José Hernández-Orallo*
Universitat Politècnica de València, Spain
Leverhulme Centre for the Future of Intelligence, UK
jorallo@dsic.upv.es

* Some slides from Cèsar Ferri, “Good, Bad, Ugly” inspiration from Peter Flach (with many thanks!!)

HUMAINT Winter school on AI and its ethical, social, legal and economic impact
JRC - Seville - February 4-8th, 2019


- **AI**: What does it mean?
- **AI**: A little bit of history
- **AI**: Main areas
- **AI**: The actors
- **AI**: What can (and can’t) it do for you?
- **AI**: The future
- **AI**: Humans they are a-changin’ too!
- Conclusions
AI: WHAT DOES IT MEAN?

Why is it so transformative?

- AI is all about **intelligence**.
  - Intelligence has made humans **dominate** the planet.
  - Intelligence is the mother of all technologies.
- AI is not yet another “new technology”.
  - AI will be the mother of all future technologies.

**So the hype!!!**
AI: WHAT DOES IT MEAN?

- An ambitious view of AI.
  
  "[Artificial Intelligence (AI) is] the science and engineering of making intelligent machines." —John McCarthy (2007)

- A more pragmatic view of AI
  
  "[AI is] the science of making machines do things that would require intelligence if done by [humans]." —Marvin Minsky (1968).

Machines do not need to be intelligent!

- Strong AI
- Weak AI
AI: A LITTLE BIT OF HISTORY

- Ramon Llull, 14th century: Ars Magna.
- Leibniz, 17th century: Ars Combinatoria: calculus ratiocinator
- Hobbes (Leviathan: “reason is nothing but reckoning”
- Babbage, Lovelace, 19th century
- Russell: “Principia Mathematica”
  - Hilbert: Can we formalise it?
  - Gödel: No, we can’t prove everything that is true.
AI: A LITTLE BIT OF HISTORY

- Alan Turing:
  - 1936: Concept of computability and UTM, Church-Turing thesis
  - 1950: The Turing Test
  - 1952: Morphogenesis

- Konrad Zuse; World’s first programmable computer: Z3
  - 1941: The input is a program! The computer changes its behaviour!
AI: A LITTLE BIT OF HISTORY

▪ Dartmouth’s summer research project in “Artificial Intelligence”:

▪ Logic and **deductive reasoning** were considered key for intelligence.

▪ Probability and **inductive reasoning** were advocated by Ray Solomonoff.
AI: A LITTLE BIT OF HISTORY

- **Symbolic AI** dominates the early decades of AI:
  - Intelligence is nothing but symbol manipulation
  - Many kinds of logics are introduced.
  - Several simple problems are solved:
  - Board games, simple theorems, …

**Symbol-grounding problem**: who gives “meaning” to the symbols?
First connexionist approach (ANN):

- The perceptron (Rosenblatt 1958): inspired by how the brain works
  - It can’t learn XOR and other simple functions!
  - They ignored that more layers could do the trick!

It deviated almost all efforts to the symbolic AI school
AI: A LITTLE BIT OF HISTORY

- First AI **Winter** (late 1970s)
  - Symbolic AI didn’t meet the expectations.
  - Connectionist AI was neglected.
  - “Learning” didn’t even appear in many AI books!
- The Expert Systems **Spring** (early 1980s)

Write the rules and let the logic make the reasoning!
AI: A LITTLE BIT OF HISTORY

- **Second AI Winter** (late 1980s)
  - Japan’s fifth generation project fails!
  - Expert systems hard to create and maintain: knowledge acquisition bottleneck!

- **First connexionist Spring** (1990s)
  - Werbos’s backpropagation (1974) takes momentum in the late 1980s
    (Rumelhart, Hinton, Williams, McClelland)

Many ANN different architectures appear
**AI: A LITTLE BIT OF HISTORY**

- **Third AI Winter (late 1990s)**
  - Hinton, LeCun and Bengio stand the freezing storm up there in Canada!

- **The Autonomous Agents Spring (2000s)**
  - This is not yet intelligent, but it’s an autonomous agent
  - Don’t like one agent, take many: multi-agent systems
  - Reinforcement learning follows the agent paradigm

---

**Strong links with game theory and decision theory.**
**AI: A LITTLE BIT OF HISTORY**

- Fourth AI **Winter** (early 2000s)
  - Funding suffers from .com bubble burst!
- The Machine Learning **spring** (2000s)
  - **Data mining.** It works in industry!
  - Everything is a kernel! No, it’s a random field! No, it’s a Gaussian process!
  - Whatever, give me data!

