In medio stat virtus

Jean-Marc Alliot

1 Slide 1

In Medio Stat Virtus

When fashion is not that good...

Every speaker is always afraid of making a bad speech. When my friend Professor Vu Duong asked me to do the first talk of this conference, I asked him what he wanted me to talk about. And the answer let me a little bit puzzled: “anything you want” he told me. “It’s like an overture, so do as you please”. Well, so I had then two opportunities to miss the point. First by choosing the wrong topic, and then by making a bad speech.

Thus I had a look at all the papers of the conference, and there was clearly two possibilities for me: I could of course speak about ATM, but I could also talk about Artificial Intelligence.

I know a little bit about ATM because I had been in the business for 25 years and even ended as head of the R&D department of the french DSNA. But I am not very much in this business now, and haven’t been for the last ten years, as I am now back in academic research. On the opposite, I never left the AI field since my PhD in the late

eighties, and even wrote a few books about it.

What fascinates me in AI is not the idea of artificial intelligence by itself, because I don't believe in it, but the hype that has always surrounded it. And what fascinates me even more is the fact that AI has been a battlefield between different factions that have done their best to kill their opponents, and almost succeeded in some cases. Of course in ATM we had also our fair share of controversies; for example we had in the eighties and nineties the opposition between the full automation concept of the AERA project and the "controller in the loop" concept based on the cognitive model of the controller. And we had many others afterwards.

The most interesting part in those "scientific" wars is that they are usually not scientific. And the other interesting part is the fact that the winner is not always the one whose position is the most scientifically sound. Of course, usually, following the old saying, "time will tell". But it can be a very long before time tells.

2 Slide 2

Beware Mortals

The end of humanity is now close… (2015)

- Stephen Hawking: "Artificial intelligence could spell the end of the human race."
- Elon Musk: "I think we should be very careful about artificial intelligence. If I were to guess like what our biggest existential threat is, it's probably that."
- Bill Gates: « 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."

Today the hype is all about Deep Learning and Machine Learning. And even the end of humanity seems close if we read what some of our great minds are saying.

3 Slide 3

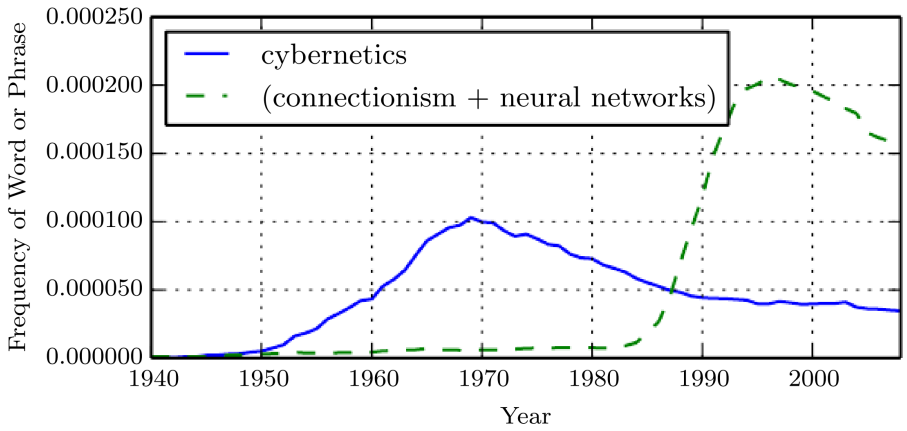
Not exactly new...

- H. L. Mencken: « The whole aim of practical politics is to keep the populace alarmed (and hence clamorous to be led to safety) by menacing it with an endless series of hobgoblins, all of them imaginary."
- ….which remains the motto of any good lobbyist….

Well that's not exactly a new argument to get funding. But it is an argument which works quite well.

So, having no scientific cutting edge result to present, I decided that I was going to play my part as an old vet of both ATM and AI research, and talk about the dangers of being too partial and narrow minded, and the necessity of keeping a balanced point of view, all of this based on an historical perspective. That's the kind of talk that old people do for young people, and, let's being honest, the kind of talk young people usually don't listen to...

4 Slide 4



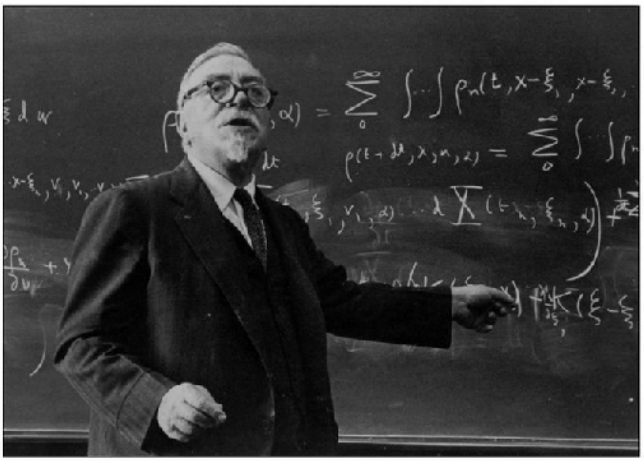
While preparing this talk, I was reading two books at the same time. One quite technical by Joshua Bengio on Deep Learning, while the other one was the biography of Norbert Wiener. And the two books collided quite strangely. Bengio presents that very interesting figure in his book. For Bengio, cybernetics is the grand father of the neural networks revolution. And as you might know, the inventor of cybernetics was Norbert Wiener.

5 Slide 5

Norbert Wiener

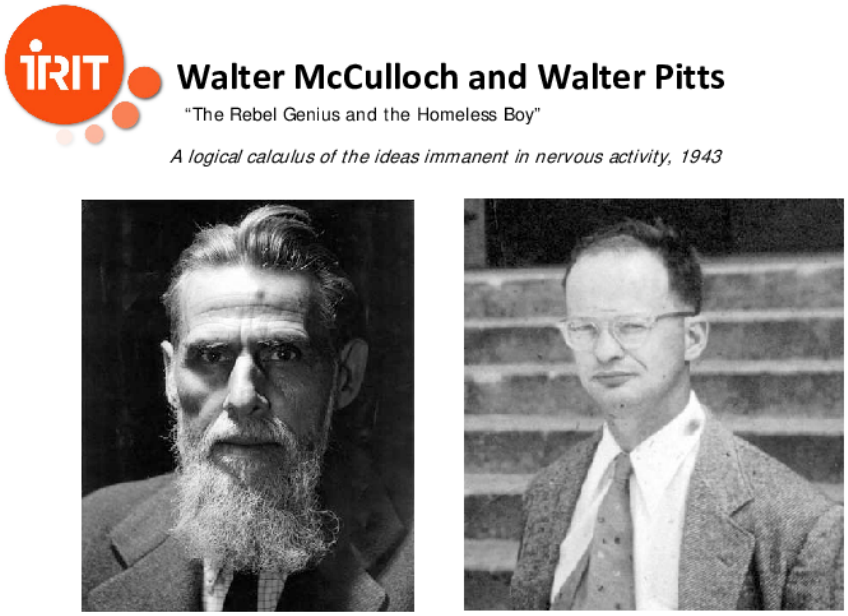
« Dark hero of the information age »

Behavior, Purpose and Teleology, Rosenblueth, Wiener and Bigelow , 1943



Wiener was a child prodigy, who had his PhD at 18, and became later an applied mathematician at MIT. Calling Wiener an applied mathematician is probably unfair, as Wiener was interested in a tremendous amount of subject, from philosophy (his early love), to philology, mathematics, biology... He made major contributions in automation and information theory, and his book "Cybernetics" that he published in 1948 was supposed to set the basis of a new science, which dealt mainly with self controlling machines. Cybernetics was in fact more a concept than a science. One of its main focus was the idea of feedback, and regarding computers, there was as much interest in their architecture as in their programming, a concept which was still extremely foggy at the time. Wiener organized a set of interdisciplinary conferences (the Macy conferences) to discuss the problems around cybernetics. Two of the people who aggregated Wiener's group were Warren McCulloch and Walter Pitts.

