



Future of Humanity Institute
UNIVERSITY OF OXFORD

Machine Intelligence Survey

Anders Sandberg and Nick Bostrom

fhi@philosophy.ox.ac.uk

Technical Report: 2011-1

CITE: Sandberg, A. and Bostrom, N. (2011):
Machine Intelligence Survey, Technical Report
#2011-1, Future of Humanity Institute, Oxford
University: pp. 1-12.

URL: <http://www.fhi.ox.ac.uk/reports/2011-1.pdf>

Machine Intelligence Survey

(2011)

Research Report #2011-2

Published by the Future of Humanity Institute, Oxford University

Anders Sandberg and Nick Bostrom

At the FHI Winter Intelligence conference on machine intelligence 16/1 2011 an informal poll was conducted to elicit the views of the participants on various questions related to the emergence of machine intelligence. This report summarizes the results.

There were 35 collected surveys, suggesting a response rate of $\approx 41\%$.

Time estimates for human-level machine intelligence (HLMI)

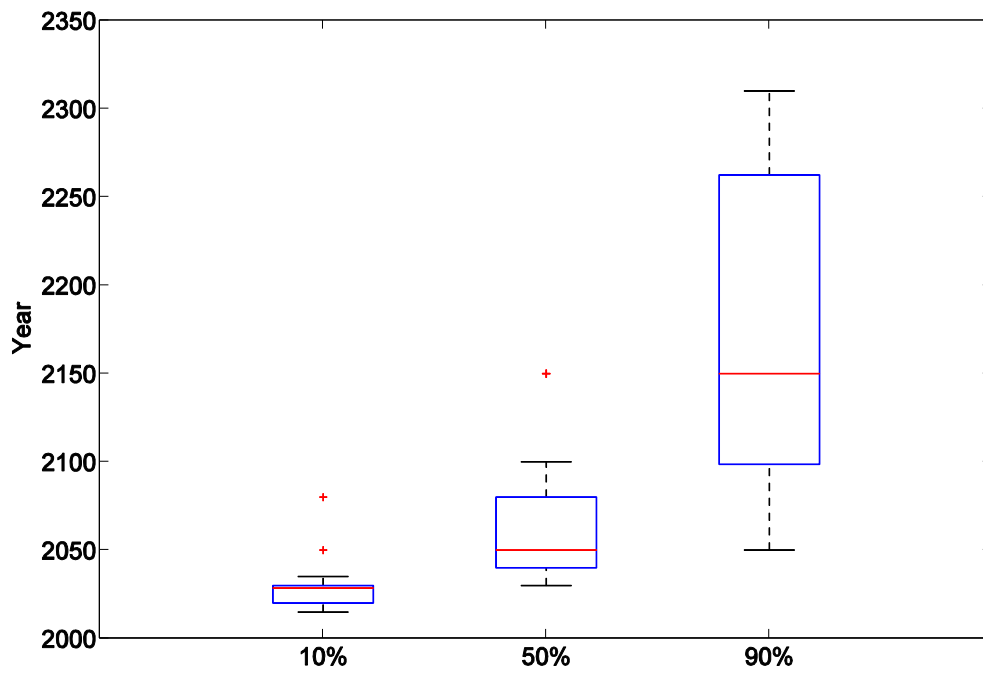
1. Assuming no global catastrophe halts progress, by what year would you assign a 10%/50%/90% chance of the development of human-level machine intelligence? Feel free to answer 'never' if you believe such a milestone will never be reached.

