Can artificial intelligence give a “mind” to machines?

Hugo Leonardo Rufiner
SINC-UNL-CONICET
lrufiner@sinc.unl.edu.ar
Organization

- Introduction
- Where are we?
- Possible scenarios
- Deep learning
- General Artificial Intelligence
Introduction

The beginning...
Many concepts and definitions in different disciplines...

PHILOSOPHY, PHYSICS, BIOLOGY, MEDICINE, PSYCHOLOGY, PSYCHIATRY, THEOLOGY, ETC.

Francisco José Soler Gil (Sevilla University)
What is the mind?

- It is the phenomenon responsible for understanding, reasoning, perception, emotion, memory, imagination, will, consciousness and other cognitive skills...
- Many of these skills are essential characteristics of the human being.
Throughout history it has been conceived in different forms:
- a substance distinct from the body,
- a part,
- a process, or
- a property.

Current dominant materialist ideas include this concept in the mind-brain identity theory and functionalism.
MIND = SOFTWARE & BRAIN = HARDWARE?

• It is in this latter sense that the duality mind-brain has been associated more strongly with "equivalent" concepts coming from the field of computer sciences.

• These analogies can be useful in some contexts, but it is "dangerous" to use them beyond their limits of application.
What is consciousness?

• Cannot be operationally defined easily, since it is a complex and subjective phenomenon (cannot be observed directly).

• Associated with: first-person experience, interiority, subjectivity, the self-perception
Science and Consciousness

>20,000 Scientific articles about conscience and still there is no consensus...

Editor “Complexity” 2017: “The main problem with the study is that consciousness as such is not a scientific concept. Consciousness is not defined as the authors admit. Consciousness is a catch-all phrase that does not explain anything. Most importantly, researchers who take a scientific point of view do not need consciousness to explain anything. Consciousness is a religious motivated concept that has been introduced by humans to make them feel better or superior to the rest of the world. The concept of consciousness has the same status as the belief that the earth is the center of the universe. Both are misconceptions of reality.”
Artificial Intelligence (AI)

• "The science and engineering of making intelligent machines." (John McCarthy, 1956).

• "The study and design of intelligent agents, where an intelligent agent is a system that perceives its environment and takes actions which maximize its chances of success." (R. B. Bernstein and W. N. Curtis, 2008)
Where we are today?
A brief historical review...
ML/ANN brief history review...

Electronic Brain

1943

1950

Perceptron

1960

Golden Age

1969

Dark Age ("AI Winter")

1970

1980

1990

ADALINE

XOR Problem

Multi-layered Perceptron (Backpropagation)

1940

S. McCulloch – W. Pitts

F. Rosenblatt

B. Widrow – M. Hoff

M. Minsky – S. Papert

D. Rumelhart – G. Hinton – R. Williams

- Adjustable Weights
- Weights are not Learned

- Learnable Weights and Threshold

- XOR Problem

- Solution to nonlinearly separable problems
- Big computation, local optima and overfitting

26-28 November
Santa Fe, Argentina
Possible scenarios

Trying to get closer to an answer to some questions...
Open questions…

• Could we equip the machines with a consciousness that would make them subjects, that is, that endow them with subjectivity? How?

• And if we can't do this, we could at least build machines that behave in all senses and contexts as a human being, but without consciousness, and therefore subjectivity?

• Or the difference between having and not having a subjective perspective it is necessarily manifested in objective differences (certain skills, certain behaviors) between machines and humans?

• And if we have to give certain insurmountable differences between beings endowed with subjectivity and beings who do not have it, in what skills or behaviors will these differences be perceived?
Three main scenarios are opened:

1) «Unexpected emergence of consciousness»: a machine equipped with intelligence at the human level in the analogical sense is having consciousness and subjectivity.

2) «Creation of a zombie»: it is possible to build a machine that is indistinguishable from a human, but without consciousness, and therefore without subjective perspective.

3) «Impossible skills»: It is possible that being endowed or not with subjectivity necessarily implies differences between what men and machines can do.
Scenario #1: «Unexpected emergence of consciousness»:

• "It is possible to build a machine equipped with intelligence at the human level, in the analogical sense, that results in consciousness and subjectivity?".
• It was one of the first hypothesis from supporters of strong AI.

• All kinds of proposals, for example (without shades!):
  – Physicalism: every material entity has some degree of consciousness (M. Kaku).
  – Biologicism: awareness involves an interaction between the brain and the body (S. Greenfield).
  – Functionalism: If you increase the algorithmic complexity, then at some point consciousness will emerge (Block, M. Minsky).
  – Comportationism: If it behaves in a particular way then it has consciousness (Turing?).
– From NC is a topic of great interest and even open.
– From CS/AI main interest goes to improve algorithms and results.
– One of the projects that goes in this direction of searching "emerging" properties is the Blue Brain.
Blue Brain project (2005)

- Reconstructing the brain piece by piece and “building” a virtual brain in a supercomputer.
- The **computing power** needed is considerable. Each simulated neuron requires the equivalent of a laptop computer.
- Different stages.
• **Simulation problem:**
  “You can simulate a nearly perfect storm in a computer, but there is no way to get wet” (well discussed by Tononi, Searle).