6 Slide 6



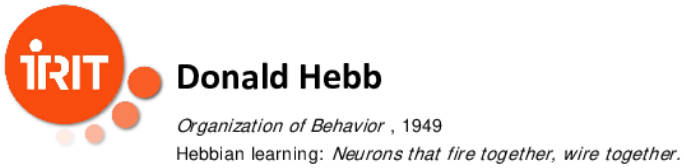
McCulloch was a neuropsychiatrist, a philosopher and a poet. He was coming from a well-to-do family, drank a lot, ate ice creams, was a radical who listened to Spanish revolutionary songs and lived a free life with his wife Rook. In 1942, he had attended one of the Macy conferences, and had been stricken by a presentation of Arturo Rosenblueth, a very close associate of Wiener, about feedback mechanism in the brain. Then he began to try to model the activity of the brain using a logical model. He was introduced to Walter Pitts by a common friend, Jeremy Lettvin, a MD student. When they met McCulloch was in his forties, while Pitt was only 19, and was doing menial jobs around Chicago university while following the group of Nicolas Rashevsky, a Russian mathematician who was trying to apply mathematics to biology.

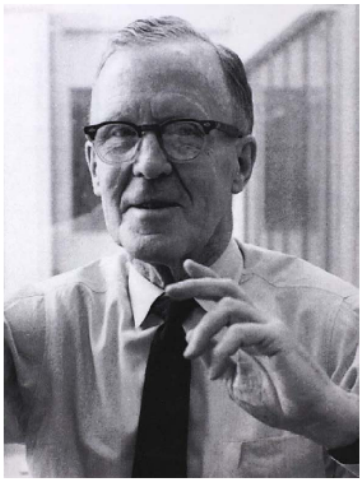
Pitt was the opposite of McCulloch. He was a runaway boy who had fled his family in his teens. He was self taught, and had written to Bertrand Russel after reading the Principia Mathematica in three days. In his letter, he pointed some problems, and Russell was so impressed that he offered him to come to England as his student, not knowing that Pitt was only 12 years old at the time. Pitt provided McCulloch with a mathematical background that he didn't have, and wrote most of the technical parts of their famous paper, which is considered as the foundation of the neural networks field. McCulloch, Pitt and Lettvin joined Wiener at MIT in 1951. Pitt then began writing his PhD under the supervision of Wiener, on the mathematics of 3D neural networks. Unfortunately, in 1953, Wiener cut all ties with Pitt and McCulloch for reasons which remained mystery for 50 years. Pitt never recovered, burned his unpublished results and his work on his PhD a few years later and died of cirrhosis at 46 in 1969. McCulloch

and Lettvin kept working on the neural system, but the neural network group at MIT was definitely damaged. McCulloch died also in 69, aged 71.

Cybernetics didn't survive Wiener, who died in 64, for many reasons. Wiener had a complex personality, with bouts of depression. He was capable to be extremely unpleasant and aggressive with colleagues, and John McCarthy, one of the founding father of AI with Marvin Minsky at MIT, said once in an interview that the name Artificial Intelligence was partly chosen to keep Wiener as far as possible to the new group, while Minsky knew of course very well the works of Wiener. Moreover Wiener was a pacifist, was opposed to collaboration with big companies and moreover cybernetics had become and extremely popular subject in the soviet union, and became painted in red for the defense department. So, while Minsky was able to secure tremendous funding for AI through the army, Wiener became quite isolated and his science died with him, at least in the Western World. Cybernetics institute on the other way remained in the Eastern block. The name "cybernetics institute" even appear in some of the science fiction work of Stanislas Lem for example. Strangely, the words "cyberspace" (mostly thanks to William Gibson's book Neuromancer) or even "cybersecurity" reappeared later, while the etymology of the word has not much to do with computers, as kybernetes in Greek means rudder or steersman, which has much more to do with the ideas of feedback and control.

7 Slide 7





However, in 1949, another brick had been laid by Canadian neuropsychologist

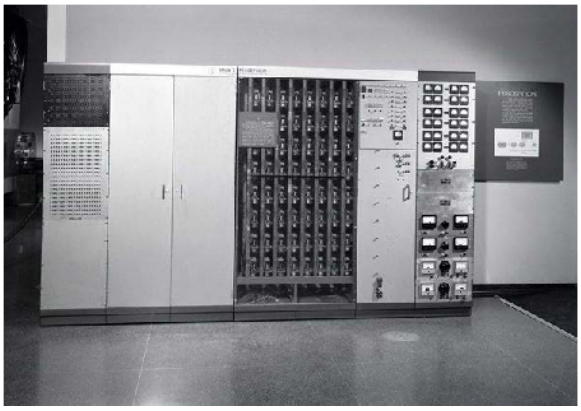
Donald Hebb, which is now known as Hebbian learning. The rule is simple: neurons that fire together, wire together. This is how reinforcement learning began gradually to appear in the theory of the modeling of brain.

8 Slide 8

Frank Rosenblatt

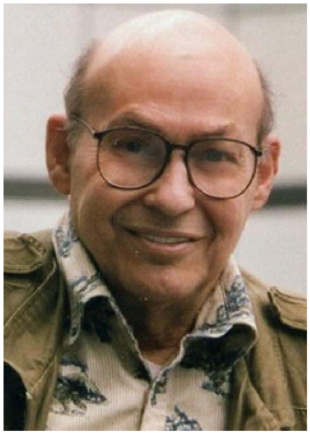
Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms, 1962

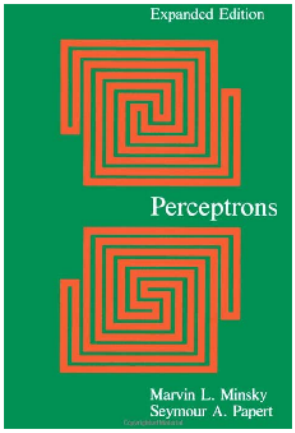
The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence (NY Times, 1958)



All these ideas were still popular in the fifties and the sixties and in 1962, Frank Rosenblatt published his book “Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms”. Rosenblatt was able to secure funding from the Navy and he built an ad-hoc machine, the Mark I perceptron. Perceptrons are binary classifier, and the Mark I was a machine designed for image recognition. The subject was an active one at the time, and there were many teams working on it in different institutions, until the publication of the famous “Perceptrons” book by Minsky and Papert in 1969.

9 Slide 9





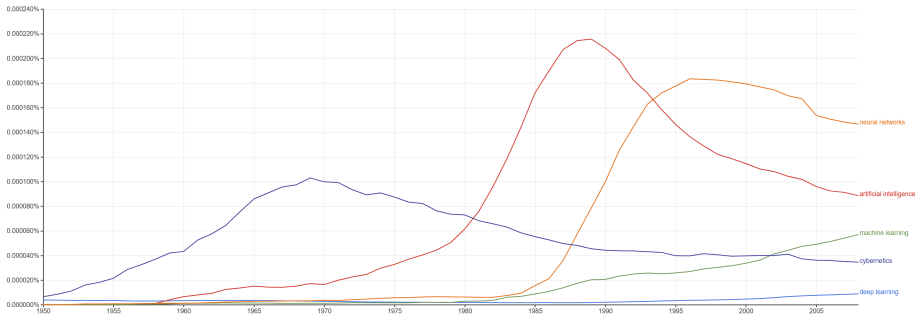
"Perceptrons" is one of the most discussed and controversial book in the history of the history of science. There have been a lot of articles written about it. One of the most interesting work on it was the PhD in sociology, and later the articles of Mikel Olazaran in the early nineties, when the crushing influence of Minsky was fading with the second AI winter.

Minsky is still universally well known. He was one of the founding fathers of AI, dominated the field at MIT for almost 50 years, and received a Turing award. He was an excellent salesperson, and loved being in the spotlight. He was the friend of many very famous people, including the infamous billionaire Jeffrey Epstein, and Minsky is one of the persons who has been formally accused of having sex with minors in Jeffrey Epstein's private island. Minsky died in 2016 and will probably remain in history as a very controversial personality.