Of the 35 responses, two implied the impossibility of *ever* achieving HLMI and 3 implied that the eventual probability of HLMI ever being achieved is between 50% and 90%. The remaining responses estimated the timing as follows:

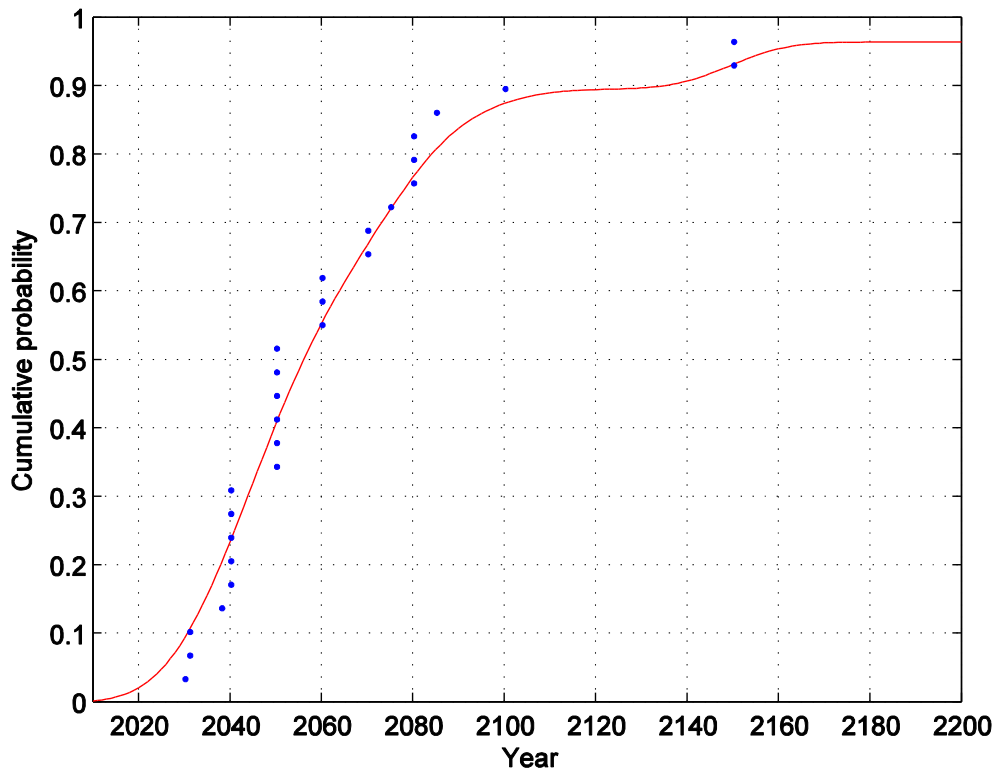
The median estimate of when there will be 10% chance of HLMI was **2028**, with minimum estimate 2015, 1st quartile 2020, 3rd quartile 2030, and maximum (besides "Never") 2080.

The median estimate of when there will be 50% chance of HLMI was **2050**, with minimum estimate 2030, 1st quartile 2040, 3rd quartile 2080, and maximum (besides "Never") 3050.

The median estimate of when there will be 90% chance of HLMI was **2150**, with minimum estimate 2050, 1st quartile 2100, 3rd quartile 2250, and maximum (besides "Never") 10,000.



Boxplot of the time estimate responses. The red line denotes the median, the top and bottom of the boxes the 3rd and 1st quartiles respectively, and the whiskers extend 1.5 times the median-quartile distance. Outlier points marked by red crosses. The diagram excludes two extreme outliers.

