation surveys, in particular [39]. It features primarily quantitative questions. Many expert elicitation studies only use a survey along these lines. However, we felt that our quantitative survey results would be insightfully complemented by qualitative follow-up questions. So, the second part of our survey consists of follow-up questions customized for each survey participant and distributed to participants via email. Unlike follow-up questions used in Delphi studies, ours were focused on clarifying and expanding original survey responses instead of working towards consensus. These follow-up questions give us a richer understanding of the meanings and reasons behind the quantitative responses given in the initial survey.

2.1 Initial Survey

The initial survey was in a single standardized form printed across 18 pages for each of the survey participants. Participants reported spending an average of 28 minutes on the survey. All questions were quantitative, closed-ended questions, asking for numbers or binary true/false or yes/no answers. However, participants were encouraged to write any clarifying or other relevant remarks anywhere on the survey instrument; most participants took advantage of this option. The full survey instrument that we used for our investigation is available online [40]. Here we briefly summarize the most important questions that were posed.

Extensive empirical research shows that experts and non-experts alike tend to be overconfident in their probability estimates [34]. In rough terms, this means that for any given parameter, individuals tend to underestimate the probability that the actual value of the parameter is quite different from their best guess estimate. For example, someone might estimate that there is a 90% chance of the parameter values being within some range when that the actual values are within the range only 50% of the time. Our survey took several steps to reduce this overconfidence bias. Before any technical questions were asked, the survey explained overconfidence so as to sensitize participants to be cautious about it when answering questions. Additionally, some questions were worded so as to reduce overconfidence.

The first set of questions elicited experts' beliefs about when AI would reach each of four milestones: passing the Turing test,¹ performing Nobel quality work, passing third grade, and becoming superhuman. These specific milestones were selected to span a variety of levels of advanced general intelligence. For each milestone, two question versions were asked – with and without additional funding – making for eight total milestone questions. These two versions explore the possibility that when the milestones are reached depends on the resources available to researchers. The amount of additional funding listed, \$100 billion per year, is obviously more than could be used for AGI research; the intent with this figure is to ensure that money would not be a scarce resource in this hypothetical AGI development scenario.

For each of the eight milestone questions, we asked the respondents to give us estimates representing 10%, 25%, 75%, and 90% confidence intervals, as well as their best estimate dates. The 10% point represents the date in which the expert estimates that there is a 10% chance that the milestone will have been achieved, and so on. We asked experts to estimate the 10% and 90% intervals first in order to reduce overconfidence. Collectively, the five points correspond at least roughly with points on participants' probability density functions for the uncertain milestone achievement dates. The five point approach represents a trade-off between gaining deeper insight into experts' beliefs and demanding more detail from the experts.

Subsequent questions requested only single point estimates instead of multiple points on a probability density function. This choice was made to keep the overall survey instrument relatively short. The subsequent questions covered four topics. Three questions asked what embodiment the first AGIs would have: physical or virtual robot bodies or a minimal text- or voice-only embodiment. Eight questions asked what AI software paradigm the first AGIs would be based on: formal neural networks, probability theory, uncertain logic, evolutionary learning, a large hand-coded knowledge-base, mathematical theory, nonlinear dynamical systems, or an integrative design combining multiple paradigms. Three questions asked the likelihood of a strongly negative-to-humanity outcome if the first AGIs were created by: an open-source project, the US military, or a private for-profit software company. Two true/false questions asked if quantum computing or hypercomputing would be required for AGI. Two yes/no questions asked if AGIs emulating the human brain conceptually or near-exactly would be conscious in the sense that humans are. Finally, 14 questions asked experts to evaluate their own expertise in several

¹ Participants expressed concern that the Turing test milestone is ambiguous, due to the numerous variations of the test. In response to this, several participants specified the Turing test variant their responses are based on. At the time of survey distribution, a verbal suggestion was given to consider the “one hour” rather than the “five minute” Turing test as some potential participants felt the latter could too easily be “gamed” by narrow-AI chatbots without significant general intelligence.

subjects: cognitive science, neural networks, probability theory, uncertain logic, expert systems, theories of ethics, evolutionary learning, quantum theory, quantum gravity theory, robotics, virtual worlds, software engineering, computer hardware design, and cognitive neuroscience.

2.2 Follow-Up Questions

After analyzing results of the initial survey, we wrote and distributed follow-up questions customized for each participant. Whereas the initial survey consisted entirely of quantitative, closed-ended questions, the follow-up questions were all qualitative, open-ended questions, asking for prose explanations and clarifications of their initial survey responses. Participants received between two and six questions. While the initial survey responses raised a large number of issues we were interested in following up on, we chose only the most important questions so as to minimize the burden on survey participants' time.

2.3 Study Participants

Participants were recruited at the AGI-09 conference. Due to the highly specialized nature of the AGI-09 conference, no conference attendees were excluded. Participants thus have a range of levels of expertise, from graduate students to senior researchers. This inclusive approach risked eliciting low-quality responses from non-experts who happen to be at the conference. On the other hand, more exclusive approaches risk missing important ideas or information and also risk biasing results in certain directions. All of the participants in our study displayed a depth of thinking demonstrative of significant AGI expertise.

Twenty-one people participated in the study. Nineteen people completed the initial survey at the AGI-09 conference; two completed it later, submitting it via postal mail. Also, 18 people (including both of those who submitted via postal mail) completed the follow-up questions. Such levels of participation are common for expert elicitation – one recent study [38] had just 7 experts; another study [41] had just 12 experts. While a larger number of participants may bring additional information, our participant group is very much adequate to yield meaningful insights.