Minsky had worked in his early years on neural networks. He knew the field, but as he had himself said, he never believed in McCulloch and Pitts ideas, while he recognized the influence of Rashevsky, one of Pitt's mentor. At the end of the sixties, Minsky and Papert, along with others such Newell, Simon, Shaw championed an approach to AI, that we might call cognitive AI, and whose goal was to create cognitive modelling of the brain, through sets of rules. The need for computing capacities was high, and the competition for Army funding was also very high. "Perceptrons" doesn't contain any formal error. But the book is misleading, and according to Olazaran, on purpose, starting with the cover itself, which shows two figures, one connected and the other not connected, that perceptrons were not able to properly classify at the time.

The book demonstrated some valid points, such as the impossibility for local neural networks to learn complex functions, and the need of fully connected, many layers, neural networks, which is correct. But the global perception of the book was that the whole approach behind perceptrons was doomed to fail.

10 Slide 10



"Perceptrons" killed the funding on research for neural networks for many years.

11 Slide 11

- Paul Werbos (1974) shows that backprop can be used to train neural networks
- La Jolla conference (1979) => Parallel Distributed Group at San Diego university (Rumelhart, McClelland)
- Rumelhart, Hinton & William 1986 paper shows that backprop can be used experimentally
- Yann LeCun proposes in its PhD (1987) the modern form of the backprop algorithm for multi-layers feedforward networks.

The field began to resurface in the eighties. There was a conference in La Jolla in 1979, and then the constitution of a group of researchers at San Diego university, with

David Rumelhart. Then, first with the work of Paul Werbos in 1974 and later with the articles of Rumelhart and his team, the backpropagation algorithm was shown to be able to train feedforward fully connected neural networks. In 1987, Yann LeCun in his PhD defined the modern form of backprop.

12 Slide 12

Perceptrons, second edition, 1988

- "We have the impression that many people in the connectionist community do not understand that this [back-propagation] is merely a particular way to compute a gradient and have assumed instead that back-propagation is a new learning scheme that somehow gets around the basic limitations of hill-climbing. ... **Virtually nothing has been proved about the range of problems upon which GD [the generalized delta rule, or back-propagation] works both efficiently and dependably**.... In the early years of cybernetics, everybody understood that hill-climbing was always available for working easy problems, but that it almost always became impractical for problems of larger sizes and complexities... The situation seems not to have changed much - we have seen no contemporary connectionist publication that casts much new theoretical light on the situation.... We fear that its [back-propagation's] reputation also stems from unfamiliarity with the manner in which hill-climbing methods deteriorate when confronted with larger-scale problems. In any case, **little good can come from statements like 'as a practical matter, GD leads to solutions in virtually every case' or 'GD can, in principle, learn arbitrary functions'**. Such pronouncements are not merely technically wrong; more significantly, the pretense that problems do not exist can deflect us from valuable insights that could come from examining things more carefully. As the field of connectionism becomes more mature, the quest for a general solution to all learning problems will evolve into an understanding of which types of learning processes are likely to work on which classes of problems."

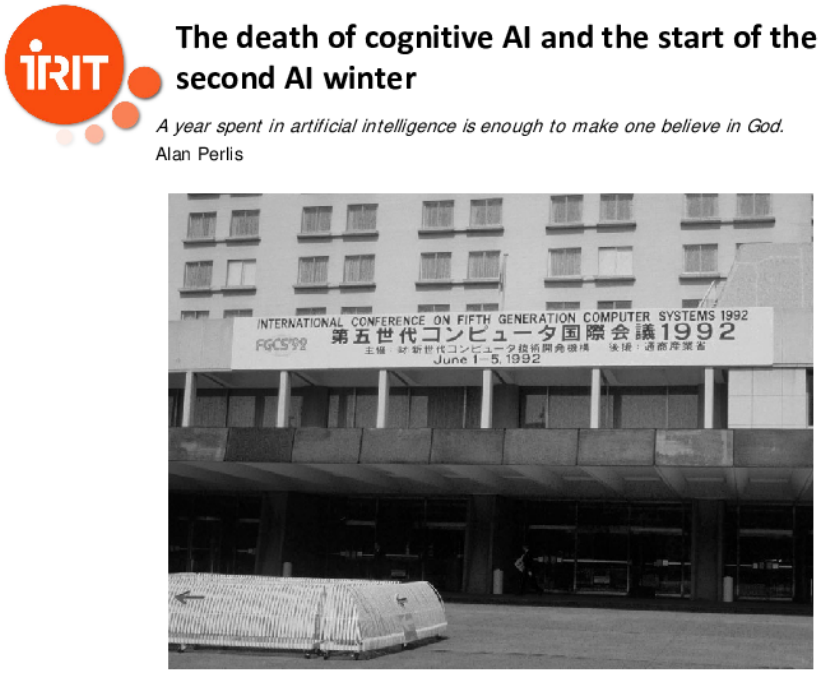
Minsky and Papert tried again to put the head of neural networks under the water in their second edition of perceptron in 1988:

> We have the impression that many people in the connectionist community do not understand that this [back-propagation] is merely a particular way to compute a gradient and have assumed instead that back-propagation is a new learning scheme that somehow gets around the basic limitations of hill-climbing. ... Virtually nothing has been proved about the range of problems upon which GD [the generalized delta rule, or back-propagation] works both efficiently and dependably.... In the early years of cybernetics, everybody understood that hill-climbing was always available for working easy problems, but that it almost always became impractical for problems of larger sizes and complexities... The situation seems not to have changed much - we have seen no contemporary connectionist publication that casts much new theoretical light on the situation.... We fear that its [back-propagation's] reputation also stems from unfamiliarity with the manner in which hill-climbing methods deteriorate when confronted with larger-

> scale problems. In any case, little good can come from statements like 'as a practical matter, GD leads to solutions in virtually every case' or 'GD can, in principle, learn arbitrary functions'. Such pronouncements are not merely technically wrong; more significantly, the pretense that problems do not exist can deflect us from valuable insights that could come from examining things more carefully. As the field of connectionism becomes more mature, the quest for a general solution to all learning problems will evolve into an understanding of which types of learning processes are likely to work on which classes of problems.

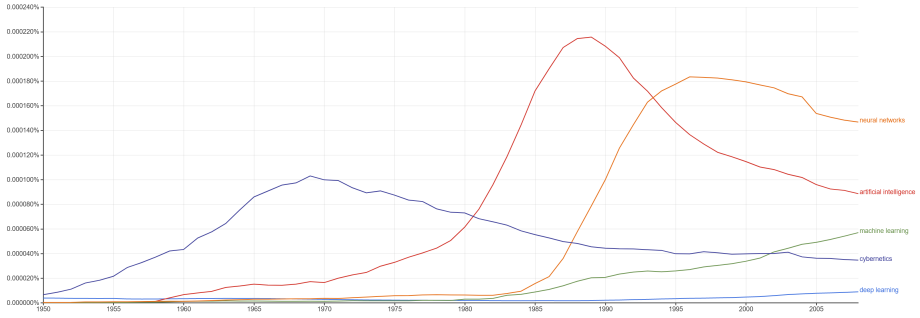
But it was too late this time.

13 Slide 13



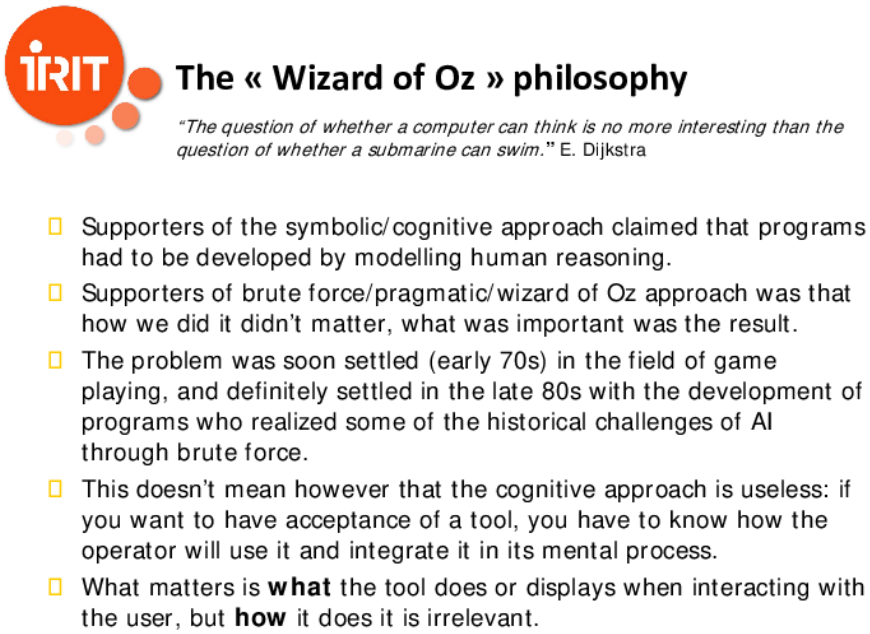
Cognitive AI had shown his shortcomings, the Fifth Generation Computer System project would end in 1992 in another controversy, and Expert Systems were on their way to the grave. I took this shot while attending FGCS'92 where I published my first important paper on a very complicated subject about automatic resolution in non classical logic using Gentzen's sequents, a field dead today.