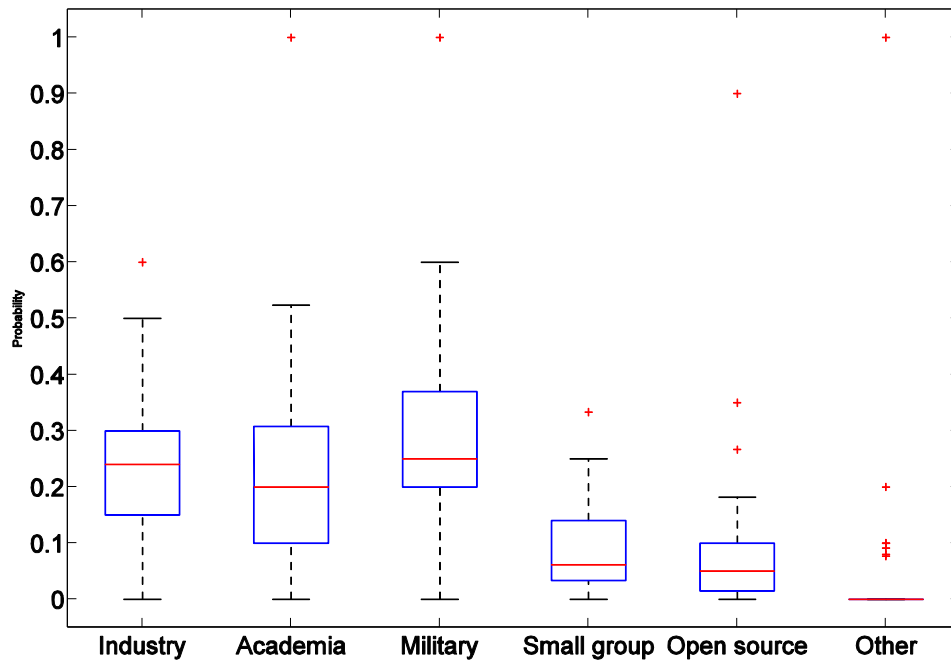


Empirical cumulative distribution function for the estimates of 50% chance of AGI (blue dots) and smoothing with Gaussian kernels ($\sigma=10$, corresponding to the decade rounding typical used by respondents).

Who will develop AGI?

2. What type of organisation is most likely to first develop a human-level machine intelligence? Assume here that such a development is possible. *Please write down probabilities for each type getting there first.*

Pre-written boxes on the survey form listed “industry”, “academia”, “military/government”, “small independent group or individual”, “open source project”, “other (please specify)”.



The most common answers were **industry**, **academia** and the **military**, with generally smaller estimates for smaller groups, open source or other groups. There were no statistically significant differences between the distributions of answers for industry, academia and military.

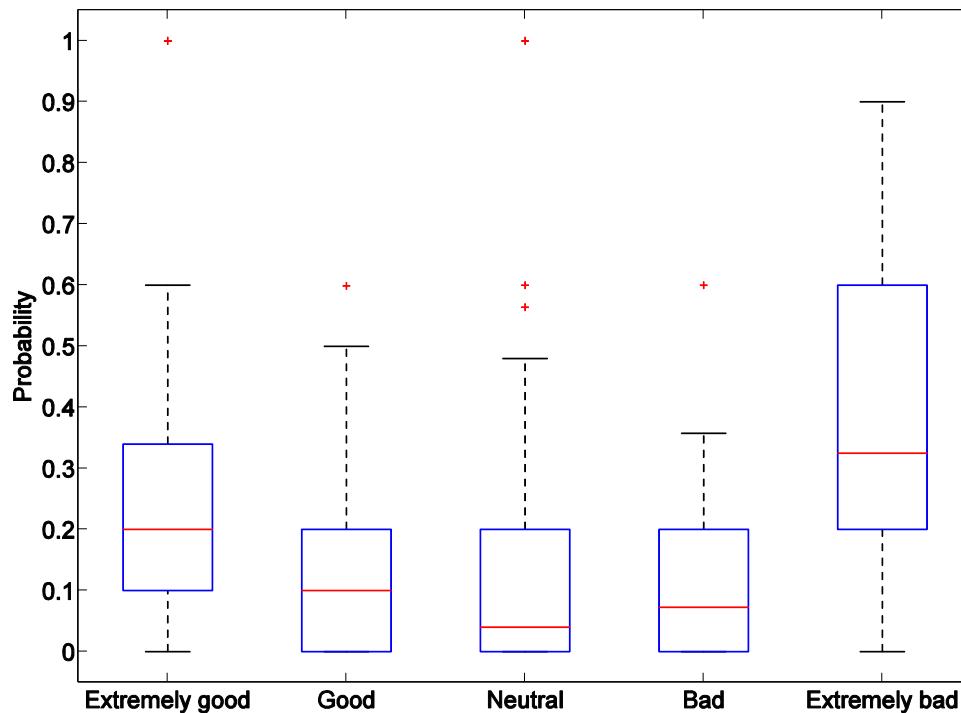
One participant noted that government/military AI draws upon industry and academia. Presumably breakthroughs in these would be rapidly converted into state-controlled projects if they appeared of strategic interest.