It should be emphasized here that the purpose of our study is ultimately to gain insight about AGI itself, not about AGI experts. Thus our sample of AGI experts must provide rich insights about AGI, but it does not need to be in some way representative of the broader population of AGI experts. After all, there is no guarantee that an accurate assessment of the opinions of the broader on AGI population correlates with the realities of AGI. It is entirely possible that the deepest insights about AGI come from outliers from the broader population of experts. Indeed, the very existence of disagreement among experts is grounds for reflection on the state of our knowledge. As our results make clear, our sample succeeds quite well at providing rich insight about AGI. While we believe our sample to be at least somewhat representative of the broader AGI expert population, this is of secondary importance.

Study participants have a broad range of backgrounds and experience, all with significant prior thinking about AGI. Eleven are in academia, including six Ph.D. students, four faculty members, and one visiting scholar, all in AI or allied fields. Three lead research at independent AI research organizations and three do the same at information technology organizations. Two are

researchers at major corporations. One holds a high-level administrative position at a relevant non-profit organization. One is a patent attorney. All but four participants reported being actively engaged in conducting AI research.

We are not listing the names of participating experts for several reasons. First, prior expert assessments vary in their approach to anonymity. Many studies do not list experts' names at all (e.g. [42-43]). Many other studies list experts' names but do not link specific names to specific responses (e.g. [35-39, 41]). Second, due to the sensitive nature of AGI, several of our experts requested that their names not be listed. The others participated on the condition that their names not be linked to their responses. Listing experts' names and discussing results in this way could enable some readers to link specific names to specific responses.

3. Results

Although the majority of the experts who participated in our study were clearly AGI timing optimists, there was a significant minority of timing pessimists. While we don't know how the views of study participants compare to those of the rest of the conference attendees, it seems unsurprising to find an optimist majority at an AGI conference. Most AGI pessimists may see little reason to bother attending an AGI conference, although this is not universal – some known AGI pessimists do regularly attend AGI conferences. Finally, all experts in our study, including those who could be classified as timing pessimists, gave at least a 10% chance of some AI milestones being achieved within a few decades.

Additionally, the amount of agreement between the experts varied considerably from one question to another. On several questions, experts disagreed strongly – more so than is typically found in expert assessments – indicating a lack of professional consensus and the presence of multiple competing mental models of the future of AGI. Meanwhile, on other questions, much more agreement existed, though rarely full consensus.

3.1 Milestone Timing Without Additional Funding

Much of the initial survey and follow-up questions focused on the timing of four AI milestones: passing the Turing test, performing Nobel-quality research, passing the third grade, and gaining dramatically superhuman capabilities. The range of the best guess estimates given for these milestones without massive additional funding is summarized in Fig. 2. This figure clearly shows the dichotomy between timing optimists and timing pessimists. It also shows that a substantial number of AGI experts expect that key milestones will be reached within upcoming decades.

The overlap in the best guess estimates for the milestones was surprising. Using human cognitive development as a model, one might think that being able to do Nobel level science would take much longer than being able to conduct a social conversation, as in the Turing Test. It is true that many complex problems can be solved effectively by artificial expert systems. But performing Nobel level science requires more than just complex problem solving: it also requires creativity in the formulation of new research questions and ideas, something that expert systems

cannot achieve. Thus we might expect that performing Nobel level science would be more difficult than passing the Turing test. However, some experts thought it would be quite different for an AGI. One expert observed that “making an AGI capable of doing powerful and creative thinking is probably easier than making one that imitates the many, complex behaviors of a human mind – many of which would actually be hindrances when it comes to creating Nobel-quality science”. This expert observed that “humans tend to have minds that bore easily, wonder [sic] away from a given mental task, and that care about things such as sexual attraction, all [of] which would probably impede scientific ability, rather than promote it”. To successfully emulate a human, a computer might have to disguise many of its abilities, masquerading as being less intelligent, in certain ways, than it actually was. This is one reason that the Turing Test milestone might be reached later than the Nobel-quality science milestone.