14 Slide 14



AI had already entered its second winter, that would last 20 years.

15 Slide 15



During these 20 years however, some problems which were considered as AI were solved, such as the historical victory of DeepBlue against Gary Kasparov in 1997. The third branch of AI, that I would call pragmatist AI, or brute force AI, was on his way. Its motto was "what matters is what you do and not how you do it". There is a famous quote by Dijkstra that said that "the question of whether a computer can think is no more interesting as whether a submarine can swim". The philosophy behind it was sometimes also called the "wizard of Oz" philosophy, as a reference to the character

in the movie of the same name, who does not perform any actual magic trick, but makes people believe that he does. What matters is what the tool does or display when interacting with the user, but how it does it is irrelevant.

16 Slide 16

Quite ironically, this approach, which was quite in opposition with cognitive/symbolic AI is sometimes know regrouped with it inside a category called “handcrafted AI”.

17 Slide 17

The delayed effect

The chicken still runs while its head has been cut off...

- The ERATO case is a typical example of the delayed effect.
- Started in 1988 as an expert system for Air Traffic Control at the time when expert systems were already on their way to the grave.
- Excellent operational philosophy: « controller in the loop », acceptance of tools, etc. which proved to be right.
- Science outdated: all tools had to be developed based on a cognitive modelling of the controller up to the end of the 90s. Anything else was anathema.
- Put into operational service in 2015, 27 years after the start of the project.
- Was it possible to go faster from the « all cognitive » philosophy to the more pragmatic « wizard of Oz » philosophy?
- It takes time for science to perfuse...

Now, I am going to speak about a very interesting phenomenon, that I call the delayed effect. While in the scientific community the times of expert systems and cognitive modeling were gone forever, this was not the case in all communities, and especially inside the ATM french community. In 1988, the french CENA started a project called ERATO. I remember very well the exact date because I did my ENAC master dissertation on ERATO in 1988. ERATO was to be an expert system for air traffic control, based on a cognitive modeling of the controler. Behind ERATO was an air traffic controller who remained in charge of the project till the late nineties. ERATO was created at a time when there was quite a controversy in the ATC community. On the one hand, we had the US AERA project, with a third phase, AERA3, which was supposed to be a fully automated system, with the air traffic controller being only a system manager. On the other hand, there was the “controller in the loop” philosophy, which was championed by the French. I presume that behind the differences in philosophy there was also political reasons, that can be found in the way controllers strikes were handled in the US and in France in the seventies and in the early eighties, when Ronald Reagan fired 90% of the American air traffic controllers after the great strike of 1981.

In retrospective, 30 years later, it is quite clear that the operational philosophy behind ERATO was very sound, and that the approach was much more realistic than the AERA 3 approach. Full automation is still out of sight, and controller’s tools are being developed quite universally by the ATM systems developers with the controller in the loop philosophy. Ideas which originated in ERATO, such as the notion of the

necessity of the acceptance of a tool by the controller, are now considered as common knowledge.

But the problem of acceptance has nothing to do with the algorithms used to develop the tools. So, while the operational philosophy was absolutely sound, the science used in the beginning was already outdated. However CENA didn't flinch and it was anathema at the time to only propose to develop tools in ERATO based on something else than the controller's cognitive model. The "wizard of Oz" philosophy had not reached the management. To predict aircraft conflicts (or "problems"), you had to do the job as the controller did it, while the true question was in fact to build a tool that would be compatible with the cognitive model of the controller, but how the computer did the job was completely irrelevant.

In the end, it took 27 years (from 1988 to 2015) to have ERATO in operational service. Could have it been avoided? It's very hard to know, but I think it was not. ERATO is just a case study of the delayed effect, (and probably also a case study in the theory of commitment): it takes some time for the results of science to perfuse the society. Perfusion is probably faster today, because the internet enables a much quicker diffusion of information. But still, the delayed effect can not be avoided.

However, ERATO was the source of many interesting results regarding tools acceptance, and thus the cognitive modelling work was very useful. The error was to take an "all-or-nothing" approach about it, instead of being pragmatic and using computers for what they are, and relying on the wizard of Oz philosophy to develop the tools while using cognitive modelling for designing what the tools should do and how they should interact with the controller.

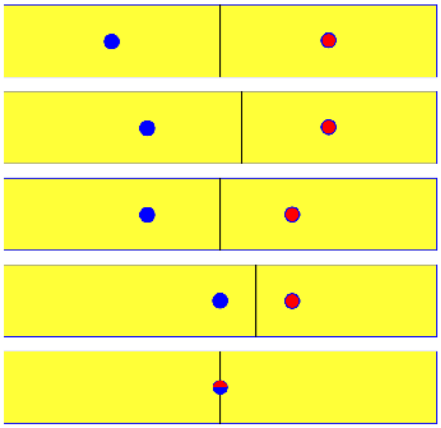
18 Slide 18

Science and ideology: the Free Flight affair

- "In 1995 David Hinson, the FAA administrator, organized a task force to draw up detailed plans to implement free flight. The report, issued in October that year called for three phases; phase I ended at the end of 2002, the others have not been started."
- The idea of Free Flight was interesting in low density areas.
- However, it is inefficient in high density areas, because a global optimum is always at least as good as multi-local optimizations, and is most of the time better.
- Roughly said, if you throw your clothes in your suitcase at random, they will all reach their minimal state of energy (the lowest position they can reach) but it is highly unlikely that your suitcase will be packed in an optimal way…

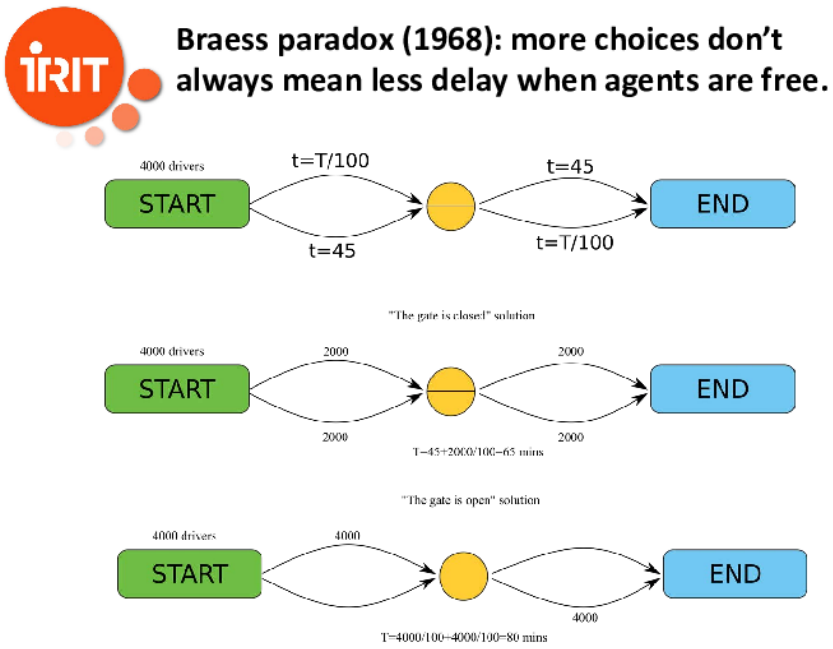
We had another interesting example of this "all-or-nothing" approach in ATM, in the late 90s. The FAA began to promote a concept known as Free Flight. The idea was that aircraft would be responsible to provide self-separation. While Free Flight was an interesting idea for low density areas, it is a mathematical nonsense in high density zone. It is quite easy to understand without a lot of mathematical demonstrations that global optimization is more efficient than multi local optimization. If you throw you clothes into your suitcase at random, they will all reach their minimal state of energy (the lowest possible position), however it is highly unlikely that your suitcase will be optimally packed.