The end of 2000s sets the stage for “the new AI”
AI: A LITTLE BIT OF HISTORY

- The timelines are different and subjective
- It’s mostly about public perception
- It’s mostly about funding
- It’s mostly about interests
- And a little bit about science

A lesson for AI forecasting
AI: A LITTLE BIT OF HISTORY

- Other battles brought different weapons!
  - The evolutionary route to AI: do like nature has done (must it be inefficient?)
  - Embodied AI: sensors and actuators (can there be intelligence without a body?)
  - The fuzzy route to AI: soft computing, computational intelligence, …
  - The solver route to AI: SAT solvers, modulo theories, …

Most of them are seen as techniques that can be combined with some other paradigms
AI: MAIN AREAS

- Knowledge representation:
  - Ontologies (e.g., semantic nets: IS-A and HAS)
  - Different kinds of logics: first-order, description logics, ...
  - Inference possible:
    - Tweety is a bird → Tweety can fly
AI: MAIN AREAS

- Heuristic search:
  - A*, ...
  - Beam Search
  - Hill Climbing & Gradient Descent
  - Montecarlo Tree Search
  - Simulated annealing
  - Ant colonies, swarms, …
AI: MAIN AREAS

- Planning & Scheduling:
  - Forward / Backward chaining
  - Temporal planning
  - Probabilistic planning
  - Markov Decision Processes
  - Constraint Satisfaction Problems

- Towers of Hanoi
AI: MAIN AREAS

- Machine Learning:
  - Linear models
  - Decision trees
  - Distance-based methods
  - Inductive programming
  - Neural networks (shallow and deep)
  - Bayesian and probabilistic approaches
  - Kernel methods and SVM
  - ...

- Supervised Learning
  - Classification
  - Regression
  - Ranking

- Unsupervised Learning
  - Clustering
  - Association Mining
  - Segmentation
  - Dimension Reduction

- Reinforcement Learning
  - Decision Process
  - Reward System
  - Recommendation Systems
AI: MAIN AREAS

- Pattern Recognition:
  - Machine Vision (filters, spectral analysis, …)
  - Speech Recognition (spectral analysis, HMM, …)

Now, most of the “old feature-based approaches” have been replaced by deep learning approaches.
AI: MAIN AREAS

- Natural Language Processing
  - Language understanding
  - Summarisation
  - Retrieval
  - Machine translation
  - Tagging
  - Sentiment analysis
  - …
AI: MAIN AREAS

- Deep Learning: it’s just a particular kind of machine learning technique!
- But it can learn representations...
**AI: MAIN AREAS**

- **Deep Learning:**
  - Usually supervised (especially at the end)
  - But they can be unsupervised
  - Can also be used to generate (e.g. GAN)
AI: MAIN AREAS

- Deep learning for everything?

The Holy Trinity of Deep Learning:
big data, big models, big compute

- IJCAI2018 debate, AAAI2019 debate, ...
  - Gary Marcus’s papers (the AI Contrarian)
  - Compositionality
    - Program induction (Tenenbaum et al., Science 2015)
  - Many breakthroughs really combine old AI and new AI (e.g., AlphaGo).
AI: THE ACTORS

- Who are the actors in AI?
AI: THE ACTORS

- The Researchers
  - Many still in academia.
  - Young and bright:
    - Now in the Tech Giants
  - Brain drain
- Major (massive) venues:
  - IJCAI, AAAI, NeurIPS, ICML, …
AI: THE ACTORS

- The Tech Giants
  - “Don’t be Evil” they say
- An oligopoly:
  - Alphabet
  - Amazon
  - Microsoft
  - Facebook
  - Apple
  - Baidu
  - Tencent
AI: THE ACTORS

- The Superpowers
  - USA
  - China
  - E.U.
  - Russia

AI Race Narrative. Putin: “the nation that leads AI will rule the world”
AI: THE ACTORS

- Good and Bad in romance

https://www.partnershiponai.org/

A heartbroken ugly might be dangerous: they claim they represent the people!
AI: WHAT CAN (AND CAN'T) IT DO FOR YOU?

Don’t look at the breakthroughs!

- Jump?
- Play board games
- Paint?
- Recognise sketches?
- Recommend?
- Recognise faces?
- Win TV quizzes?
- Play video games?
AI: WHAT CAN (AND CAN'T) IT DO FOR YOU?

- Better: look at the applications in your domain

- Tell me the future!
- Personal assistants are wonderful!

- Better than my doctor!
- No more driving! Or accidents!

- Buy me what I see!

A mix of success, unmet expectations and outright failures
AI: WHAT CAN (AND CAN'T) IT DO FOR YOU?

Brynjolfsson and Mitchell’s 21 item rubric:


1. Information needed to complete the task (inputs) and outputs specified in machine-readable format
2. Task information is recorded or recordable by computer
3. Task does not require a wide range of complex outputs (mental and/or physical)
4. Task feedback (on the success of outputs) is immediate or available through plentiful historical data.
5. The task output is error tolerant
6. It is not important that outputs are perceived to come from a human
7. Task does not require complex, abstract reasoning
8. Task is principally concerned with matching information to concepts, labels, predictions, or actions
9. Task does not require detailed, wide-ranging conversational interaction with a person
10. Task is highly routine and repeated frequently
11. Task is describable with rules
12. There is no need to explain decisions during task execution
13. Task convertible to answering choice questions, ranking, predicting a number, or grouping objects
14. Long term planning is not required to successfully complete the task
15. The task requires working with text data or might require working with text in the future:
16. The task requires working with image/video data now or in the future:
17. The task requires working with speech data or might require working with speech data in the future:
18. The task requires working with other types of data (other than text, image/video, and speech):
19. Many components of the task can be completed in a second or less
20. Each instance, completion, or execution of the task is similar to the other instances in how it is done
21. Actions in the task must be completed in a very specific order, and practicing to get better is easy
AI: WHAT CAN (AND CAN'T) IT DO FOR YOU?