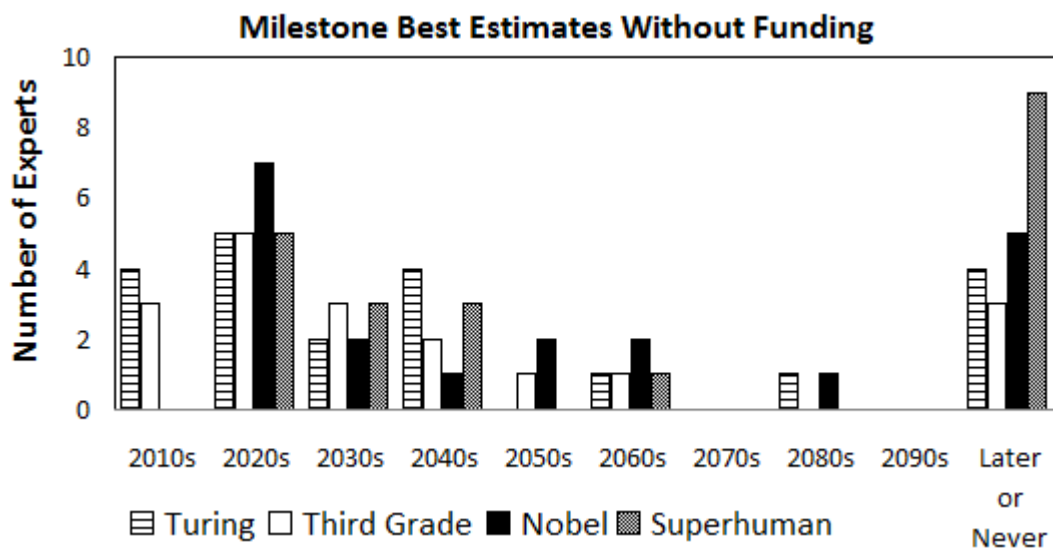


Fig. 2: Milestone best estimate guesses without massive additional funding. Estimates are for when AI would achieve four milestones: the Turing Test (horizontal lines), third grade (white), Nobel-quality work (black), and superhuman capability (grey).

We computed median estimates for each milestone (Table 1). The medians would be only slightly lower if these less optimistic respondents were excluded from the analysis. Medians were used, instead of means, because these are less susceptible to distortion via statistical outliers.²

3.2 Milestone Estimates with Massive New Funding

Another interesting result concerns the role of funding – the predicted impact on the AGI field of a hypothetical \$100 billion per year funding infusion. The range of the best guess estimates given for these milestones with massive additional funding is summarized in Fig. 3. Most experts estimated that a massive funding increase of this nature would cause the AI milestones to

² In this case, the sample size is so small that it’s hard to rigorously speak about “outliers,” but it’s clear qualitatively that the very long estimates given by some experts, including those who thought that AGI might never be achieved, would have a disproportionate effect upon the means.

be reached sooner. However, for many of these experts, the difference in milestone timing with and without massive funding was quite small – just a few years. Furthermore, several experts estimated that massive funding would actually cause the AI milestones to be reached later. One reason given for this is that with so much funding, “many scholars would focus on making money and administration” instead of on research. Another reason given is that “massive funding increases corruption in a field and its oppression of dissenting views in the long term”. Of those who thought that funding would make little difference, a common reason was that AGI progress requires theoretical breakthroughs from just a few dedicated, capable researchers, something that does not depend on massive funding. Another common reason was that the funding would not be wisely targeted. Several noted that funding could be distributed better if there was a better understanding of what paradigms could produce AGI, but such an understanding is either not presently available or not likely to be understood by those who would distribute funds.

Milestone	10%	25%	50%	75%	90%
Turing Test	2020	2030	2040	2050	2075
Nobel Science	2020	2030	2045	2080	2100
Third Grade	2020	2025	2030	2045	2075
Super Human	2025	2035	2045	2080	2100

Table 1. Median estimates for each AGI milestone at each level of confidence.

Because of the lack of agreement on a single paradigm, several experts recommended that modest amounts of funding should be distributed to a variety of groups following different approaches, instead of large amounts of funding being given to a “Manhattan Project” type crash program following one approach. Only one expert recommended concentrating funds in a fewer number of groups. Similarly, several observed that well-funded efforts guided by a single paradigm had failed in the past, including the Japanese Fifth Generation Computer Systems project. While most experts did not address the question of how to distribute funds across groups, our initial impression is that there would be some consensus for distributing funds to a larger variety of groups. This result could have major implications for how available funds are to be allocated.

Several experts gave technical reasons for why funding would not have a significant impact. On this, one said, “AGI requires more theoretical study than real investment.” Another said, “I believe the development of AGI’s to be more of a tool and evolutionary problem than simply a funding problem. AGI’s will be built upon tools that have been developed from previous tools. This evolution in tools will take time. Even with a crash project and massive funding, these tools will still need time to develop and mature.” Given that these experts are precisely those who would benefit most from increased funding, their skeptical views of the impact of hypothetical massive funding are very likely sincere.

3.3 Milestone Timing: Estimates of Certainty

Since we do not know for sure when milestones will be reached, it is helpful to characterize the uncertainty about this. Some insights can be gained from considering the distributions of best guess estimates of experts (as in Figs. 2 and 3) or of broader populations (as in Fig. 1). However, each individual has uncertainty about his/her own estimates. These uncertainties are not found from presentations of best guess estimates, but they can be found from the estimates across the range of confidence intervals elicited in our survey.