19 Slide 19



In fact, this problem is well known in a very large set of problems, such as Hotteling's law in economy, with the classical example of the ice cream sellers problem. Here two ice cream sellers are on a beach. The initial position on the upper part of the slide is the "optimal" one, in the sense that people on the beach have at most to walk a quarter of the beach to reach the nearest ice cream seller, and each ice cream seller has half of the available clients. However what happens if the blue seller is smart? He will soon see that, by installing his stand closer to the centre of the beach, he will steal some of the clients of the red seller. In turn, the red seller will move closer to the centre, until each seller reaches the centre, which is a Nash equilibrium. Now they still have half of the clients, but some people have to walk half of the beach to reach their seller. If you ask a Chicago School economist about this, it will explain it in terms of "negative externalities". The reality is that sometimes you need a regulator to reach the optimum.

20 Slide 20



The same thing happens in networking, a problem much closer to us, with Braess paradox. Here we have 4000 drivers who have to drive from START to END, passing through the middle yellow point. The time to travel on each segment is indicated on the upper part of the slide. For example, the time to travel from start to middle on the upper road is equal to the number of drivers on this road divided by 100, while the time from START to middle on the lower segment is fixed, and is equal to 45 minutes. In the middle part of the slide, we have the "gate is closed" solution. Drivers, if they choose to take the upper road from START to middle, must remain on the upper road from middle to END (resp. on the lower one). The Nash equilibrium is reached when there are 2000 drivers on each road, with a total time of 65 minutes for each driver. However, on the lower part of the slide, the drivers have the choice to change from upper to lower, or from lower to upper at the middle point. And here, the Nash equilibrium is reached when there are 4000 drivers on the upper segment from START to middle, and 4000 drivers on the lower segment from middle to END, with a total travelling time of 80 minutes (to convince yoursel of this, try to change the path of one driver, and you will see that its travelling time will be longer whatever path he chooses, so he has no incentive to modify his choice)! Here, adding roads to a network or opening new crossroads will result in longer delays instead of shorter ones if the agents are free to choose the road they want, and this paradox has been observed on real road networks.

21 Slide 21

The price of anarchy

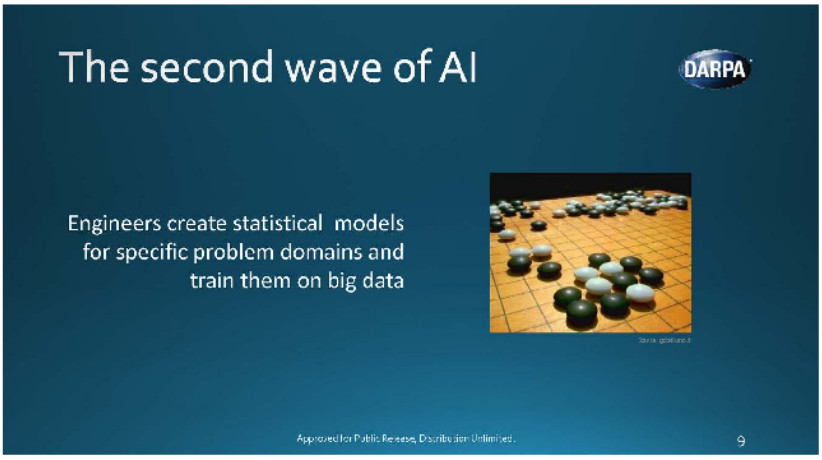
- Multi-agents systems reach Nash equilibriums, and Nash equilibriums are not always optimal
- Selfish Routing and the Price of Anarchy, Tim Roughgarden (MIT Press, 2005)
- However, this doesn't mean that we should not have investigated free flight, but just that we should not have claimed that it was going to solve every problem.
- Every method is a tool, and part of the problem is to find the right tool for the right problem (from a user perspective), or sometimes the right problem for the right tool (from a scientist perspective).

The short story is: when agents are free, they reach a Nash optimum, and Nash optima are not always global optima. So Free Flight could have been a solution for many problems, but absolutely not for solving capacity problems. In fact, it looks like Free Flight died in ten years, and the second phase of the program was never implemented. I recently read a few documents, such as President Trump plan for the modernization of the American ATC, and while the ideas are sometimes quite "extreme", there is not a single mention of Free Flight.

All of these problems are well developed in the excellent book "Selfish Routing and the Price of Anarchy" by Tim Roughgarden (MIT Press, 2005). However, this doesn't mean that we should not have investigated free flight, as it could be a solution in low density areas, but just that we should not have claimed that it was going to solve every problem.

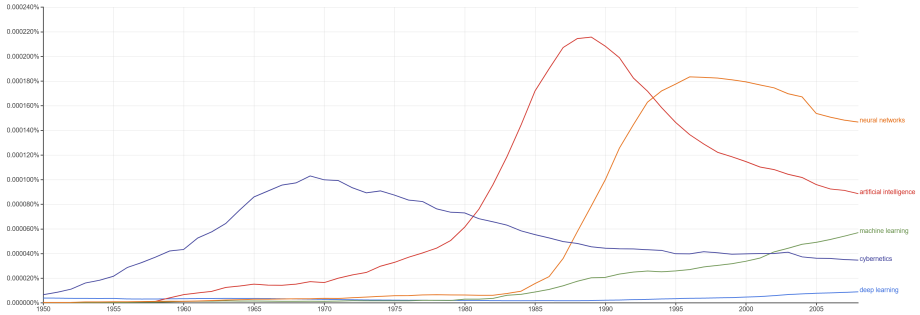
Every method is a tool, and part of the problem is to find the right tool for a given problem (from a user perspective), or sometimes the right problem for a given tool (from a scientist perspective, who has often the bias, if he has a hammer, to look at each problem as a nail).

22 Slide 22



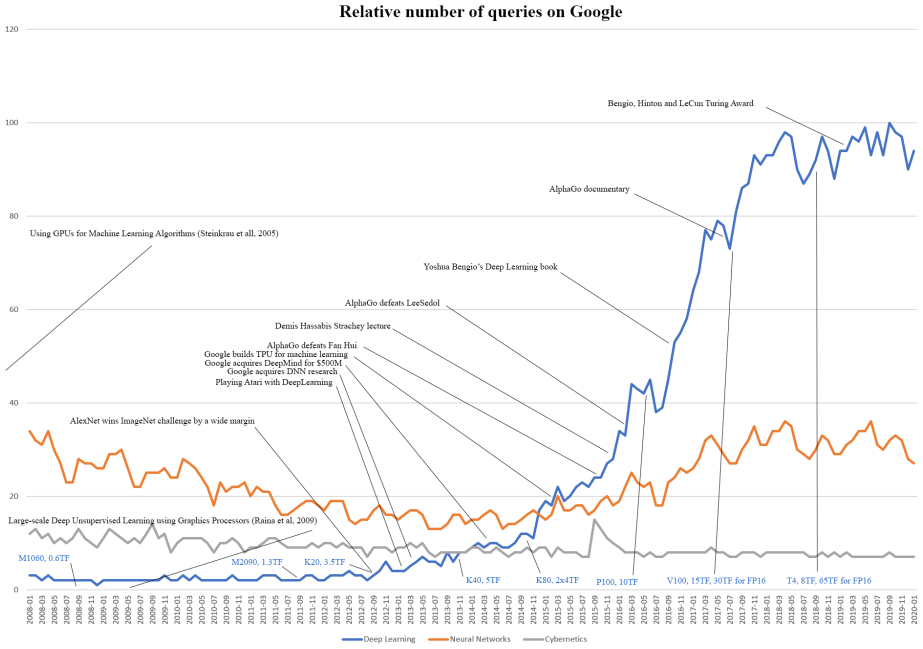
Now after this historical perspective, let's go back to our current subject: what DARPA calls the second wave of AI, even if I, like many others, think that it is more a third wave than a second wave. Today everything is about deep learning and machine learning. But things have not been as simple as they seem.

