THE WALL STREET JOURNAL.

http://www.wsj.com/articles/whats-next-for-artificial-intelligence-1465827619

LIFE | IDEAS | THE FUTURE OF EVERYTHING

What's Next for Artificial Intelligence

The best minds in the business—Yann LeCun of Facebook, Luke Nosek of the Founders Fund, Nick Bostrom of Oxford University and Andrew Ng of Baidu—on what life will look like in the age of the machines
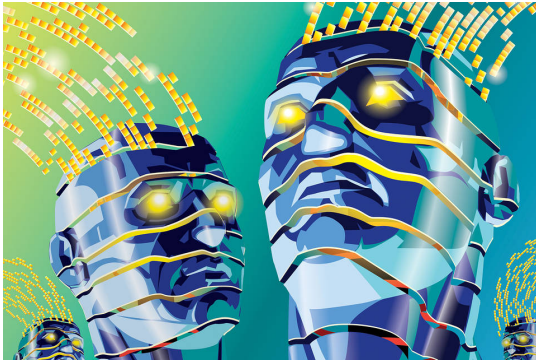

ILLUSTRATION: DANIEL HERTZBERG

Updated June 14, 2016 1:14 a.m. ET

HOW DO YOU TEACH A MACHINE?

Yann LeCun, director of artificial-intelligence research at Facebook, on a curriculum for software

The traditional definition of artificial intelligence is the ability of machines to execute tasks and solve problems in ways normally attributed to humans. Some tasks that we consider simple—recognizing an object in a photo, driving a car—are incredibly complex for AI. Machines can surpass us when it comes to things like playing chess, but those machines are limited by the manual nature of their programming; a $30 gadget can beat us at a board game, but it can't do—or learn to do—anything else.

This is where machine learning comes in. Show millions of cat photos to a machine, and it will hone its algorithms to improve at recognizing pictures of cats. Machine learning is the basis on which all large Internet companies are built, enabling them to rank responses to a search query, give suggestions and select the most relevant content for a given user.

MORE FUTURE OF EVERYTHING »

- The Next Frontier in Nordic Vacationing? Norway Itself
- What Are You Printing for Dinner?
- Can Bitcoin Technology Solve the Migrant Crisis?
- The Future of Everything: June 2016

Deep learning, modeled on the human brain, is infinitely more complex. Unlike machine learning, deep learning can teach machines to ignore all but the important characteristics of a sound or image—a hierarchical view of the world that accounts for infinite variety. It's deep learning that opened the door to driverless cars, speech-recognition engines and medical-analysis systems that are sometimes better than expert radiologists at identifying tumors.

Despite these astonishing advances, we are a long way from machines that are as intelligent as humans—or even rats. So far, we've seen only 5% of what AI can do.

IS IT TIME TO RETHINK YOUR CAREER?

Andrew Ng, chief scientist at Chinese Internet giant Baidu, on how AI will impact what we do for a living

Truck driving is one of the most common occupations in America today: Millions of men and women make their living moving freight from coast to coast. Very soon, however, all those jobs could disappear. Autonomous vehicles will cover those same routes in a

faster, safer and more efficient manner. What company, faced with that choice, would choose expensive, error-prone human drivers?

There's a historical precedent for this kind of labor upheaval. Before the Industrial Revolution, 90% of Americans worked on farms. The rise of steam power and manufacturing left many out of work, but also created new jobs—and entirely new fields that no one at the time could have imagined. This sea change took place over the course of two centuries; America had time to adjust. Farmers tilled their fields until retirement, while their children went off to school and became electricians, factory foremen, real-estate agents and food chemists.

'We're about to face labor displacement of a magnitude we haven't seen since the 1930s. '

Truck drivers won't be so lucky. Their jobs, along with millions of others, could soon be obsolete. The age of intelligent machines will see huge numbers of individuals unable to work, unable to earn, unable to pay taxes. Those workers will need to be retrained—or risk being left out in the cold. We could face labor displacement of a magnitude we haven't seen since the 1930s.

In 1933, Franklin Roosevelt's New Deal provided relief for massive unemployment and helped kick-start the economy. More important, it helped us transition from an agrarian society to an industrial one. Programs like the Public Works Administration improved our transportation infrastructure by hiring the unemployed to build bridges and new highways. These improvements paved the way for broad adoption of what was then exciting new technology: the car.

