The University of Georgia

AI Newsletter

Institute for Artificial Intelligence
The University of Georgia
Athens, GA 30602-7415 U.S.A.

Winter 2015

Super-AI years away, if ever, but people still worry

2014 was an unusual year for artificial intelligence, and 2015 is shaping up to be more of the same. It's not that a grand theoretical breakthrough occurred. AI has advanced steadily for several years, and it turns out that many of the fundamental advances having an impact today were made years ago.

No, what made 2014 significant is the way AI appeared to make advances into society's collective consciousness. Perhaps because of the presence of real-time speech understanding, machine translation, and image recognition in consumer products, and also because of a flurry of purchases of AI and robotics companies by Google and Facebook, AI-related technologies seemed to suddenly become very noticeable. There was, we think, a collective realization in 2014 that AI is out of the lab and having a pronounced effect on how people live. In a sense, AI has become a part of ordinary life.

People have begun to think about what might happen next. Some people—in cases, very influential people—have even expressed worry, warning that AI poses an "existential threat" to humanity, and this worry has been amplified by the press.

Below we recount some of the more noteworthy news items from the year. Afterwards, we give our own take, and we present results of an impromptu survey we distributed via email to students, faculty, alumni, and others associated in some way with the AI Institute.

What Elon Musk Said

Elon Musk, CEO of Tesla and SpaceX, is arguably the person who touched off the flurry of recent articles on AI as potential threat. When asked during an October 2014 interview at MIT whether he had plans to enter the AI field, Musk's response raised eyebrows.[1]

"I think we should be very careful about artificial intelligence. If I had to guess at what our biggest existential threat is, it's probably that. So we need to be very careful with artificial intelligence. I'm increasingly inclined to think that there should be some regulatory oversight, maybe at the national and international level, just to make sure that we don't do something very foolish."

"With artificial intelligence we're summoning the demon. You know those stories where there's the guy with the pentagram, and the holy water, and he's like—Yeah, he's sure he can control the demon? Doesn't work out."

In an earlier (June) interview on CNBC, Musk indicated that he wanted to **"to keep an eye on"** AI. **"It is sort of like the Monty Python thing: Nobody expects the Spanish inquisition. It's just — you know, but you have to be careful."**

Musk would later donate $10,000,000 to the Future of Life Institute (FLI), an organization "working to mitigate existential risks facing humanity". The FLI's website states that human-level AI is currently its primary focus. Musk, physicist Stephen Hawking, AI researcher Stuart Russell, and philosopher Nick Bostrom are all on its advisory board.

Elon Musk

The Spanish Inquisition

---

[1] http://webcast.amps.ms.mit.edu/fall2014/AeroAstro/index-Fri-PM.html

What Stephen Hawking Said

Hawking and the FLI also made AI-related news in 2014. Specifically, in a December BBC interview, Hawking expressed his concerns about AI.[2]

"The primitive forms of artificial intelligence we already have, have proved very useful. But I think the development of full artificial intelligence could spell the end of the human race. Once humans develop artificial intelligence, it would take off on its own, and redesign itself at an ever increasing rate. Humans, who are limited by slow biological evolution, couldn't compete, and would be superseded."

Earlier in the year, Hawking, together with Stuart Russell and MIT physicists Frank Wilczek and Max Tegmark (Tegmark co-founded the FLI), wrote an opinion piece published on both The Huffington Post and in The Independent.[3] They indicate that dismissing super-intelligent machines as science fiction would be **"potentially our worst mistake in history."**

Stephen Hawking

"One can imagine such technology outsmarting financial markets, out-inventing human researchers, out-manipulating human leaders, and developing weapons we cannot even understand. Whereas the short-term impact of AI depends on who controls it, the long-term impact depends on whether it can be controlled at all."

To them, we're not taking the possibility seriously enough: **"If a superior alien civilisation sent us a message saying, "We'll arrive in a few decades," would we just reply, "OK, call us when you get here – we'll leave the lights on"? Probably not – but this is more or less what is happening with AI."**
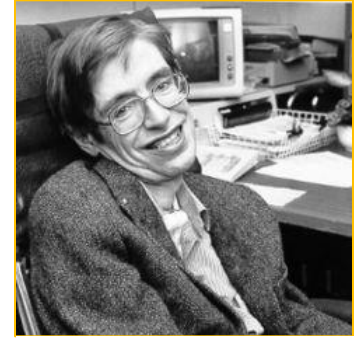
The Future of Life Institute's Open Letter

All four authors are also signatories of the FLI's "Research Priorities for Robust and Beneficial Artificial Intelligence: an Open Letter", published in January 2015.[4] The rather modest letter, which currently has over 6000 signatories (anyone can sign online), can be summed up with the following extract: **"Because of the great potential of AI, it is important to research how to reap its benefits while avoiding potential pitfalls."**

A 12-page accompanying document lists areas and risks that research should prioritize. Most of them are reasonable. Issues such as privacy, the ethics of autonomous weapons, and the potential effects of AI on labor markets are all discussed. Regarding long-term priorities, *control* is a significant theme. We need to ensure that the AI systems we develop function in predictable and controllable ways and cannot be hijacked by unauthorized parties. Only at the end of the document is an "intelligence explosion" mentioned, and even there the document appears to express some skepticism (or at least uncertainty) about whether it could actually happen. It's indicated that research into it would nevertheless be worthwhile.