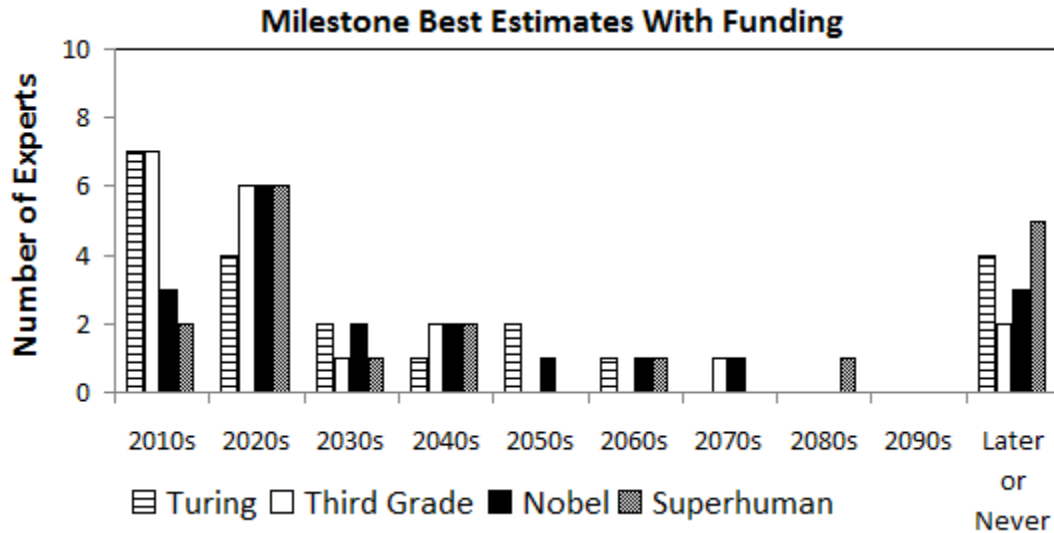


Fig. 3. Milestone best estimate guesses with massive additional funding. Estimates are for when AI would achieve four milestones: the Turing Test (horizontal lines), third grade (white), Nobel-quality work (black), and superhuman capability (grey).

Participants showed a broad range of certainty levels in their timing estimates. Several participants showed little uncertainty, indicating that they believed it was very likely that milestones would be achieved within a narrow range of dates. The participants giving the narrowest ranges were all strong timing optimists. Meanwhile, many of the participants showed much uncertainty. Some of these individuals gave substantial probabilities for both early and late milestone achievement dates. For these individuals, the timing optimist/pessimist dichotomy does not readily apply. Other individuals showing much uncertainty could be classified as timing pessimists in that they considered it most likely that milestones would be achieved much later. The fact that these individuals are not certain about their pessimism could explain why they thought that attending a conference on AGI would be worthwhile. One expert estimated that there was at least a 25% chance that the Turing Test would never be achieved, but he also thought that there was a 10% chance that it would be achieved by the year 2020. Another estimated that there was a 25% chance it would not be achieved until the year 3000, but that there was a 10% chance it would happen by 2050. The stronger optimists at the conference were much more certain of their opinions. One estimated that there was a 90% chance that the Turing Test would be achieved by 2016 and a 10% chance that it would be achieved as soon as 2010. The range of variation in the estimates for the four milestones without massive additional

funding is shown in Fig. 4. Dates after 2100 are removed to improve resolution in the rest of the plots.

There was substantial agreement among the more optimistic participants in the study. Most thought that there was a substantial likelihood that the various milestones would be passed sometime between 2020 and 2040. Sixteen experts gave a date before 2050 as their best estimate for when the Turing test would be passed. Thirteen experts gave a date between 2020 and 2060 as their best estimate for when superhuman AI would be achieved. (The others all estimated later dates, ranging from 2100 to never.) The opinions of the optimists were similar to or perhaps somewhat more optimistic than Kurzweil's well-known projections.

As noted above, optimism in this sense is about the timing of AI milestones. It does not imply a belief that achieving AGI would be a good thing. To the contrary, one can be optimistic that AGI will happen soon yet believe that AGI would have negative outcomes. Indeed, several experts reported this pair of beliefs. Results about the likelihood of negative outcomes are discussed further below.

3.4 Milestone Order

There was little agreement among experts on the order in which the four milestones (Turing test; third grade; Nobel; superhuman) would be achieved. The only area of consensus was that the superhuman milestone would be achieved either last or at the same time as other milestones. Meanwhile, there was significant divergence regarding the order of the other three milestones. One expert argued that the Nobel milestone would be easier precisely because it is more sophisticated: to pass the Turing test, an AI must "skillfully hide such mental superiorities". Another argued that a Turing test-passing AI needs the same types of intelligence as a Nobel AI "but additionally needs to fake a lot of human idiosyncrasies (irrationality, imperfection, emotions)". Finally, one expert noted that the third grade AI might come first because passing a third grade exam might be achieved "by advances in natural language processing, without actually creating an AI as intelligent as a third-grade child". This diversity of views on milestone order suggests a rich, multidimensional understanding of intelligence. It may be that a range of milestone orderings are possible, depending on how AI development proceeds.

The milestone order results highlight the fact that many experts do not consider it likely that the first human-level AGI systems will closely mimic human intelligence. Analogy to human intelligence would suggest that achieving an AGI capable of Nobel level science would take much longer than achieving an AGI capable of conducting a social conversation. However, as discussed above, an AGI would not necessarily mimic human intelligence. This could enable it to achieve the intelligence milestones in other orders.

3.5 Technical Approaches

Our initial survey asked about eight technical approaches used in constructing AI systems: formal neural networks, probability theory, uncertain logic, evolutionary learning, a large hand-coded knowledge-base, mathematical theory, nonlinear dynamical systems, and integrative designs combining multiple paradigms. For each, it requested a point estimate of the odds that

