Certificate I: Understanding AI and Machine Learning in Africa

Course AIML01: Artificial Intelligence – Past, Present, and Future

Module 1: What is AI, where did it come from, and where is it taking us?

Lecture 3: The End of the AI Winter

Carnegie Mellon University
Africa

Learning Objectives

1. Identify the reasons for the end of the AI winter
2. Explain how the field of AI enlarged to embrace techniques from other fields
3. Explain the impact of machine learning and learning based on deep neural networks
4. Identify landmark successes in probabilistic knowledge engineering
5. Explain the need for trustworthy, explainable techniques and the challenges this poses for AI based on deep learning

Lecture Contents

Factors that contributed to the end of the AI Winter

- Research in the late 1990s laid the foundations for the end of the AI winter and the beginning of a period of great success of AI today

- For example, this period saw the achievement of two key results:

  - Convolutional neural networks (CNNs) by LeCun et al. (1998)
  - Long short-term memory (LSTM) by Hochreiter and Schmidhuber (1997)

- But some other developments were needed before the power of these results could be harnessed

We will explain these terms in more detail in Module 2, Lecture 2

For the moment, we will just remark that LSTMs are a very effective form of recurrent neural network and CNNs are an advanced form of trainable perceptron, both of which we mentioned in the previous lecture

Factors that contributed to the end of the AI Winter

These were

- Artificial neural networks with deeper network topologies

  Networks with many more layers than had been used in the mid-1980s to mid-1990s

- New learning techniques

  We will say more about these later in this course and cover them in detail in subsequent courses

- Much greater computing power in the form of graphic processor units (GPUs)

  Originally developed to allow much faster computer graphics, the operations they perform are also ideally suited to those needed in artificial neural networks

- Much larger datasets to train the networks

  This was a major breakthrough, enabled to a great extent by the world-wide web
  For example, in a recognition or classification task, you typically need 1000 training examples for each class you want to recognize.

Factors that contributed to the end of the AI Winter

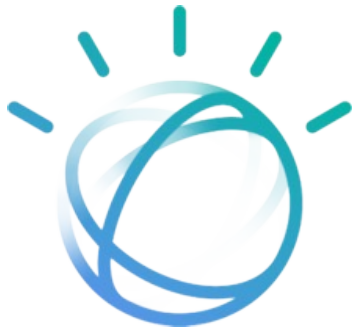
Since 2011, AI based on deep learning had been very successful in many difficult applications, such as

- Computer vision
- Robotics
- Autonomous driving
- Natural language processing
- Sentiment analysis
- Medical imaging
- Several other domains

Factors that contributed to the end of the AI Winter

- Also in this period, there were very successful probabilistic approaches to AI

- For example, the Watson system from IBM
  - Named after Thomas J. Watson, the founder of IBM
  - Watson won the TV show *Jeopardy!* in 2011
  - By answering rich natural language questions over a very broad domain of topics
  - Beating two human champions



Watson's avatar
https://en.wikipedia.org/wiki/Watson_(computer)

Factors that contributed to the end of the AI Winter

- The success of Watson was the result of probabilistic knowledge engineering

This is a mathematical way of dealing with uncertainty about the knowledge

You can read about the AI behind Watson in (Ferrucci et al., 2010)

- Integrated many knowledge sources
- Exploited many techniques for
  - Search
  - Hypothesis formulation
  - Hypothesis evaluation

Many possible solutions exist: one of the goals of AI is to constrain the search for the best solution

A hypothesis is a conjecture about what might the the correct solution or answer to the problem
Watson evaluates many hypotheses and selects the best one

Factors that contributed to the end of the AI Winter

- AI had finally come of age
  - It was able to provide reliable solutions
  - To complex problems
  - In many application domains
- John McCarthy once remarked that

"As soon as it works, no one calls it AI anymore"

Quoted in (Meyer, 2011)

This was no longer true, but it was also no longer clear exactly what AI was

The Diversification of AI

Many techniques which had traditionally not been part of AI, for example

– data science
– optimization theory
– control theory

were now included under the umbrella of AI

When you read "AI – The Revolution Hasn't Happened Yet" by Michael I. Jordan (2019) after the last lecture, you will have been struck by the following

"The developments now being called AI arose mostly in the engineering fields associated with low-level pattern recognition and movement control, as well as in the field of statistics, the discipline focused on finding patterns in data and on making well-founded predictions, tests of hypotheses, and decisions."

These techniques helped to make it more robust and more reliable, especially in narrowly-specificed applications

The Need for Explainable, Trustworthy AI

But there is a price to pay for this success

The expansion of AI represents a departure from its original focus on strong AI:

Human-centred AI

The basis for decisions is understandable by and communicable to humans

Human-level AI

Exhibits the same versatility and flexibility that humans do

For more details, see AI at a Crossroads by Luc Steels (2020)

The Need for Explainable, Trustworthy AI

In contrast, the highly-successful AI that derives from the behaviorist tradition

- connectionism (i.e. artificial neural networks)
- probabilistic data science models
- statistical machine learning