• It would be the same with ANNs, unless you build them physically and only for some configurations (according to IIT).
Scenario #2: «creation of a zombie».

• “It is possible to build a machine that is indistinguishable from a human, but without consciousness, and therefore without subjective perspective? .”
• This would imply that it can pass the Turing test without consciousness, making only a very good simulation...

• Given the limitations of the TT other tests have been proposed such as the College student...

• There are many supporters of this view (Tononi, Searle).
Interview with Eugene Goostman, the Fake Kid Who Passed the Turing Test

Doug Aamoth  @daamoth  June 9, 2014

Chatbot Eugene Goostman supposedly passed the legendary Turing Test on Sunday, tricking 33% of a panel of judges into believing he was a real boy during the course of a five-minute chat conversation.

The milestone conveniently occurred 60 years to the day after Alan Turing passed away; Turing bet that by the year 2000, computers would be intelligent enough to trick humans into thinking they were real 30% of the time.
• Although they are indistinguishable from humans can be **dangerous** as well...MIT Media Lab: **Ethics of AI Course** (Joichi Ito, 2018)

https://m.xataka.com/entrevistas/doy-clases-de-eticade-inteligencia-artificial-y-esto-eso-que-enseno-a-futuros-ingenieros
Scenario #3: «impossible skills».

- “Would be possible that the presence or absence of subjectivity necessarily involves differences between what humans and machines can do?”. 
Let's hypothesize some skills that we could consider "related" with consciousness...

For example:
- Paint, draw, criticize, talk, play the piano, imagine, write a book, socialize, have empathy, read the mind, believe, love, laugh, ...
The images combine the content of photographs with styles of famous works of art.

A) Original photograph that represents the Neckarfront in Tübingen, Germany.

B) Output painting with style of "The Shipwreck of the Minotaur" (J.M.W. Turner, 1805).

Scenario #3: Paint/draw

- Google 'Inceptionism' art sells big at San Francisco Auction

- You've never seen 'Starry Night' like this before.

- Sarah Cascone, 2016

https://news.artnet.com/market/google-inceptionism-art-sells-big-439352
Scores of creativity for 1710 paintings of "Artchive dataset". Each point represents a painting.

- Horizontal axis: year of creation of the painting.
- Vertical axis: score of creativity (scaled).
- Miniatures: paintings which scores relatively high or low compared to its neighbours.

A. Elgammal and B. Saleh “Quantifying Creativity in Art Networks”, 6th Int. Conf. on Comp. Creativity (ICCC), 2015, USA.
• Sample by sample synthesis with a particular type of CNN.
• TTS and audio results of MOS quality evaluation for different techniques.
• Music generated after training with a dataset of classical piano music.

Comparison StackGAN and 1 stage GAN to generate 256 × 256 images:

(a) Given the text descriptions, the Stage-I of StackGAN outlines rough forms and the basic colors of objects, producing low resolution images.

(b) Stage II of StackGAN takes the results of phase I and the descriptions of text as input, and generates high resolution images with photorealistic detail.

GAN: Generative adversarial networks

Part 1

"The Malfoys!" said Hermione. Harry was watching him. He looked like Madame Maxime. When she strode up the wrong staircase to visit himself.

"I’m afraid I’ve definitely been suspended from power, no chance—indeed?" said Snape. He put his head back behind them and read groups as they crossed a corner and fluttered down onto their ink lamp, and picked up his spoon. The doorbell rang. It was a lot cleaner down in London.

Hermione yelled. The party must be thrown by Krum, of course.

---

Harry Potter: wrote by AI
Max Deutsch, 2016
He trained a LSTM using text from the first four Harry Potter books.
Then, he asked it to produce a chapter based on what it learned.
Here there is a part of a chapter 1 (added a bit of format to help readability).

“Racist” Chatbot:

Tay was an AI chat robot released by Microsoft via Twitter in March 23, 2016. He started to publish offensive tweets that forced Microsoft to close the service only 16 hours after its release.

According to Microsoft, this was caused by trolls who "attacked" the service when the bot made responses based on their interactions with people on Twitter (empathy?).