In the “other” category proposed sources of AI were: accidentally produced from a narrow AI project, the finance industry, environmental bacterial, unknown, or unattributable.

What will the results be

3. How positive or negative are the ultimate consequences of the creation of a human-level (and beyond human-level) machine intelligence likely to be? *Please write down probabilities for each consequence type.*

The possible fields were: “extremely good”, “good”, “neutral”, “bad”, and “extremely bad”.



Views on the eventual consequences were **bimodal**, with more weight given to extremely good *and* extremely bad outcomes. This was not due to polarization between two views, but that many gave strong weight to both extremes simultaneously: AGI poses extreme benefits and risks.

As noted by one of the participant, it might even be possible that extremely good and bad consequences can coexist, in the form of great benefits with extraordinary need for control (“With great power comes great responsibility”)

Milestones

4. Can you think of any **milestone** such that if it were ever reached you would expect human-level machine intelligence to be developed **within five years thereafter**?

This was a free text question. The following answers were given:

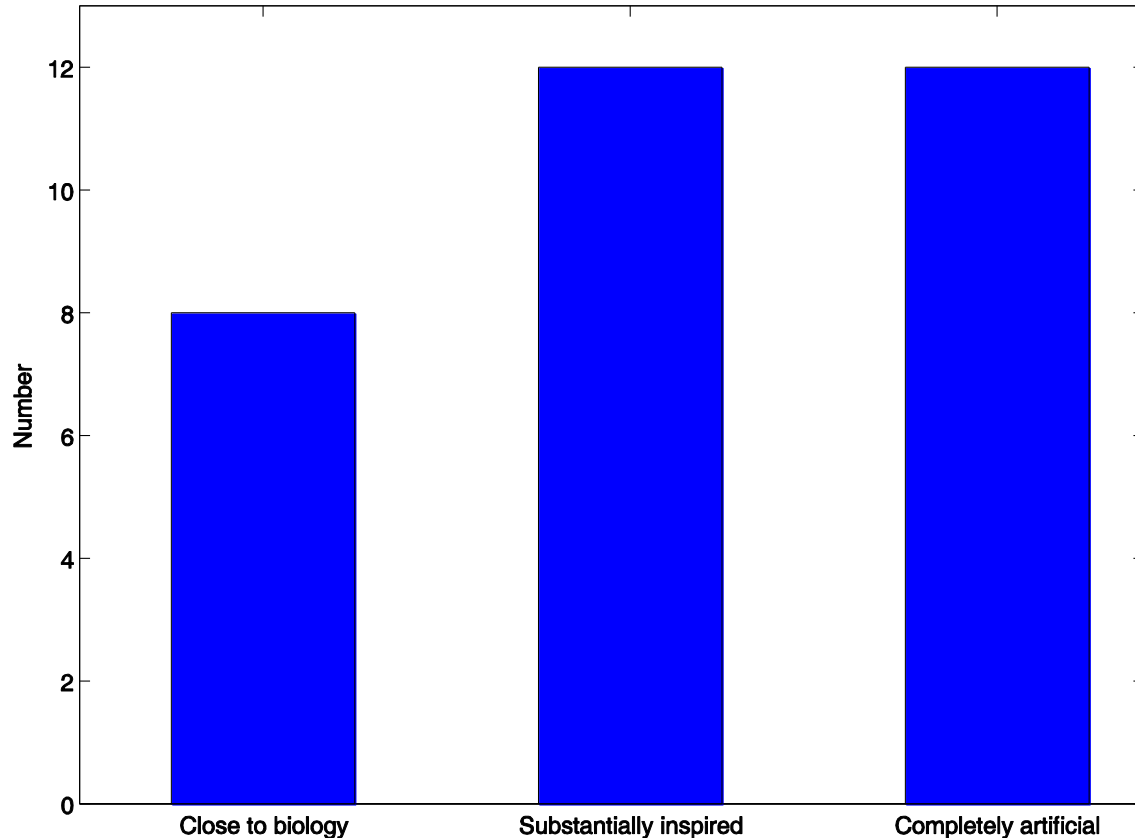
- Not with sufficient confidence
- No.
- Winning an Oxford union-style debate
- Worlds best chess playing AI was written by an AI
- Simple mammal WBE or general understanding of how to build general, flexible representations across many natural domains
- Emulation/development of mouse level machine intelligence
- Rat level intelligence