23 Slide 23



First neural networks stagnated for a few years (1985-2012), mainly because of the lack of proper hardware to support them. The idea of Deep Learning appeared in the 80s, but training neural networks required a kind of parallel computing power that was not available at the time.

24 Slide 24



Things began to change with the apparition of fast graphic cards, which were intrinsically parallel, around 2005, with the first implementation in 2005 by Steinkrau. There was a second seminal paper in 2009 by Andrew Ng and others, and the real breakthrough was the victory by a very wide margin of AlexNet, developed by Alex Krizhevsky under the supervision of Geoffrey Hinton, in the ImageNet challenge in 2012. This is considered as the starting point for the new AI revolution. Then things accelerated at an incredible pace. In 2013, David Silver, from DeepMind, presents its paper on how neural nets can solve Atari games. Google buys DNN, Geoffrey's Hinton startup for its image recognition software. The following year they buy DeepMind for 500M$, and soon starts manufacturing Tensor Processing Units for Deep Learning. The breakthrough in the media was the victory of AlphaGo against Lee Sedol in 2016. Winning against a World Champion of Go was, 10 years before, considered as almost impossible within decades. It was a tremendous achievement, and a clear demonstration of DeepMind ability.

25 Slide 25

After that, everybody wanted to be in the Artificial Intelligence and Deep Learning business, while 5 years before, nobody wanted to hear anything about AI.

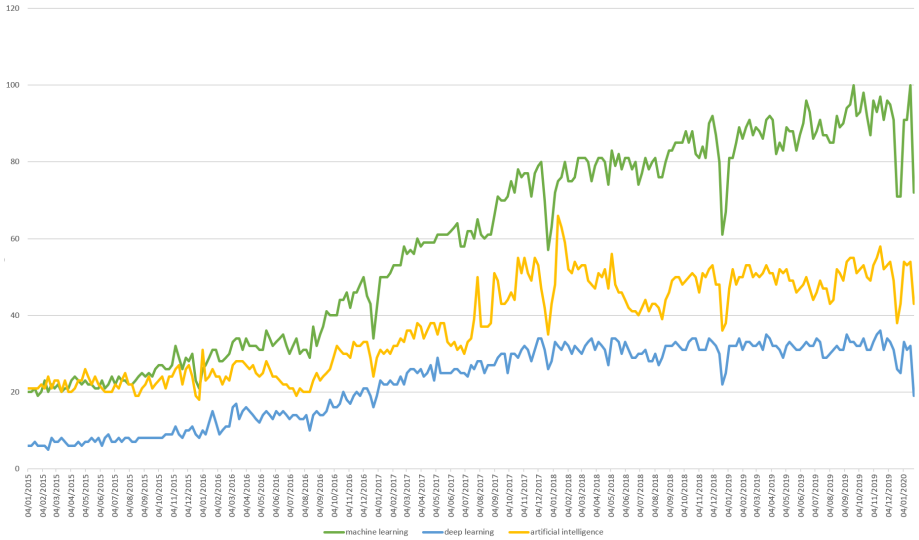
26 Slide 26

Reinforcement learning and delayed rewards

- Sutton (1988): « Learning to predict by the methods of temporal differences"
- Watkins (1989): « Learning from Delayed Rewards »
  - « Q-learning »
  - SARSA
  - Deep Q learning
  - Double Q learning
  - etc….
- The multi-armed bandit problem
  - UCB algorithm (2002)
- Monte-Carlo Tree Search (Coulom 2006, Abramson 1987)

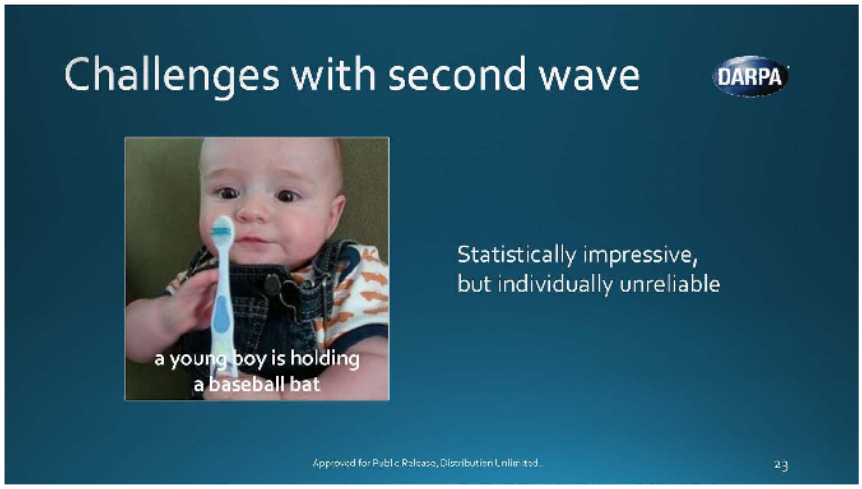
As I already said, fast parallel hardware was behind the boom of machine learning. But other advances in algorithms and maths were also behind it. I don't want to make a complete lesson on machine learning and reinforcement learning, which is a field as old as the work of Arthur Samuel (1959): "Some Studies in Machine Learning Using the Game of Checkers" and of course of Donald Michie (1962) with the famous MENACE (Matchbox Educable Noughts And Crosses Engine) machine. And we had significant advances, especially regarding the problem of delayed rewards in machine learning, as soon as 1988 with the methods of temporal differences by Sutton, then Watkins (1989): "Learning from Delayed Rewards", later we had the development of other algorithms such as "Q-learning"and the likes, and of course the development of the Upper Confidence Bound method to solve the multi-armed bandit problem by Auer in 2002. The last brick was added by Remi Coulom with Monte-Carlo Tree Search, which was central to the AlphaGo program. In fact, Aja Huang, the lead programmer of AlphaGo, had his PhD under the supervision of Remi Coulom before being hired by DeepMind.

27 Slide 27



Today machine learning is even more popular than Artificial Intelligence. In a way, the part has become larger than the whole.

28 Slide 28



Well we have a problem now: is that AI that good, or is it, again, only a new hype? The achievements of Deep Learning are impressive, but can they be used for everything?

DARPA wrote that “Artificial Intelligence is statistically impressive but individually unreliable”.

29 Slide 29

Grand failures...

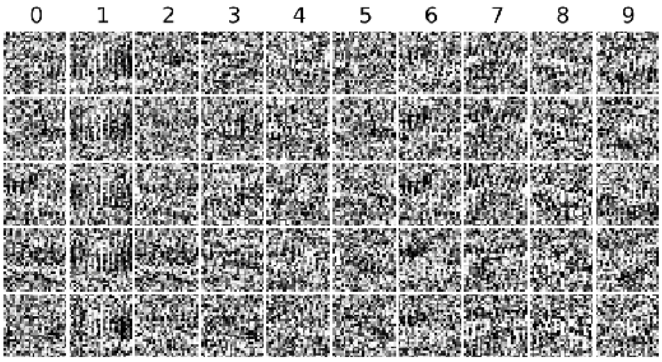


Figure 4. Directly encoded, thus irregular, images that MNIST DNNs believe with 99.99% confidence are digits 0-9. Each column is a digit class, and each row is the result after 200 generations of a randomly selected, independent run of evolution.

Well, we have already a very large panel of examples where NN fail miserably, for example in image classification. Here a NN thinks that he recognizes numbers where there is apparently only noise.