- Rule of thumb:
  - If it can be done subconsciously (< 1sec), then AI can do it.
  - System 1 and system 2 (D. Kahneman’s “Thinking, Fast and Slow”)
  - If there’s lots of data, and knowledge is not crucial, then data science will probably create good patterns and models.

Today’s specialities! Perception and data crunching
AI: WHAT CAN (AND CAN'T) IT DO FOR YOU?
AI: THE FUTURE

- Simple extrapolations?

- Anthropocentric views?
  - Tegmark’s “Life 3.0”
AI: THE FUTURE

- Be careful with your predictions: be vague or be wrong!
  - White-collar jobs were in danger! (Steven Pinker “The Language Instinct”, 1995)
  - "Most fears of automation are misplaced. As the new generation of intelligent devices appears, it will be the stock analysts and petrochemical engineers and parole board members who are in danger of being replaced by machines. The gardeners, receptionists, and cooks are secure in their jobs for decades to come”

- Risk of automation (Frey and Osborne “The Future of employment” 2017):
  - “Financial Analysts” (0.23), “Chemical engineers” (0.017), “judges, magistrate judges, and magistrates” (0.4).
  - “Landscaping and groundskeeping workers” (0.95), “receptionists” (0.96), “cooks” (0.96)

Who’s right?
First the crystal ball: Tell me more about the future!

“I’m sorry Dave, I’m afraid I can’t do that”
AGI: Artificial General Intelligence

- Term introduced in the 2000s as opposed to Artificial Narrow Intelligence, but not properly defined.
- AGI/ANI not the same exactly as Strong AI / Weak AI
- Some people use AGI as synonym to HLAI (or HLMI)
- Even more poorly defined!!
Let’s use AGI in a less anthropocentric way:

Good performance on a wide range of environments
(up to a level of resources/difficulty)

Key notion: “generality”

How can we define notions of generality?
AI: THE FUTURE

- Superintelligence
  - The definition takes human as a reference (yet again)
  - It’s very much based on the notion of HLAI.
  - As it is ill-defined, the predictions are not verifiable

https://futureoflife.org/superintelligence-survey/
Cyborgs and uploads

- Why should humans keep the biological substrate?
- Common as a “natural” trend in augmentation
- Transhumanism

Many issues!
AI: THE FUTURE

▪ The singularity is coming and Kurzweil is near!
▪ Technological singularity (Good 1965):

“Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an ‘intelligence explosion,’ and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control”.

It’s just a hypothesis. What about the scale of intelligence and the physical resources?
AI: THE FUTURE

- Let’s do things right!
  - Asilomar Conference 2017: Beneficial AI 2017
    - Principles: https://futureoflife.org/ai-principles/

- Safety issues, privacy, inclusion, global standards, competition, ..
AI: THE FUTURE

Let’s do things right! AGI version

- Puerto Rico Conference 2019: Beneficial AGI 2019
- Future is always bright in the Caribbean: https://futureoflife.org/beneficial-agi-2019/

Destination, Safety, Strategy and Governance.
AI: HUMANS THEY ARE A-CHANGIN’ TOO

- AI and digitisation is changing our minds.
- Future AI extenders can change our capabilities too:
  - Memory processes
  - Sensorimotor interaction
  - Visual processing
  - Auditory processing
  - Attention and search
  - Planning, decision-making and acting
  - Comprehension and expression
  - Communication
  - Emotion and self-control
  - Navigation
  - Conceptualization
  - Quantitative and logical reasoning
  - Mind modeling and social interaction
  - Metacognition
Augmentation:

Positive effects: very empowering
- Dealing with cognitive decline of an aging population
- Equalising cognitive abilities

Negative effects: potential risks
- Atrophy and safety (google effect, dependency)
- Moral status and personal identity (take the money but don’t steal my phone)
- Responsibility and trust (it’s not my fault, it’s the gadget)
- Interference and control (Siri: “Why would you bring another woman back to our flat?”)
- Education and assessment (Cognitive extenders not allowed in the exam)
AI: HUMANS THEY ARE A-CHANGIN’ TOO

▪ “Where AI is heading?” Is this the right question?

Space of possible behaving systems / minds (Sloman 1984)
AI: HUMANS THEY ARE A-CHANGIN’ TOO

- Where is intelligence heading?
  - Human, artificial, hybrid, collective, …

Atlas of intelligence

Natural Behaviour

Human Behaviour

Artificial Behaviour
CONCLUSIONS

- The discussion about AI is too narrow and too focused on:
  - The volatile **present**:
    - Deep learning
    - The breakthroughs
    - The tech giants
  - The uncertain **future**:
    - Jobs, superintelligence and the singularities
- Try to look **beyond** this duality
  - Humans are changing too
  - It’s not AI, but intelligence what we need to track
  - Measurement is key for this.... More on Wednesday!!!!