What Nick Bostrom Said

Nick Bostrom, a University of Oxford professor and director of the Future of Humanity Institute (FHI), is also a central figure in the recent news. Bostrom has researched "existential risk" for years, and his 2014 book *Superintelligence: Paths, Dangers, and Strategies* discusses the probability and consequences of creating intelligences that far outstrip the capabilities of humans (i.e., superintelligences). In his view, there is some likelihood that it could happen during this century. Echoing Hawking and Musk, many of the possible scenarios result in the destruction of humanity. The book lays out ways in which superintelligence might arise, how it might run amok, and how its evolution might be controlled **"to achieve a survivable and beneficial outcome".** Bostrom thinks the issue is of paramount importance. **"This is quite possibly the most important and most daunting challenge humanity has ever faced. And—whether we succeed or fail—it is probably the last challenge we will ever face."**

What Bill Gates Said

Bill Gates, too, has expressed fears. When asked "How much of an existential threat do you think machine superintelligence will be…?"[5] Gates offered the following: **"I am in the camp that is concerned about super intelligence. First the machines will do a lot of jobs for us and not be super intelligent. That should be positive if we manage it well. A few decades after that though the intelligence is strong enough to be a concern. I agree with Elon Musk and some others on this and don't understand why some people are not concerned."**

---

[2] http://www.bbc.com/news/technology-30299992, http://www.bbc.com/news/technology-30290540

[3] http://www.huffingtonpost.com/stephen-hawking/artificial-intelligence_b_5174265.html

[4] http://futureoflife.org/misc/open_letter

[5] http://www.reddit.com/r/IAmA/comments/2tzjp7/hi_reddit_im_bill_gates_and_im_back_for_my_third/

What Marvin Minsky Said

Marvin Minsky, a founder of AI, also appears at least mildly concerned. When asked in 2014 whether AI would become "more than pre-programmed humanness", Minsky indicated that we can't predict it but that we should be prepared for several major advances. [6] "**Somewhere down the line, some computers will become more intelligent than most people.**"

"**I think some of the things we've been talking about sound like science fiction or fantasy or romance or whatever, but they're going to happen, or some of them are going to happen, and it would be good for more people to think more seriously about what could happen in the next 100 years and what options we should be aiming toward.**"
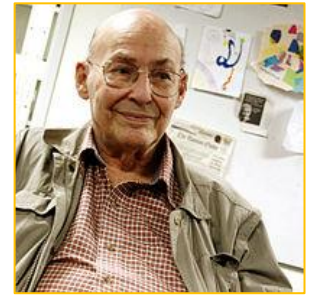
Marvin Minsky

The Atomic Bulletin: "Three Minutes and Counting"

Even the *Bulletin of Atomic Scientists* has a few things to say. In January 2015, it announced that it was moving the hands of the Doomsday Clock (created in 1947 to indicate how close humanity was to its destroying itself) forward 2 minutes to 11:57PM.[7] The primary reasons are the continued alteration of the Earth's climate and attempts by the US and Russia to modernize their respective nuclear arsenals. Nevertheless, "Emerging technological threats" also warranted a mention, and cyber-attacks (which could destabilize governments or economies) and artificial intelligence are the only two threats specifically named.

Our Own Survey, and a Few Comments

Prompted by these and similarly alarming articles, we surveyed our own students, faculty, alumni, and associates to hear what they thought. The survey questions are based in part on those asked by Nick Bostrom at a conference in 2014. They and the responses (we received a little over 70) are given below.

- Do you believe artificial intelligence will at some point pose a significant threat to the existence of human life?

    YES (32%), NO (68%)

- Do you believe artificial intelligence technologies, taken as a whole, ultimately will be beneficial or detrimental to society?

    Beneficial (93%), Detrimental (7%)

- By what year do you expect to see an artificial intelligence that can match human ability in almost all key cognitive tasks?

    Never (7%) 2025 (6%) 2050 (24%) 2075 (8%) 2100 (10%) 2125 (10%) 2150 (8%) 2200 (4%) >2200 (18%) No Opinion (4%)

- By what year do you expect to see a "superintelligence", that is "an intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills"?

    Never (27%) 2025 (0%) 2050 (6%) 2075 (11%) 2100 (6%) 2125 (4%) 2150 (4%) 2200 (9%) >2200 (23%) No Opinion (10%)

- Do you believe it is now worthwhile to fund initiatives to research potential negative societal effects of AI?

    YES (66%), NO (34%)

Summarizing the responses, almost half expect to see human-level AI within 100 years. Fewer people think a superintelligence is possible (27% said never), but the majority appear to think that it will also happen at some point (though maybe in the distant future). And while a third of respondents indicate that AI poses an existential threat, and two thirds say that research to minimize negative effects is worthwhile, the vast majority (93%) think that AI will ultimately be beneficial to society.