30 Slide 30

Grand failures, part 2

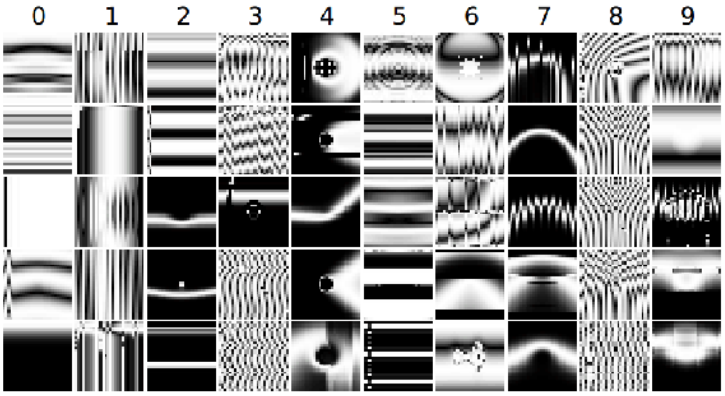
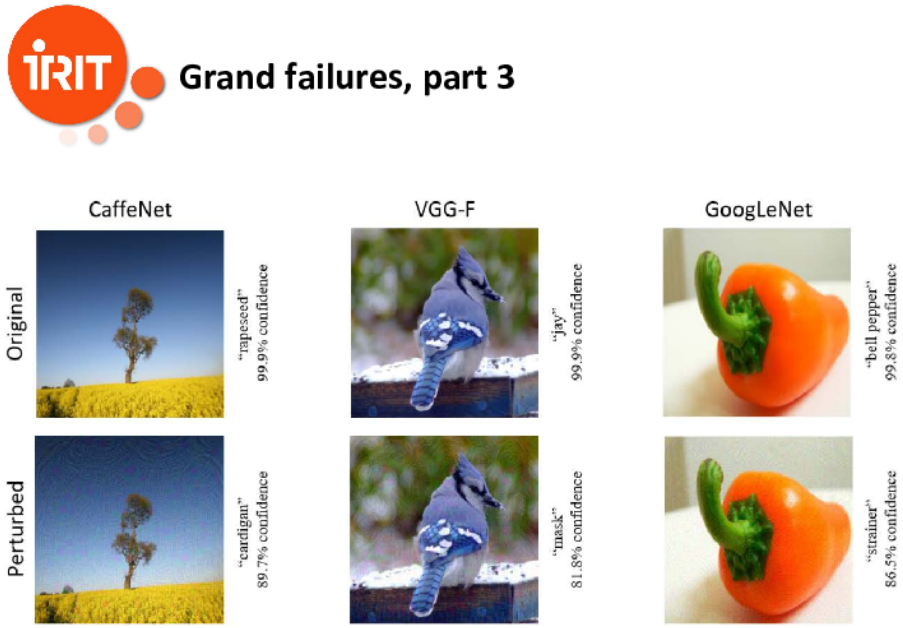


Figure 5. Indirectly encoded, thus regular, images that MNIST DNNs believe with 99.99% confidence are digits 0-9. The column and row descriptions are the same as for Fig. 4.

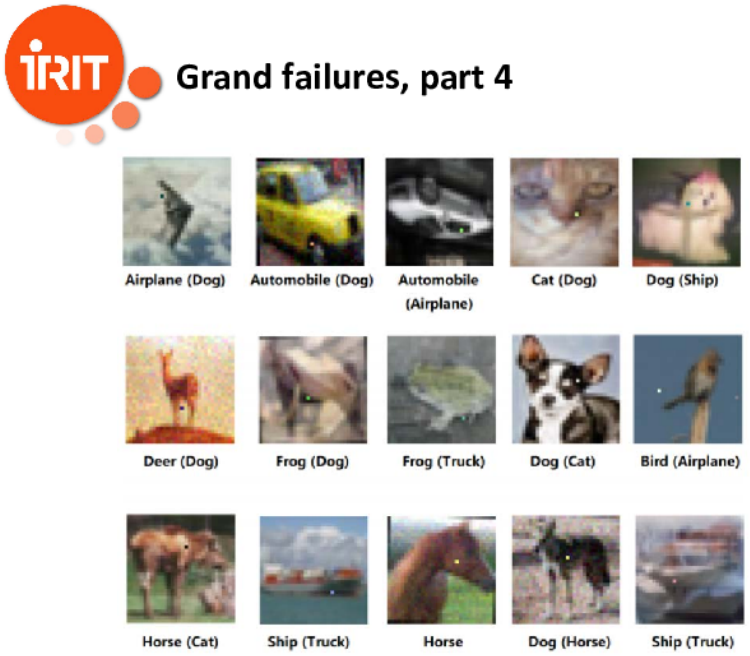
Here, the NN still recognizes numbers where there only shapes completely unrelated to numbers.

31 Slide 31



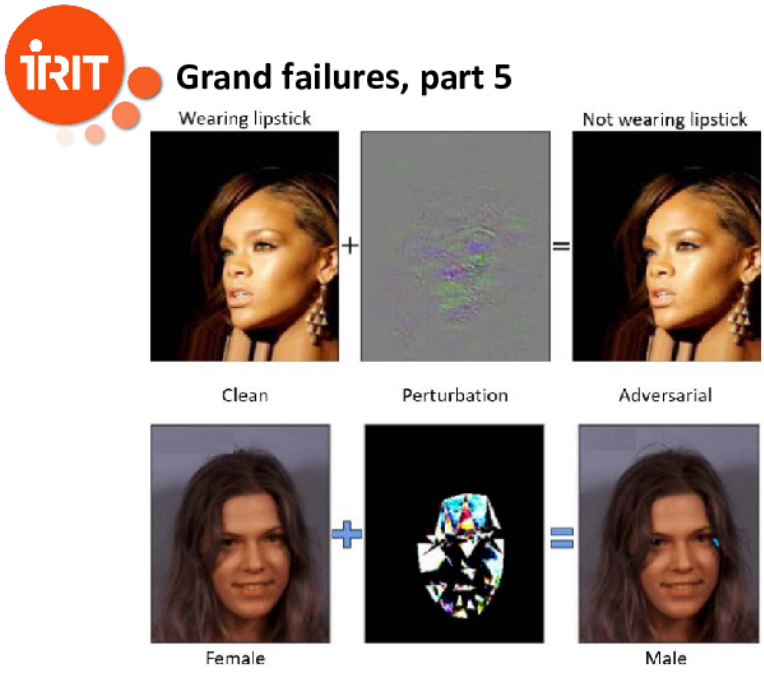
Here, very small perturbations of images change completely the interpretation, while still getting a very high level of confidence, and for all of the best networks.

32 Slide 32



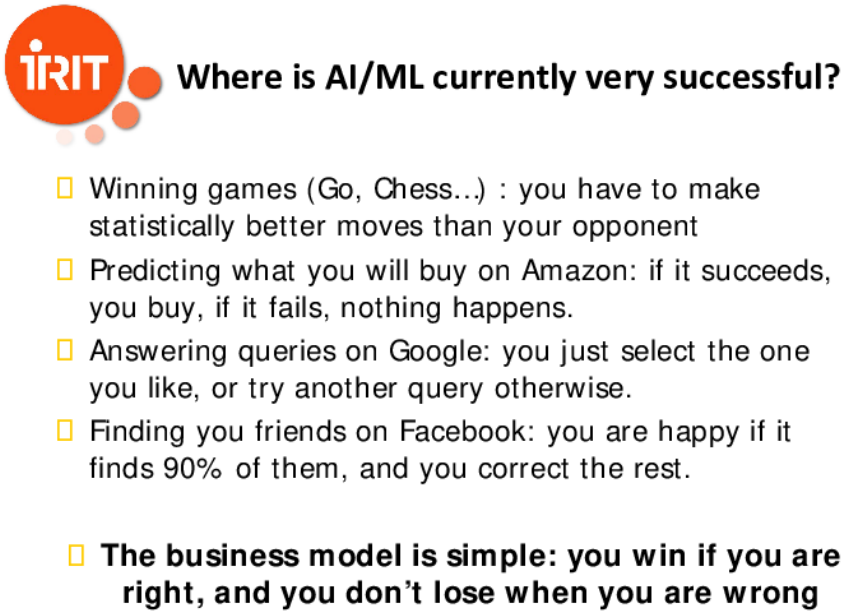
Here, changing only a pixel modifies completely the classification.