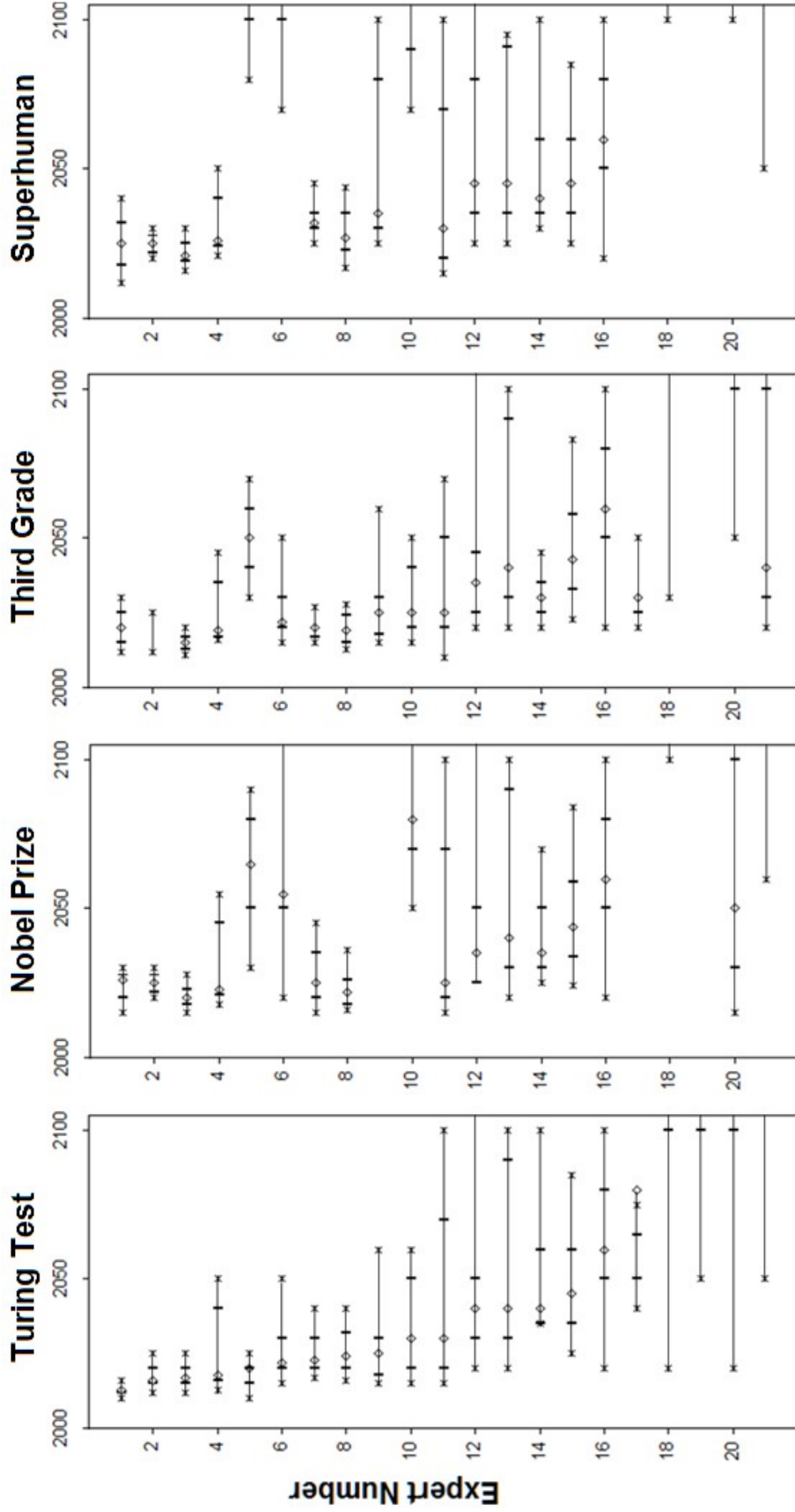


Fig. 4. Estimates for when AI would achieve four milestones without massive additional funding. Estimates are given at five confidence intervals. Asterisks represent 10% and 90% confidence bounds. Horizontal lines represent 25% and 75% confidence bounds. Diamonds represent best guess estimates. In each plot, participants are ordered by their Turing Test best guess estimates.

the approach would be critical in the creation of human-level AGI. The survey also asked about the odds that physical robotics, virtual agent control, or minimal text- or voice-based embodiments would play a critical role.

The most striking finding in this section was a strong majority opinion in favor of integrative designs. Of the 18 respondents who answered this question, 13 gave a probability of 0.5 or greater; the mean response across all 18 was 0.63.³ None of the specific technical approaches mentioned received strong support from any more than a small plurality of respondents. Probability theory received the strongest support, with a mean probability of 0.35 and with 7 out of 20 respondents answering the question giving a probability of 0.5 or greater.

The question about robotics vs. other forms of embodiment received a range of responses (Fig. 5). There were responses of 0.9, 0.94, 0.89, and 0.60 for physical robot embodiment, but the mean response was only 0.28. The handful of participants who felt robotics was crucial were all relatively optimistic, as seen by their mean date of 2034 for the Turing Test without massive additional funding. The virtual agents option received a similar but considerably less strong response, with top responses of 0.80, 0.65, and 0.60 and a mean response of 0.34. While some of the virtual worlds enthusiasts were optimistic about timing, this was much less of a strong pattern than in the robotics case. The preliminary impression one obtains is that a few researchers are highly bullish on robotics as the correct path to AGI in the relatively near term, whereas the rest feel robotics is probably not necessary for AGI. Meanwhile, a larger number of researchers are moderately optimistic about the role of virtual worlds for AGI, but with milder opinions about the importance of the approach.

3.6 Impacts of AGI

In science fiction, intelligent computers frequently become dangerous competitors to humanity, sometimes even seeking to exterminate humanity as an inferior life form. Several researchers have recently expressed similar concerns [27, 44]. Motivated by these concerns, we asked experts to estimate the probability of a negative-to-humanity outcome occurring if an AGI passes the Turing test. Our question was broken into three parts, for each of three possible development scenarios: if the first AGI that can pass the Turing test is created by an open source project, by the United States military, or by a private company focused on commercial profit.