It's not shown in the above numbers, but the current and former faculty members that responded appear less optimistic about the chances of human-level AI. Of the 11 respondents identifying themselves as faculty or former faculty, 3 (27%) said that human-level AI will never happen, 3 expect it by 2050, 1 (9%) by 2125, and another 3 said after 2200. As for a so-called superintelligence, 6 respondents (55%) think that it will never happen. 1 (7%) thinks it could happen by 2050, another 1 by 2100, and 2 think it will probably happen but not until after 2200. 1 respondent had no opinion about either human-level or a superintelligent AI. Regarding the risks of AI, 9 respondents (approximately 82%) indicated that AI will not pose an existential threat to humanity. All respondents indicated that AI technologies will ultimately be beneficial to society, while 7 (64%) indicated that research into minimizing negative effects is nevertheless worthwhile.

Though it's difficult to draw conclusions from an unscientific survey with a small response rate, the faculty responses at least run counter to the more dire warnings expressed by Musk, Hawking, and Bostrom.

---

[6] http://www.jpost.com/Business/Business-Features/For-artificial-intelligence-pioneer-Marvin-Minsky-computers-have-soul-352076

[7] http://thebulletin.org/three-minutes-and-counting7938

This doesn't mean we shouldn't think about where we're going with AI. One of the survey respondents commented, **"Excessive reliance on machines we do not fully understand is always hazardous."** Arguably, this is the real issue, and it perhaps explains why even some of those firmly convinced that real AI won't happen think it's nevertheless a good idea to study AI's negative effects. It's not unreasonable to think that advances in technologies, including AI, will cause changes to society detrimental to some, and that in cases we will be unable to predict how the systems we create will act. This happens all the time—no AI needed. None of that rises to the level of existential threat, however, and one can take these issues seriously while simultaneously dismissing the more sensational headlines about AI (and perhaps focusing on more pressing existential threats).

Letter from the Director

Hello again. Yes, we are working to get the newsletter out to you on a more regular basis but things have been pretty hectic since the summer (e.g., my daughter's wedding was in San Diego just before the fall semester started; it was hectic but fantastic). Consequently, the fall semester was sort of a blur for me starting out rushed and due to some time consuming class related issues as well as looking into how AI techniques could be used to create better random number generators. New graduate student Jesse Kuzy helped out with the random number problem a little during the summer and got us started in the right direction before becoming involved with a cool research project developing olfactory sensors directed by AI faculty member Dr. "Charlie" Li in Engineering. Shervin Shahryari, new graduate student recently awarded an Institute assistantship, is currently helping with methods to measure the quality of a generator (specifically, a multiple recursive generator). We hope to get established in this area just like we did in the Snake-In-The-Box (SIB) area. Speaking of the SIB, many MSAI alumni worked on this problem during their stay with us. Currently, soon to graduate MSAI student, Shahnawaz Khan is involved with the SIB problem. His new DNA approach has been successful in finding three new records and tying an existing record. The three new records are in dimensions 13, 15, and 17 with spread 5, 7, and 7, achieving snake lengths of 85, 57, and 103, respectively. He tied the dimension 9, spread 3 length of 63.
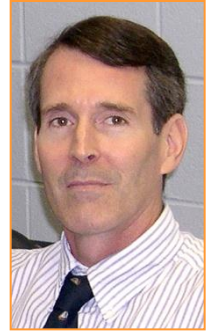
Many alumni played around with the harvest scheduling problem also. Peter Geiger is grounding his thesis on this problem as well and should be defending sometime in March. However, the breadth of topics that our current students are involved with is truly amazing. Topics include financial document sentiment analysis, diagnosing/analyzing speech disorders via acoustic phonetic analysis, robotic arm manipulation for collaborating robots, integrating logic programming with the semantic web on mobile devices, "Watson" style medical question answering, and solar radiation prediction to name just a few. Research publications stemming from these projects are outstanding and continue to bring fame and notoriety to the Institute.

Enrollment for the MSAI program continues to cruise along with about 20 to 25 active MSAI students at any given time. Undergraduate students continue to flock to the cognitive science major in record numbers. Exit (graduation) interviews reveal that undergraduates are very pleased with the cognitive science major, especially its interdisciplinary nature. This is good to hear plus it's good to hear that most have definite plans to pursue immediately after graduation. A little over half have jobs already in place or good prospects, while most of the others are keen to pursue graduate degrees including medical school.

As usual, I encourage you to let us hear from you. We are always keen to hear about your latest activities. Keep in touch!

How to Sponsor Research

As an industrial partner or associate of the IAI, you can be involved in our research in four ways:

Collaboration, especially student projects. It doesn't necessarily cost anything to work with us. We are always looking for good research topics, especially for master's theses.

Sponsored research. You can contract with us through the University of Georgia Research Foundation (UGARF) to do research for your company. *We are always looking for research sponsors.*

Donations. If you don't need confidentiality or a specific deliverable, you can simply make a gift to the IAI designated to support a particular research program. Donations are fully tax-deductible.

Consulting. You can hire faculty members or graduate students to work for you part-time as independent contractors. The University encourages this. It's a good way to do a small project with a high level of confidentiality.

The University of Georgia