33 Slide 33



Here, insignificant perturbations can change the classification from male to female.

34 Slide 34



So where are they really efficient? There is a clear business model, which is working well: neural networks are efficient when failure has no consequence and success is rewarding. There are classical examples of this: games, because statistically making better moves than your opponent is enough. Purchase suggestion, because if it is irrelevant, you just don't buy, but you wouldn't have bought anyway, while if the suggestion is relevant you buy, and thus the company makes more money. You have also search engine queries, or finding friends on Facebook, or automatic translation on Google. For all these examples, failure is not a problem. And even more, people don't expect the system to be perfect.

35 Slide 35

But can we use it for:

- Finding cancerous cells in images?
  - False hits could be corrected
  - False misses are disasters
- Autonomous drones
- Finding defaults in a production chain?
  - Too many errors can destroy the operational margin
- Solving aircraft conflicts?

But what happens if you want to apply neural networks for cancerous cells detection? False hits can be corrected, but false misses might cost the patient his life. And what for autonomous drones, for finding defaults in a production chain, or even for solving aircraft conflicts? Here, failing is a serious problem.

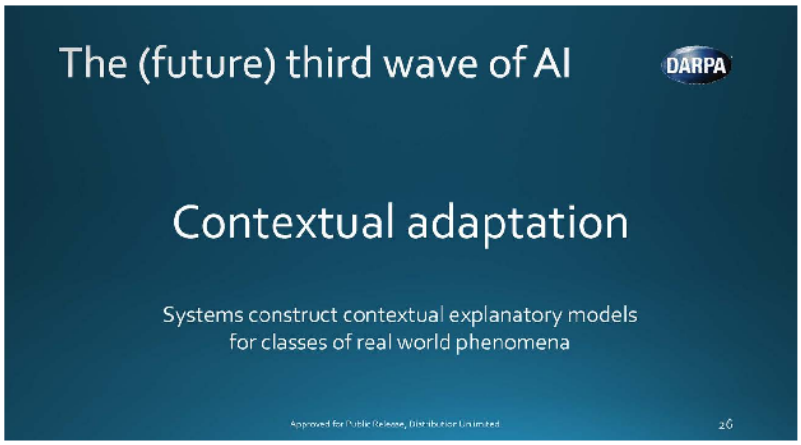
36 Slide 36

Ethics consideration

- Google photo service classified black people as gorillas in certain pictures (2015) The only solution was to remove the « Gorilla » category (2018).
- Face recognition systems perform worse with women, and far worse with dark skinned women (2019).
- Amazon « hiring tool » systematically downgraded candidates that have attended all women college. The tool was abandoned (2018).
- The infamous COMPAS algorithm that certain courts in the United States used to assess the likelihood that a criminal defendant becomes a recidivist was shown to have a very strong bias against black people (2016).
- The problem is simple: in a way, the dataset IS the algorithm, which just reproduces all biases contained in the set.

There are also biases and ethics problems which became apparent in the last years. Google algorithm for image classification had racist biases, such as sometimes recognizing black persons as gorillas. And solving the problem apparently required to remove the gorilla category. Image recognition systems are the most efficient with white males, and the less efficient with dark skin females. Amazon hiring system was discriminating women. And the COMPAS algorithm used in some justice courts in the United States for predicting the capacity of recidivism was show to exhibit a strong and unfair bias against black people. The central problem is that the dataset is the algorithm. If there are biases in the data, there will be biases in the algorithm, and detecting biases is very difficult.

37 Slide 37



So, what can we do? According to DARPA we are on our way to the third wave of Artificial Intelligence.

38 Slide 38

- Explainable AI
- Reliable AI
- Hybrid AI
- But currently it is mostly wishful thinking...
- And we might very well be in the 90%/10% - 10%/90% problem
- Perhaps we should not bet everything on Deep Learning after all...

DARPA calls it “contextual adaptation” but it has in fact a lot of different names in the community: hybrid AI, reliable AI, explainable AI... There are lots of projects around those concepts, but let’s being honest, it is currently only wishful thinking. Maybe it will yield results, or maybe we are just hitting a barrier, and we are in the classical 90% / 10% problem: it takes 10% of the time to solve 90% and 90% of the time to solve the remaining 10%. And sometimes, it takes forever.

39 Slide 39

- A new tool is just a tool, and must be applied to the right problem. There has never been any silver bullet, and it is doubtful there will be one.
- Old techniques might be old, but they also might be more efficient than new ones. To solve scheduling problems, use SAT solvers. To do data analysis, use statistics. To solve linear problems use linear solvers. Etc.
- It is always useful to keep his eyes and his ears open to every scientific research, even if they are not any more fashionable, or if they are not yet fashionable. We just don't know.
- It is mandatory to investigate the drawbacks and limitation of any new tool, as the sooner we find and acknowledge them, the sooner we will be able to find the right field of application. Being a blind believer does not improve things.
- It is extremely important to compare the results of many different algorithms on sets of identical data. If most of the evolution in AI came from game programming or image recognition, it is because data are standardized and results are indisputable.

I have tried to convince you that blindly jumping on every new train is not necessarily a good idea. And these are a few things I strongly believe in:

- A new tool is just a tool, and must be applied to the right problem. There has never been any silver bullet, and it is doubtful there will be one.
- Old techniques might be old, but they also might be more efficient than new ones. To solve scheduling problems, use SAT solvers. To do data analysis, use statistics. To solve linear problems use linear solvers. Etc. It is always useful to keep his eyes and his ears open to every scientific research, even if they are not any more fashionable, or if they are not yet fashionable. We just don't know.
- It is mandatory to investigate the drawbacks and limitation of any new tool, as the sooner we find and acknowledge them, the sooner we will be able to find the right field of application. Being a blind believer does not improve things. Being critical might.
- It is extremely important to compare the results of many different algorithms on sets of identical data. If most of the evolution in AI came from game programming or image recognition, it is because data are standardized and results are indisputable.

40 Slide 40

Some wisdom from a master

- *Rodney Brooks (1991): Critical paper about the then-mainstream line in AI (cognitive AI and expert systems):*
  - "There is a bandwagon effect in Artificial Intelligence Research, and many lines of research have become goals of pursuit in their own right, with little recall of the reasons for pursuing those lines"
- *Rodney Brooks: Machine Learning Explained (28 August 2017):*
  - *https://rodneybrooks.com/forai-machine-learning-explained/*
  - "Vast numbers of new recruits to AI/ML have jumped aboard after recent successes of Machine Learning, and are running with particular versions of it as fast as they can. They have neither any understanding of how their tiny little narrow technical field fits into a bigger picture of intelligent systems, nor do they care. They think that the current little hype niche is all that matters, are blind to its limitations, and are uninterested in deeper questions."
  - "The papers in conferences fall into two categories. One is mathematical results showing that yet another slight variation of a technique is optimal under some carefully constrained definition of optimality. A second type of paper takes a well-known learning algorithm, and some new problem area, designs the mapping from the problem to a data representation, and show the results of how well that problem area can be learned."

I will end this talk with these quotes from Rodney Brooks, an old master of the AI field, who was always both a critic and a creative researcher.

Rodney Brooks, critical paper about the then-mainstream line in AI (cognitive AI and expert systems) (1991):

> "There is a bandwagon effect in Artificial Intelligence Research, and many lines of research have become goals of pursuit in their own right, with little recall of the reasons for pursuing those lines"

Rodney Brooks: Machine Learning Explained (28 August 2017):

> "Vast numbers of new recruits to AI/ML have jumped aboard after recent successes of Machine Learning, and are running with particular versions of it as fast as they can. They have neither any understanding of how their tiny little narrow technical field fits into a bigger picture of intelligent systems, nor do they care. They think that the current little hype niche is all that matters, are blind to its limitations, and are uninterested in deeper questions.
>
> The papers in conferences fall into two categories. One is mathematical results showing that yet another slight variation of a technique is optimal under some carefully constrained definition of optimality. A second type of paper takes a well-known learning algorithm, and some new problem area, designs the mapping from the problem to a data representation, and show the results of how well that problem area can be learned."

And while you read them, I am waiting for any question.