

Great Ideas in Computing

Allan Borodin

University of Toronto CSC196
Fall 2025

Class 2: September 14 (2026)

Announcements and Todays Agends

Announcements

. A0 is posted on Quercus and the Home Page. I think the submission issue has been fixed. Has anyone subrmitted successfully?

Todays Agenda

- Unintended negative consequences of what seem to be great ideas.
- Some possible topics for the course
- Our first great idea: The von Neumann architecture

Great ideas may have negative consequences

When there are unintended, undesirable consequences for ideas we have come to accept as great ideas, is it still a great idea?

Great ideas may have negative consequences

When there are unintended, undesirable consequences for ideas we have come to accept as great ideas, is it still a great idea?

- If a technology becomes a standard, it can also become a barrier to innovation. For example, we will soon encounter our first great idea “The von Neumann” model. Some have argued that it has impeded progress on alternative computational architectures.
- Social networks allow for maintaining friendships. They also make possible the rapid spread of information and **mis-information**. It is claimed that social media companies deliberately fail to act on hate speech and political extremism while social media companies argue they are actively removing hate speech and violent discourse or that they are defending “free speech”.
- Moreover, some claim that these companies do not want to eliminate hate speech and that they often target information which in turn reinforces divisions in society. **Why do they want to target information?** (See the links to article by Starbird et al, Axios article on monitoring of AI with regard to hate speech, and a text file with some links to the issue of hate speech and bias on social media platforms.)³⁾ 15

Great ideas can have negative consequence continued

What are the ultimate dangers of AI?

Last year, I saw the following comment in one of the news feeds I look at:

Dozens of industry leaders and academics in the field of artificial intelligence have called for greater global attention to the possible threat of “extinction from AI.” A statement, signed by leading industry officials like OpenAI CEO Sam Altman and Geoffrey Hinton — the so-called “godfather” of artificial intelligence — highlights wide-ranging concerns about the ultimate danger of unchecked AI. Some experts say humanity is still a ways off from the prospect of science-fiction-like AI overlords, but the flood of hype and investment into the AI industry has led to calls for regulation now before any major mishaps occur. The growing AI arms race has already generated more immediate concerns. Lawmakers, advocacy groups and tech insiders have raised alarms about the potential dangers for AI-powered language models like ChatGPT.

Geoff Hinton's Warning and "How the Only Way to Survive Superintelligent AI"

The 'godfather of AI' reveals the only way humanity can survive superintelligent AI

See <https://www.cnn.com/2025/08/13/tech/ai-geoffrey-hinton>

Hinton has warned in the past that there is a 10% to 20% chance that AI wipes out humans. On Tuesday, he expressed doubts about how tech companies are trying to ensure humans remain "dominant" over "submissive" AI systems.

That's not going to work. They're going to be much smarter than us. They're going to have all sorts of ways to get around that," Hinton said at Ai4, an industry conference in Las Vegas.

In the future, Hinton warned, AI systems might be able to control humans just as easily as an adult can bribe 3-year-old with candy. This year has already seen examples of AI systems willing to deceive, cheat and steal to achieve their goals. For example, to avoid being replaced, one AI model tried to blackmail an engineer about an affair it learned about in an email.

Hinton's Solution

Instead of forcing AI to submit to humans, Hinton presented an intriguing solution: building “maternal instincts” into AI models, so “they really care about people” even once the technology becomes more powerful and smarter than humans.

Hinton said it's not clear to him exactly how that can be done technically but stressed it's critical researchers work on it.

“That's the only good outcome. If it's not going to parent me, it's going to replace me,” he said. “These super-intelligent caring AI mothers, most of them won't want to get rid of the maternal instinct because they don't want us to die.”

More Scary Warnings

In an Axios report, three Anthropic researchers went public (on September 7 or 8, 2026) with chilling concerns about out-of-control AI, warning it could destroy humans this decade.

Anthropic AI researcher Jacob Coxon wrote on X, after resigning yesterday to sound the alarm: "The people building AI earnestly believe that it could kill us all by the end of the decade. This is not a marketing stunt. If anything, many executives and senior researchers will couch their phrasing in the press to sound sensible - but I hear the same people express their fears privately. No other human activity poses this level of danger."

Anthropic alignment-science lead Evan Hubinger responded: "Jacob is correct here — we really do earnestly believe AI could kill all humans! I personally think it is $> 10\%$ within the next decade. I believe Anthropic is trying its best, but we do not yet have a plan to solve alignment for superintelligence and are not clearly on track to." (When I went to bed, this tweet had 1 million views. When I woke up, it had more than 10 million, and still soaring.)

Continuation of Scary Warning

Samuel Marks, Anthropic scalable-oversight lead, added: "AI developers believe their technology could cause human extinction (or similarly bad outcomes). This could happen in the next few years. In general, the more senior the employee, the more concerned they are."