This set of questions marked another instance in which the experts lacked consensus (see Fig. 6). Five experts estimated a less than 20% chance of a negative outcome, regardless of the development scenario. Four experts estimated a greater than 60% chance of a negative outcome, regardless of the development scenario. Only four experts gave the same estimate for all three development scenarios. Finally, several experts were more concerned about the risk from AGI itself, whereas others were more concerned that AGI could be misused by humans who controlled it.

³ In this section the survey did not require that the probabilities all add up to one, because the technical approaches are not necessarily mutually exclusive. When the numbers are normalized such that each participant's set of probabilities add up to one, then only 2 experts give integrated design a probability of 0.5 or greater, but 12 still rank it the most likely, including 1 who gave the highest probability to "none of the above". After normalization, the average for integrated design is 0.32; all other approaches ranged 6% (mathematical theory) and 12% (probability theory).

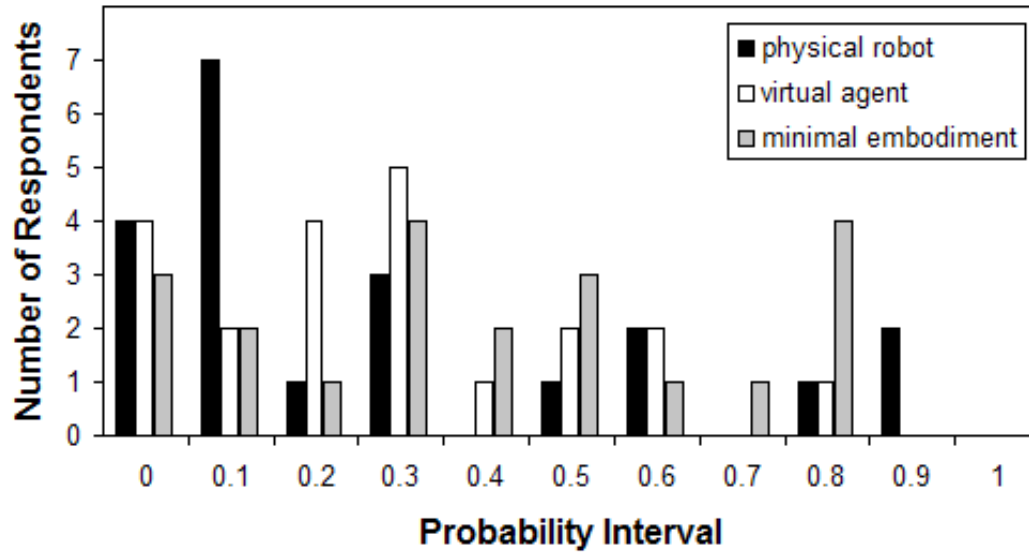


Fig. 5. Number of respondents for different embodiment approaches. Respondents are grouped into probability intervals for each of three different embodiment approaches: physical robot (black), virtual agent (white), and minimal embodiment (grey). Intervals are labeled on the horizontal axis by their lower bound. For example, the interval labeled “0.1” represents the interval 0.1 through 0.199.

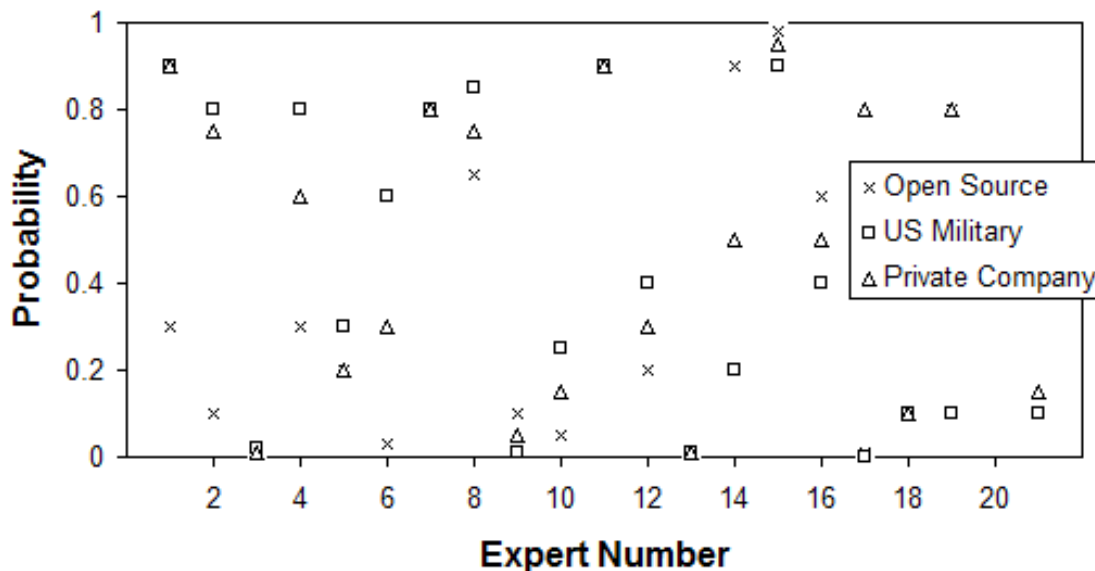


Fig. 6. Probability of a negative-to-humanity outcome for different development scenarios. The three development scenarios are if the first AGI that can pass the Turing test is created by an open source project (x's), the United States military (squares), or a private company focused on commercial profit (triangles). Participants are displayed in the same order as in figure 4, such that Participant 1 in figure 6 is the same person as Participant 1 in figure 4.