We need to update the New Deal for the 21st century and establish a trainee program for the new jobs artificial intelligence will create. We need to retrain truck drivers and office assistants to create data analysts, trip optimizers and other professionals we don't yet know we need. It would have been impossible for an antebellum farmer to imagine his son becoming an electrician, and it's impossible to say what new jobs AI will create. But it's clear that drastic measures are necessary if we want to transition from an industrial society to an age of intelligent machines.

AI: JUST LIKE US?
How intelligent machines could resemble their makers

The next step in achieving human-level ai is creating intelligent—but not autonomous—machines. The AI system in your car will get you safely home, but won't choose another destination once you've gone inside. From there, we'll add basic drives, along with emotions and moral values. If we create machines that learn as well as our brains do, it's easy to imagine them inheriting human-like qualities—and flaws. But a "Terminator"-style scenario is, in my view, immensely improbable. It would require a discrete, malevolent entity to specifically hard-wire malicious intent into intelligent machines, and no organization, let alone a single group or a person, will achieve human-level AI alone. Building intelligent machines is one of

The greatest scientific challenges of our times, and it will require the sharing of ideas across countries, companies, labs and academia. Progress in AI is likely to be gradual—and open. *—Yann LeCun*

HOW TO MASTER THE MACHINES
Nick Bostrom, founding director of the Future of Humanity Institute at Oxford University, on the existential risk of AI. Interviewed by Daniela Hernandez.

Can you tell me about the work you're doing?

We are interested in the technical challenges related to the "control problem." How can you ensure that [AI] will do what the programmers intended? We're also interested in studying the economic, political and social issues that arise when you have these superintelligent AIs. What kinds of political institutions would be most helpful to deal with this transition to the machine- intelligence era? How can we ensure that different

ILLUSTRATION: DANIEL HERTZBERG

stakeholders come together and do something that can lead to a good outcome?

Much of your work has focused on existential risk. How would you explain that to a 5-year-old?

I would say it's technology that could permanently destroy the entire future for all of humanity. For a slightly older audience, I would say there's the possibility of human extinction or the permanent destruction of our potential to achieve value in the future.

What are some of the strategies you think will help mitigate the potential existential risks of artificial intelligence?

Work on the control problem could be helpful. By the time we figure out how to make machines really smart, we should have some ideas about how to control such a thing, how to engineer it so that it will be on our side, aligned with human values and not destructive. That involves a bunch of technical challenges, some of which we can start to work on today.

Can you give me an example?

There are different ideas on how to approach this control problem. One line of attack is to study value learning. We would want the AI we build to ultimately share our values, so that it can work as an extension of our will. It does not look promising to write down a long list of everything we care about. It looks more promising to leverage the AI's own intelligence to learn about our values and what our preferences are.

Values differ from person to person. How do we decide what values a machine should learn?

Well, this is a big and complicated question: the possibility of profound differences between values and conflicting interests. And this is in a sense the biggest remaining problem. If you're optimistic about technological progress, you'll think that eventually we'll figure out how to do more and more.

We will conquer nature to an ever-greater degree. But the one thing that technology does not automatically solve is the problem of conflict, of war. At the darkest macroscale, you have the possibility of people using this advance, this power over nature, this knowledge, in ways designed to harm and destroy others. That problem is not automatically solved.

How might we be able to deal with that tension?

I don't have a simple answer to that. I don't think there's an easy technofix.

Wouldn't a self-programming agent be able to free itself from the shackles of the control systems under which we place them? Humans do this all the time already, to some extent, when we act selfishly.

The conservative assumption would be that the superintelligent AI would be able to reprogram itself, would be able to change its values, and would be able to break out of any box that we put it in. The goal, then, would be to design it in such a way that it would choose not to use those capabilities in ways that would be harmful to us. If an AI wants to serve humans, it would assign a very low expected utility to an action that would lead it to start killing humans. There are fundamental reasons to think that if you set up the goal system in a proper way, these ultimate decision criteria would be preserved.