Their warnings came just days after top OpenAI leaders, including CEO Sam Altman, said AI is speeding into a scary, uncontrollable phase.

Warnings from inside the AI giants point to an epic shared dilemma: Slow down and risk falling behind, or press ahead and risk losing control. The companies are full speed ahead, even as they practically beg for regulation or a global pause.

The big picture: Calling this unprecedented would be a gross understatement. You basically have the fastest-growing companies in human history warning their products could harm or even destroy humanity.

How Much is Possibly Hype

Critics say Anthropic and OpenAI are hyping their products to raise their valuations and invite regulation that would benefit them alone as the dominant incumbents.

But we (the Axios reporters) have been talking with dozens of people inside these companies for months, and they've sounded increasingly spooked and concerned. Given they see models not yet released to the public, it seems reckless not to take them seriously.

End of Axios article

I am posting (on the web page) an essay written by Dario Amodei, the CEO of Anthropic calling for a slowdown of the AI/LLM technology. Elon Musk and Sam Altman say they agree. The article appeared this weekend in the New York Times. It is worth reading the article

Can we control the technology? How much is hype? Haven't all *technological revolutions* come with dire warnings which didn't come to pass? What do you think?

Predictions and the nature of research

Niels Bohr, Mark Twain, Yogi Berra: “Predicting the future is hard because it hasn’t happened yet” and “it’s tough to make predictions, especially about the future” Who knows who said it first?

- Predictions about computing and what and can and can’t be done within some predicted time frame are often wrong. I will post a link to the 1955 Dartmouth summer project on artificial intelligence [AI](#) (where the term seems to have first appeared). Turing’s 1950 article provided what is now called the Turing Test as to “whether or not it is possible for a machine to show intelligent behaviour”. The Dartmouth project suggests that the indicated challenges for AI could be done over the summer.
- I will post a link to an article by John Backus about the view (that he proved wrong) that source level languages could never be nearly as efficient as machine code.
- There is also an article giving what one individual calls the “7 Worst Tech Predictions of All Time.” But note that the quote “I think there is a world market for maybe five computers”, attributed to T.J. Watson may never have happened.

Possible Topics

- What responsibilities do computer professionals have for the impact (and possible misuse) of the technology?
- What is a computer? The von Neumann architecture. Digital vs analogue. What were the alternatives? What else is possible (parallel, quantum)?
- The genius of Alan Turing; A mathematical definition of *computable functions*. Interpreters. Non computable functions.
- How did computers and computing become a commodity? The amazing advances in software and algorithms came along with advances in hardware (cost, speed, memory size, physical size, power) and communication (cost, capacity and speed).
- “Killer applications” such as word processors, email, search engines, navigation systems, games).
- The internet; packet routing. TCP/IP.

More possible topics

- Fortran, the first commercial source level language (with an efficient compiler). First compiler often attributed to Grace Hopper. John Backus vs the prevailing view that compiled code would be too slow compared to machine code.
- How search engines work? Will search engines survive now that we have large language models?
- A local great idea: NP completeness. What is and what is not *efficiently computable*.
- Complexity based cryptography; public key cryptography; digital signatures. Captchas.
- Another local great idea: deep neural networks and the success of machine learning (ML). What (if anything) is the limitation of ML?
- HCI (human computer interaction). The mouse. Menus, click, paste and drag. Visualization.

And some more possible topics

- Information theory: the genius of Claude Shannon. Error correcting codes; compression.
- Social networks and the spread of information (and mis-information, conspiracies, etc). Targetting information to different communities. (What is a social network community?)
- Open Source. Wikipedia. Blogs
- Relational data bases.
- General methods for solving optimization problems; for example, greedy algorithms, local search linear programming and dynamic programming. Dijkstra's algorithm and navigation systems.
- Distributed System primitives : mutual exclusion, consensus
- Differential privacy; extracting useful statistical information without sacrificing individual information.
- Algorithmic mechanism design; automated auctions. Algorithmic social choice; (e.g., voting, fair allocation).

What great ideas (past or current) have we missed?

Our first great idea

Our first topic/great idea for discussion: Digitilization and the von Neumann architecture.

- By the mid 40s the first computers were being built. (I am not going to talk say about Babbage and the recent implementation of Babbages 1850 machine.)
- The earliest computers had fixed programs and were not general purpose machines.
- The stored program computer. The conceptual idea is associated with von Neumann who first described the model in a paper "First Draft of a Report on the EDVAC" dated June 30, 1945.
- It is clear that the stored program idea is present in Turing's 1936 paper which we will discuss later. But the Turing model was not meant to be a model of an actual computer.

We ended the Monday class just mentioning the last slide. This is where we will begin the next class.