Interesting insights can be found in the experts' orderings of the riskiness of the development scenarios. Of the 11 experts who gave different estimates for each of the three scenarios, 10 gave the private company scenario the middle value. Of these 10, 6 gave the US military scenario the highest value and 4 gave it the lowest value. Thus the open source scenario and the US military scenario tend to be perceived as relatively safe or relatively dangerous, but experts are divided on which is which. Experts who estimated that the US military scenario is relatively safe noted that the US military faces strong moral constraints, has experience handling security issues, and is very reluctant to develop technologies that may backfire (such as biological weapons), whereas open source development lacks these features and cannot easily prevent the "deliberate insertion of malicious code". In contrast, experts who estimated that the open source scenario is relatively safe praised the transparency of open source development and its capacity to bring more minds to the appropriate problems, and felt the military has a tendency to be destructive.

Several experts noted potential impacts of AGI other than the catastrophic. One predicted that "in thirty years, it is likely that virtually all the intellectual work that is done by trained human beings such as doctors, lawyers, scientists, or programmers, can be done by computers for pennies an hour. It is also likely that with AGI the cost of capable robots will drop, drastically decreasing the value of physical labor. Thus, AGI is likely to eliminate almost all of today's decently paying jobs." This would be disruptive, but not necessarily bad. Another expert thought that, "societies could accept and promote the idea that AGI is mankind's greatest invention, providing great wealth, great health, and early access to a long and pleasant retirement for everyone." Indeed, the experts' comments suggested that the potential for this sort of positive outcome is a core motivator for much AGI research.

4. Conclusion

Despite decades of setbacks, the development of AI with the general intelligence of humans remains an important possibility both for the AI research community and for society at large. Given the paucity of hardware and software data enabling detailed extrapolations regarding the future of such AI, expert opinion is an important source of information. In this article, we have presented the most careful assessment of expert opinion on AGI to date. Given the long history of AI researchers making erroneous predictions, our results must be interpreted cautiously, but they nonetheless add much to our understanding of the future of AGI.

In the broadest of terms, our results on AGI timing concur with those of the two previous studies mentioned in the Introduction [28-29]. All three studies suggest that significant numbers of interested, informed individuals believe it is likely that AGI at the human level or beyond will occur around the middle of this century, and plausibly even sooner. Due to the greater depth of the questions, our survey also revealed some interesting additional information, such as the disagreement among experts over the likely order of AGI milestones and the relative safety of different AGI development scenarios. The experts' suggestions regarding funding (in particular, skepticism about the importance of funding and a preference for distributing funds among a broad range of research groups) are also potentially valuable.

Our results reinforce the idea that there is a significant likelihood of AGI being achieved within upcoming decades. This idea has profound societal implications given the strong potential for AGI to act as a transformative agent, either to massively help humanity or to destroy it. The experts in our study, like others who have commented on the issue, were divided on whether the impacts of AGI would be positive or negative for humanity. Our results add depth to our understanding of AGI impacts by exploring multiple development scenarios. However, the jury remains out regarding how to achieve positive outcomes and avoid negative outcomes. This is a crucial area for future research.

Currently, AGI experts hold a rich diversity of views, a situation which challenges our ability to make confident predictions about the future of AGI, but is not surprising given the state of development of the field. It would be interesting to carry out similar studies in future and see how expert views evolve and converge or diverge as AI technology advances.

Acknowledgements

We thank all of the experts who participated in the study reported here. Additionally, Granger Morgan and Aimee Curtright provided helpful feedback in our survey design; David Hart and Bruce Klein assisted in the development and delivery of the survey; Josh Hall, Thomas McCabe, Michael Vassar, Pei Wang, and two anonymous reviewers provided helpful comments on an earlier draft of this article.

References

- [1] D. Crevier, *AI: The Tumultuous History of the Search for Artificial Intelligence*, Basic Books, New York, 1993, p.108-109.
- [2] P. McCorduck, *Machines Who Think: 25th Anniversary Edition*, A.K. Peters, Natick, MA, 2004.
- [3] B. Goertzel, C. Pennachin (Eds.), *Artificial General Intelligence*, Springer Verlag, New York, 2007.
- [4] B. Goertzel, P. Wang (Eds.), *Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms*, IOS Press, Amsterdam, 2007.
- [5] M. Hutter, P. Hitzler, B. Goertzel (Eds.), *Proceedings of the Second Conference on Artificial General Intelligence*, Atlantis Press, Los Angeles, 2009.
- [6] P. Wang, B. Goertzel, S. Franklin (Eds.), *Artificial General Intelligence 2008*, IOS Press, Amsterdam, 2008.
- [7] N. Cassimatis, E.T. Mueller, P.H. Winston, Achieving human-level intelligence through integrated systems and research: introduction to this special issue. *AI Magazine* 27(2) (2006) 12-14.
- [8] <http://www.aaai.org/Conferences/AAAI/2010/aaai10ii.php> (Accessed 3 July 2010)
- [9] <http://agi-conf.org> (Accessed 3 July 2010)
- [10] <http://www.dartmouth.edu/~ai50> (Accessed 3 July 2010)
- [11] <http://www.ai50.org> (Accessed 3 July 2010)
- [12] <http://sss.stanford.edu> (Accessed 3 July 2010)