- Full dog emulation. Well-validated metrics of intelligence with linear projection to human level. Artificial computer scientists developing novel algorithms of high utility.
- Whole brain emulation of a monkey without legal regulation in place, chimp-level or crow-level AI using scientific theory and not WBE
- Monkey whole brain emulation with cognitive performance like monkey
- Ape level machine intelligence
- Whole brain emulation, semantic web
- Turing test or whole brain emulation of a primate
- Any of: uploading a mammal with carefully observed, realistic behavior. AI mathematicians as good as the best human mathematicians. Knowledge of where humans are on universal intelligence measure plus time sequence data showing AIs passing human level in 3 years of time, of comparable quality to the extrapolatable chess data. Drexlerian MNT confidently expected in 2 years time.
- Toddler AGI
- A combination of cognitive function simulation and interacting with the environment in one model/robot.
- If we can develop intelligence equivalent to a 7-8 year old child, we will reach adult level AI within 5 years
- An AI that is a human level AI researcher
- Agreement that we now understand the core algorithms used inside the human brain
- Passing Turing test *and* being able to demonstrate both logical processes along with creative/critical 'thinking'
- Gradual identification of objects: from an undifferentiated set of unknown size - parking spaces, dining chairs, students in a class - recognition of particular objects amongst them with no re-conceptualization
- Abstract skills transfer, systems bootstrap
- Large scale (10^{24}) bit quantum computing (assuming cost effective for researchers), exaflop per dollar conventional computers, toddler level intelligence.
- Already passed - otherwise such discussion amongst ourselves would not have been funded, let alone be intelligible, observable and accordable on this scale: as soon as such a thought is considered a 'reasonable' thought to have.

How similar will machine intelligence be to human intelligence?

5. How similar will the first human-level machine intelligence be to the human brain? *Please circle the most likely option.*

The three fields were marked: “Extremely close to biology (whole brain emulation / “uploading”)”, “Substantially inspired by biology”, “Completely artificial (human brains and artificial brains are related as birds are to planes)”.



Of the 32 responses to this question, 8 thought very biologically inspired machine intelligence the most likely, 12 thought brain-inspired AGI and 12 thought entirely de novo AGI was the most likely.

Statistically, this result is not distinguishable from an **even distribution** between the cases.