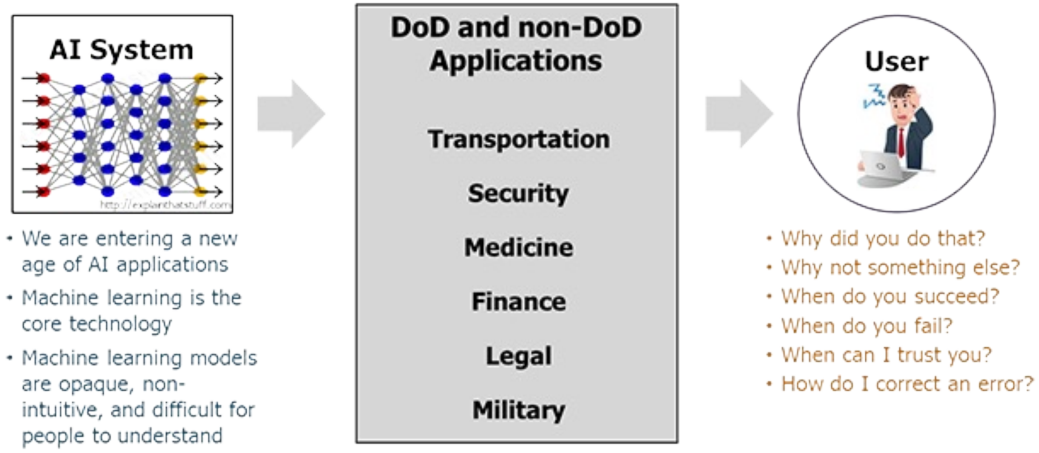
are black box systems

- the basis for their decision-making is not open to scrutiny by humans in any meaningful way
- at least not at present

The Need for Explainable, Trustworthy AI

As a result, much time and effort today is being spent to make this approach to AI

- More explainable … XAI (eXplainable AI)
- More trustworthy
- This is not a trivial problem



The Need for XAI
https://www.darpa.mil/program/explainable-artificial-intelligence

The Need for Explainable, Trustworthy AI

- Deep learning AI achieves exceptional performance due to its statistical nature

- But it is can make errors if presented with

  Data in the training set that isn't representative of majority of the data

  – Outliers

  – Data that are not drawn from the distribution on which the system is trained

Data that aren't modelled by the training set,
that is, the data are not captured by a probability distribution of the data.
This simply means that the model would attach a probability of zero that that data,
i.e., it would infer that the data never occurs.
This, of course, is the problem: it can occur when the AI system is being used and,
consequently, it is likely to lead to an error in the result.

The Need for Explainable, Trustworthy AI

This means that the data set is not representative of all the possible instances that can occur

- Such cases may be very common
- When systems are trained on data sets that exhibit implicit or explicit bias
  - The bias is modelled during training
  - The systems inevitably operate in a biased manner when subsequently used in applications
- Even if the bias in the original data set was not intended

The Need for Explainable, Trustworthy AI

- We return to the issue of trustworthy AI in Module 4, Lecture 3

- In the next module, we take a closer look at the constituent approaches of AI:

  - Symbolic AI
  - Connectionist AI
  - Statistical machine learning

This should be viewed as a preview of the material to be covered in the remaining courses in this program

Lecture Summary

1. Breakthrough developments in neural networks laid the foundation for the end of the AI winter and the success of AI in the 2000s

2. Other enabling factors included the use of neural networks with many layers, new learning techniques, large datasets to train the networks, and GPU technology to facilitate the training

3. AI absorbed many techniques from other engineering disciplines including statistical data science, machine learning, control theory, and optimization

4. This migration away from human-centred and human-level AI, one of the core motivations behind strong AI, is both the source of much of its success and the creation of a new challenge to make these approaches to AI both explainable and trustworthy

Recommended Reading

Cangelosi A and Vernon D (2022). "Artificial Intelligence: Powering the Fourth Industrial Revolution", in *EPS Grand Challenges: Physics for Society at the Horizon 2050*, coordinated by the European Physical Society.
http://vernon.eu/publications/2022_Cangelosi_Vernon.pdf

Jordan, M. I. (2019). Artificial Intelligence — The Revolution Hasn't Happened Yet. Harvard Data Science Review, 1(1).
https://doi.org/10.1162/99608f92.f06c6e61

Steels L (2020) AI at a Crossroads.
https://www.muhai.org/blog/12-artificial-intelligence/12-ai-at-a-crossroads

References

Ferrucci, D, Brown, E, Chu-Carroll, J, Fan, J, Gondek, D, Kalyanpur, A, Lally, A, Murdock, JW, Nyberg, E, Prager, J, Schlaefer, N, and Welty, C (2010). Building Watson: An Overview of the DeepQA Project, AI Magazine.
https://www.aaai.org/Magazine/Watson/watson.php

Hochreiter, S and Schmidhuber, J (1997). Long Short-Term Memory, Neural Computation 9(8):1735-1780.
https://www.bioinf.jku.at/publications/older/2604.pdf

LeCun, Y, Bottio, L, Bengio, Y, Haffner, P (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11): 2278-2324.
http://yann.lecun.com/exdb/publis/pdf/lecun-98.pdf.

Meyer B (2011) John McCarthy. CACM Blogs.
https://cacm.acm.org/blogs/blog-cacm/138907-johnmccarthy/fulltext