- [13] <http://www.singularitysummit.com> (Accessed 3 July 2010)
- [14] J. Markoff, Optimism as Artificial Intelligence Pioneers Reunite, *New York Times*, December 7, 2009.
- [15] <http://mmp.cba.mit.edu> (Accessed 3 July 2010)
- [16] R. Kurzweil, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*, Viking Press, New York, 1998.
- [17] R. Kurzweil, *The Singularity Is Near: When Humans Transcend Biology*, Viking Penguin, New York, 2005.
- [18] J. Harris, Intel Predicts Singularity by 2048, *TechWatch*, August 22, 2008, <http://www.techwatch.co.uk/2008/08/22/intel-predicts-singularity-by-2048> (accessed 17 March 2010).
- [19] K. Auletta, *Googled: The End of the World as We Know It*, Penguin Press, New York, 2009, p.327.
- [20] S. Bringsjord, M. Zenzen, *Superminds: People Harness Hypercomputation, and More*, Kluwer, Dordrecht, The Netherlands, 2003.
- [21] H.L. Dreyfus, *What Computers Can't Do: The Limits of Artificial Intelligence*, MIT Press, Cambridge, MA, 1972.
- [22] H.L. Dreyfus, *What Computers Still Can't Do: A Critique of Artificial Reason*, MIT Press, Cambridge, MA, 1992.
- [23] J. Searle, *The Rediscovery of the Mind*, MIT Press, Cambridge, MA, 1992.
- [24] W.B. Hurlbut, W.H. Durham (Eds.), *Brain, Mind and Emergence*, Unpublished conference proceedings, Stanford University, 2003.
- [25] R. Penrose, *Shadows of the Mind*, Oxford University Press, Oxford, 1994.
- [26] S. Hameroff, *Ultimate Computing: Biomolecular Consciousness and Nanotechnology*, North Holland, Amsterdam, 1987.
- [27] E. Yudkowsky, Artificial intelligence as a positive and negative factor in global risk, in: N. Bostrom, M. Cirkovic (Eds.), *Global Catastrophic Risks*, Oxford University Press, Oxford, 2008.
- [28] M.H. Maker, *AI@50, Engaging Experience*, July 13-15, 2006, <http://www.engagingexperience.com/ai50> (accessed 17 March 2010).
- [29] B. Klein, When will AI Surpass Human-Level Intelligence? *AGI-World*, August 5, 2007, <http://www.novamente.net/bruce/?p=54> (accessed 17 March 2010).
- [30] H. de Garis, B. Goertzel, *The Second International Conference on Artificial General Intelligence (AGI-09)*, *AI Magazine* 30(4) (2009) 115-116.
- [31] R. Moss, S.H. Schneider, Uncertainties, in: R. Pachauri, R. Taniguchi, K. Tanaka (Eds.), *Guidance Papers on the Cross Cutting Issues of the Third Assessment Report of the IPCC*, World Meteorological Organisation, Geneva, 2000.
- [32] R.M. Cooke, *Experts in Uncertainty: Opinion and Subjective Probability in Science*, Oxford University Press, Oxford, 1991.
- [33] M.A. Meyer, J.M. Booker (Eds.), *Eliciting and Analyzing Expert Judgement: A Practical Guide*, Academic Press Limited, London, UK, 1991.
- [34] M.G. Morgan, M. Henrion, *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press, Cambridge, UK, 1990.
- [35] A.E. Curtright, M.G. Morgan, D.W. Keith, Expert assessments of future photovoltaic technologies, *Environ. Sci. Tech.* 42(24) (2008) 9031-9038.

- [36] E. Baker, H. Chon, J. Keisler, Carbon capture and storage: combining expert elicitations to inform climate policy. *Climatic Change* 96 (3) (2009) 379-408.
- [37] E. Baker, H. Chon, J. Keisler, Advanced solar R&D: combining economic analysis with expert elicitations to inform climate policy. *Energy Econ.* 31 (2009) S37-S49.
- [38] E. Baker, H. Chon, J. Keisler, Battery technology for electric and hybrid vehicles: expert views about prospects for advancement, *Technol. Forecast. Soc. Change* (2010) doi:10.1016/j.techfore.2010.02.005.
- [39] M.G. Morgan, P. Adams, D. Keith, Elicitation of expert judgments of aerosol forcing. *Climatic Change* 75(1-2) (2006): 195-214.
- [40] <http://sethbaum.com/research/agi> (Accessed 3 August 2010)
- [41] K. Zickfeld, A. Leverman, D.W. Keith, T. Kuhlbrodt, M.G. Morgan, S. Rahmstorf, Expert judgements on the response of the Atlantic meridional overturning circulation to climate change. *Climatic Change* 82(3-4) (2007) 235-265.
- [42] N.W. Arnell, E.L. Tompkins, W.N. Adger, Eliciting information from experts on the likelihood of rapid climate change, *Risk Analysis* 25(6) (2005) 1419-1431.
- [43] W. Bruine De Bruin, B. Fischhoff, L. Brilliant, D. Caruso, Expert judgments of pandemic influenza risks, *Global Public Health* 1(2) (2006) 178-193.
- [44] B. Joy, Why the future doesn't need us, *Wired Magazine*, April 2000.