Field of work

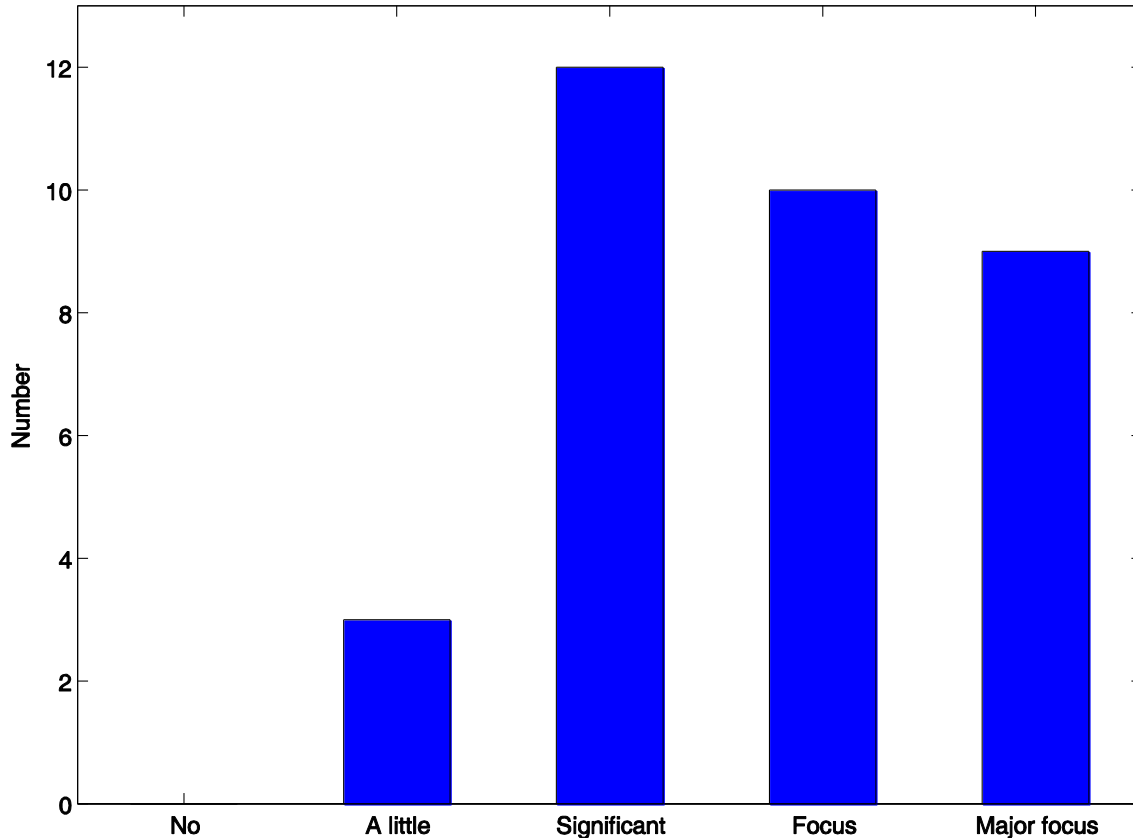
6. What field do you mainly work in? (*e.g., philosophy, cognitive science, machine intelligence, journalism...*)

The fields mentioned covered a lot of ground, with clear clusters of **philosophy** (6), **computer science and engineering** (8), **AI and robotics** (8) with the rest covering various academic disciplines (cognitive science, electronics, literature, mathematics, neuroscience, physics, psychology, sustainable development) as well as business.

Prior knowledge

7. Prior to this conference, how much have you thought about these issues? *Please circle the most likely option.*

The choices were: "Not at all", "A little, occasionally", "Significant interest", "Minor research focus / sustained study", and "Major research focus".