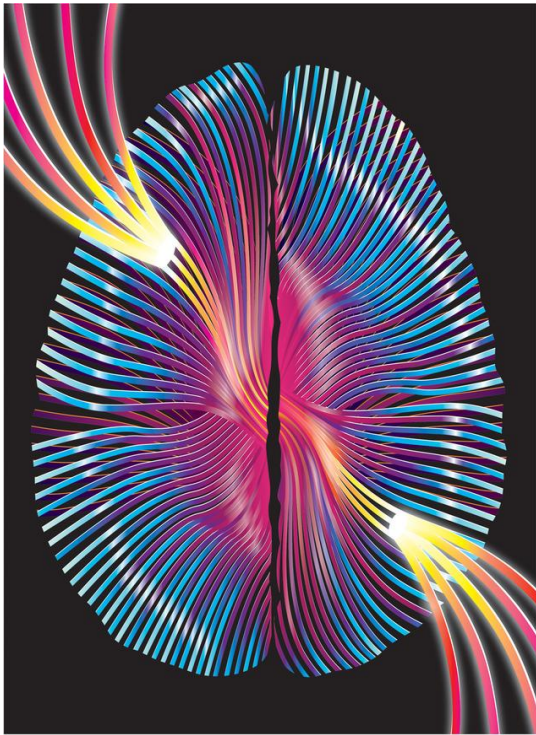## LET'S IMPROVE THE MINDS WE HAVE

ILLUSTRATION: DANIEL HERTZBERG

Luke Nosek, co-founder of PayPal and the Founders Fund, on the need to train our brains before the artificial ones arrive

Earlier this year, the korean Go champion Lee Sedol played a historic five-game match against Google's AlphaGo, an artificially intelligent computer program. Sedol has 18 world championships to his name. On March 19, 2016, he lost to software.

High-performance computing today is unprecedentedly powerful. Still, we remain stages away from creating an artificial general intelligence with anywhere near the capabilities of the human mind. We don't yet understand how general, human-level AI (sometimes referred to as AGI, or strong AI) will work or what influence it will have on our lives and economy. The scale of impact is often compared to the advent of nuclear technology, and everyone from Stephen Hawking to Elon Musk to the creator of AlphaGo has advised that we proceed with caution.

The nuclear comparison is charged but apt. As with nuclear technology, the worst-case scenario for strong AI—malevolent superintelligence turns on humanity and tries to kill it—would be globally devastating. Conversely, the optimistic predictions are so blindingly positive (universal economic prosperity, elimination of disease) that we may be biased by both undue fear and optimism.

Strong AI could help billions of people lead safer, healthier, happier lives. But to design this machine, engineers will need a better understanding—greater than that of anyone alive today—of the complex social, neurological and economic realities faced by a society of intelligent humans and machines. And if we upgrade the minds we already have, we'll be better equipped to conceptualize, build and coexist with strong AI.

We can divide the enhancement of human intelligence into three stages. The first, using technology like Google Search to augment and supplement the human mind, is well under way Compare a fifth-grader with a library card in 1996 to a fifth-grader on the Google home page in 2016—just keystrokes from much of human knowledge.

If stage one involves supplementing the mind with technology, then stage two is about amplifying the mind directly. Adaptive learning software personalizes education and makes adjustments to lessons in real time. If a student is excelling, the pace will increase. If he or she is struggling, the program might slow down, switch teaching styles or signal to the instructor that assistance is needed. Adaptive learning and online education could mean the end of one-size-fits-all education. Integration with virtual and augmented reality could also amplify intelligence in unexpected ways.

Stage three of intelligence enhancement involves a fundamental transformation of the mind. Transcranial magnetic stimulation, or TMS, is a noninvasive, FDA-approved treatment in which an electromagnetic coil is applied to the head. TMS is currently being used to treat post-traumatic stress disorder, autism and drug-resistant major depression. Sample sizes at such facilities as the Brain Treatment Center in Newport Beach, Calif., and the University of Louisville in Kentucky are small and the duration of impact unknown, but high percentages of individuals—up to 90% for a trial with 200 higher-functioning autistic patients—have shown improvement. Initial signs indicate that TMS could be effective for a wide, seemingly unrelated range of neurological conditions. If we can positively affect injured or non-neurotypical brains, we may not be

far from improving connections in healthy brains and enhancing intelligence in a generalized way.

Strong AI appears to be on the horizon, but for now the human mind is the only one we have. Enhancing our own intelligence is the first step toward creating—and successfully coexisting with—the intelligent machines of the future.

YOU CAN'T TEACH (MACHINES) COMMON SENSE

At least not yet. And it's the biggest barrier to true artificial intelligence.

Predictive learning, also called unsupervised learning, is the principal mode by which animals and humans come to understand the world. Take the sentence "John picks up his phone and leaves the room." Experience tells you that the phone is probably a mobile model and that John made his exit through a door. A machine, lacking a good representation of the world and its constraints, could never have inferred that information. Predictive learning in machines—an essential but still undeveloped feature —will allow AI to learn without human supervision, as children do. But teaching common sense to software is more than just a technical question—it's a fundamental scientific and mathematical challenge that could take decades to solve. And until then, our machines can never be truly intelligent. —*Yann LeCun*