Coherent averaging estimation autoencoders applied to evoked potentials processing

Iván E. Gareis, Leandro D. Vignolo, Ruben D. Spies, Hugo L. Ruffner

ABSTRACT

The success of machine learning algorithms strongly depends on the feature extraction and data representation stages. Classification and estimation of small repetitive signals masked by relatively large noise usually requires recording and processing several different realizations of the signal of interest. This is one of the main signal processing problems to solve when estimating or classifying P300 evoked potentials in brain-computer interfaces. To cope with this issue we propose a novel autoencoder variation, called Coherent Averaging Estimation Autoencoder with a new multiobjective cost function. We illustrate its use and analyze its performance in the problem of event related potentials processing. Experimental results showing the advantages of the proposed approach are finally presented.

Scenario #3:

- About the future possibilities opinions are more divided...
- On the other hand, more and more human intelligence skills can gradually be emulated and are also used in more complex tasks...
- The truth today is that it has not been demonstrated that a single machine is aware and in fact I am sure that these algorithms are not.
Scenario #3:

- How do I know? I make a confession, I feel like a professional magician ...

[Link](https://www.quora.com/How-does-Criss-Angel-do-his-massive-levitation-illusions-live-or-can-it-not-be-done-live)
Deep Learning
Showing some "tricks" ...
Bioinspiration: microscopy of the cerebral cortex

- You can easily see many neurons organized in multiple layers...
There was no "substantial" change, the conjunction of several "simultaneous" factors
Deep Architectures (DA)

- There are theoretical results that suggest that to represent high levels of abstractions needed DA (instead of shallow ones).
- These consist of multiple levels of non-linear operations (such as artificial neural networks with many layers).
- For the DA searching in the space of parameters is a very difficult optimization problem (local minima, noise, chaos, instability, etc.).
- New algorithms have been proposed with great success, defeating the State of the art techniques in many areas (computer vision, ASR, etc.).
1. **Shallow** = feature extraction (manual) + classification (few layers).

2. **Deep** = full "end-to-end“ system (automatic) + many layers.

3. **Training** = greedy layer to layer, unsupervised + last layer supervised + fine adjustment.
Auto-encoders

Input layer | Hidden layer | Output layer
---|---|---
\( \chi_1 \) | \( a_1 \) | \( \eta_1 \)
\( \chi_2 \) | \( a_2 \) | \( \eta_2 \)
\( \chi_3 \) | \( a_3 \) | \( \eta_3 \)
\( \chi_4 \) | \( a_4 \) | \( \eta_4 \)
\( \chi_5 \) | \( a_5 \) | \( \eta_5 \)
\( \chi_6 \) | \( a_6 \) | \( \eta_6 \)

"bottleneck"

Representation Space

Original

Reconstructed
Pre-training: boost performance
Deep hierarchical features

Deep neural networks learn hierarchical feature representations.
General Artificial Intelligence
Re-using old tricks...
Why GENERAL AI?

• “Narrow” AI:
  – Currently, AI algorithms have been designed, trained and optimized by human engineers to achieve a single specific task.
  – They surpass the human skills, but they can not extend capabilities to new domains.
  – This limits the re-use and increases the amount of data to train them, and leaves them without generalization capability or development of "common sense".
• **General AI:**

  – Will be able to overcome these limitations, learn and propose creative solutions for a wide range of tasks from multiple domains.
MODELO DE APRENDIZAJE POR REFUERZO

Ingreso de datos sin procesar → Ambiente → Mejor Acción → Selección de Algoritmo → Agente → Estado → Recompensa → Mejor Acción → Salida

26-28 November
Santa Fe, Argentina
Without domain specific knowledge!
• AlphaGo (Google DeepMind) vs Lee Sedol:
  – Match of 5 Go games among the 18 times world champion Lee Sedol and AlphaGo, played in Seoul, from 9 to 15 March 2016.
  – AlphaGo won all but the fourth game; all games were won by resignation.

Comparable with historic chess match between Deep Blue and Garry Kasparov in 1997.
Final thought...

“But their idols are silver and gold, made by human hands. They have mouths, but cannot speak, eyes, but cannot see. They have ears, but cannot hear, noses, but cannot smell. They have hands, but cannot feel, feet, but cannot walk, nor can they utter a sound with their throats. Those who make them will be like them, and so will all who trust in them.”

(Psalm 115, 4-8)


Bibliography

• Kurt Hornik (1991) "Approximation Capabilities of Multilayer Feedforward Networks", Neural Networks, 4(2), 251–257.
• Tononi; Boly; Massimini; Koch, "Integrated information theory: from consciousness to its physical substrate". Nature Reviews Neuroscience. 17 (7): 450–461 (2017).
• https://www.closertotruth.com/series/consciousness-entirely-physical
Thanks!
lirofiner@sinc.unl.edu.ar

FICH-UNL ~ FI-UNER
SINC – Lab. de Cibernética
CONICET

26-28 November
Santa Fe, Argentina
Thank you