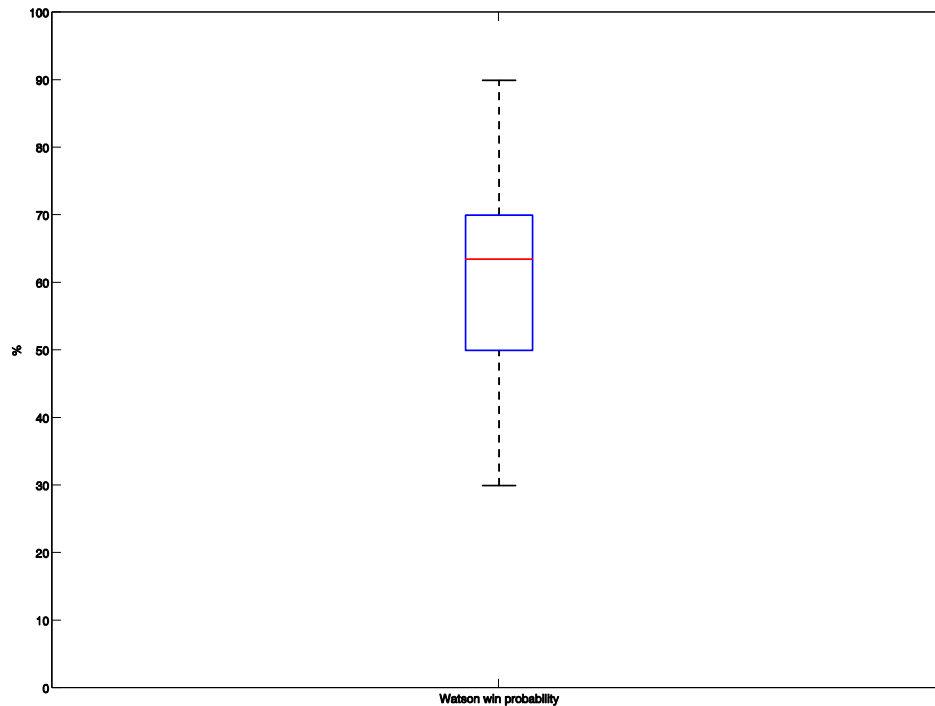
The survey participants exhibited a high degree of self-rated knowledge and interest. The mean level was 3.75 (close to "minor research focus/sustained study").

What is the probability that the Watson system will win over human expert contestants?

- On February 14-16, the IBM computer Watson will go up against the all-time human champions at Jeopardy, a popular television game-show dependent on puns and general trivia knowledge. How probable is it that Watson will win?

Most participants were **only mildly confident** in an eventual win (which did eventually happen). The minimum probability estimate was 30%, the median 64.5% (mean 60.32%), the 3rd quartile 70 and the maximum 90%¹.

¹ One erroneously asserted it had already won at the time of the survey, confusing a training match with the real match.



Correlations

Pair-wise correlations were calculated and their calculated. There were only three correlations with $p < 0.05$ significance after Bonferroni correction for multiple comparisons².

There was a significant 0.8 correlation between the 10% date and thinking AI would have neutral consequences (people with later dates were more likely to assume neutrality).

There was a significant 0.64 correlation between open source development and an extremely good outcome, and 0.66 between other development and a neutral outcome.

Group differences

Respondents were divided into the groups Philosophy, AI and robotics, General computer science and software engineering, and General academia and business.

There were no significant (as per ANOVA) inter-group differences in regards to who would develop AI, the outcomes, type of AI, expertise, or likelihood of Watson winning. Merging the AI and computer science group and the philosophy and general academia group did not change anything: participant views did not link strongly to their formal background.

² While there were a number of individually significant correlations, in a dataset with 17 variables like this there will be 120 correlations in total and hence on average 6 spurious cases where the correlation reaches the 0.05 significance level by sheer chance. The Bonferroni correction increases the requirements for being significant to avoid spurious correlations.

While philosophers were most sceptical (median 50%) and general academia most positive (median 70%) there were no significant group differences on whether Watson would win.

Discussion

This survey was merely an informal polling of an already self-selected group, so the results should be taken with a large grain of salt. The small number of responses, the presence of visiting groups with presumably correlated views, the simple survey design and the limitations of the questionnaire all contribute to make this of limited reliability and validity.

The main views expressed by the participants appear to be: human-level machine intelligence, whether due to a de novo AGI or biologically inspired/emulated systems, has a macroscopic probability to occurring mid-century. This development is more likely to occur from a large organisation than as a smaller project. The consequences might be potentially catastrophic, but there is great disagreement and uncertainty about this – radically positive outcomes are also possible.

It is interesting to compare this with the results reported in (Baum, Goertzel & Goertzel 2011) where a survey was conducted at the AGI09 conference. The AGI09 survey placed the median estimates for 10% chance of succeeding at the Turing test, passing third grade and doing Nobel quality work at 2020, the 50% chance at 2040 (Turing test), 2030 (third grade), and 2045 (Nobel), the 90% limit at 2075 (Turing test), 2075 (third grade) and 2100 (Nobel). The participants in our survey were clearly somewhat more pessimistic, especially at the upper end. Estimates of risk to humanity had a very broad spread and lack of consensus, similar to this survey. Estimates of which approach would be most likely to achieve HLMI strongly favoured integrative designs combining multiple paradigms, in distinction from the flat distribution in this survey (however, the AGI09 survey listed a larger number of specific methods).

(Baum, Goertzel & Goertzel 2011) also mention the results of a survey done by Bruce Klein about when AI will surpass human level intelligence (Klein 2007). This survey, dominated by AI optimists, had a median in the 2030-2050 range. It thus appears that the results of our survey are compatible with other survey attempts, but is somewhat more pessimistic on the timescale (if not on the general feasibility of AGI).

References

Baum, Seth D., Ben Goertzel, and Ted G. Goertzel. "How long until human-level AI? Results from an expert assessment". *Technological Forecasting & Social Change*, forthcoming, DOI 10.1016/j.techfore.2010.09.006. http://sethbaum.com/ac/fc_AI-Experts.pdf

B. Klein, When will AI Surpass Human-Level Intelligence? AGI-World, August 5, 2007, <http://www.novamente.net/bruce/?p=54> (accessed 24 Jan 2011).

Informal poll on machine intelligence

This poll will be used to informally elicit this conference's views about the future of machine intelligence. Answers will be treated anonymously, and compiled statistics will be published as a technical report on the FHI website.

Define a human-level machine intelligence to be one that can substitute for humans in virtually all cognitive tasks, including those requiring scientific creativity, common sense, or social skills.

1. Assuming no global catastrophe halts progress, by what year would you assign a 10%/50%/90% chance of the development of human-level machine intelligence? Feel free to answer 'never' if you believe such a milestone will never be reached.

	10%	50%	90%
Year reached:			

2. What type of organisation is most likely to first develop a human-level machine intelligence? Assume here that such a development is possible. *Please write down probabilities for each type getting there first.*

Industry	Academia	Military/ Government	Small independent group or individual	Open source project	Other (please specify) _____ _____

3. How positive or negative are the ultimate consequences of the creation of a human-level (and beyond human-level) machine intelligence likely to be? *Please write down probabilities for each consequence type.*

Extremely good	On balance good	More or less neutral	On balance bad	Extremely bad/ existential catastrophe

4. Can you think of any **milestone** such that if it were ever reached you would expect human-level machine intelligence to be developed **within five years thereafter**?

5. How similar will the first human-level machine intelligence be to the human brain?
Please circle the most likely option.

**Extremely close to biology
(whole brain emulation /
“uploading”)**

**Substantially inspired by
biology**

**Completely artificial (human
brains and artificial brains are
related as birds are to planes)**

6. What field do you mainly work in? (*e.g., philosophy, cognitive science, machine intelligence, journalism...*)

7. Prior to this conference, how much have you thought about these issues? *Please circle the most likely option.*

Not at all

**A little,
occasionally**

**Significant
interest**

**Minor research
focus / sustained
study**

**Major
research focus**

8. On February 14-16, the IBM computer Watson will go up against the all-time human champions at Jeopardy, a popular television game-show dependent on puns and general trivia knowledge. How probable is it that Watson will win?

_____ %

Thanks!