Carnegie Mellon University Africa
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

Welcome to Lecture 3 of Module 1 in which we will discover why AI and machine learning became so successful in the 2000s.

In this lecture, we will discuss the factors that contributed to the end of the AI Winter. We will explore the implications of this and the consequent diversification of techniques that now fall under the umbrella of AI. We will also discuss the reasons why these developments created an urgent need for explainable, trustworthy AI. We will finish up by summarizing what we have covered and identifying the articles that you should read to consolidate what you have learned.

We have five learning objectives, so that, after studying the material covered in this lecture, you should be able to do the following.

1. Identify the reasons for the end of the AI winter.

2. Explain how the field of AI subsequently evolved.

3. Explain the need for trustworthy, explainable techniques and the challenges this poses for AI based on different forms of machine learning.

Slide 1 Welcome to Lecture 3 of Module 1 in which we will discover why AI and machine learning became so successful in the 2000s.

Slide 2 In the previous lecture, we noted that 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

Two of the most important breakthroughs are the development of the Convolutional Neural Network by Yann LeCun and the long short-term memory (LSTM) by Sepp Hochreiter and Jürgen Schmidhuber

We will discuss CNNs and LSTMs 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.

However, some other developments were needed before the power of these results could be harnessed

Slide 3 These were

1. Artificial neural networks with deeper network topologies

2. New learning techniques

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

4. Much larger datasets to train the networks

These deep networks have many more layers than had been used in the mid-1980s to mid-1990s.

We will say more about the new learning techniques later in this course and we will cover them in detail in subsequent courses.

GPUs were originally developed to allow much faster computer graphics but the operations they perform are also ideally suited to those needed in artificial neural networks.

The ability to generate very large datasets 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.

Slide 4 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

Slide 5 This period also saw the development of some landmark probabilistic approaches to AI

Perhaps the most celebrated of which is the Watson system from IBM (named after its founder Thomas J. Watson) and which won the TV show *Jeopardy!* in 2011, beating two human champions in answering rich natural language questions over a very broad domain of topics.

Slide 6 The success of Watson was the result of probabilistic knowledge engineering

that integrated many knowledge sources

and exploited many techniques for

search, hypothesis formulation, and hypothesis evaluation

The probabilistic approach uses probability theory to deal in a principled way with uncertainty about the knowledge

Many possible solutions for a problem or answers to a question exist: one of the goals of AI is to constrain the search for the best solution or answer.

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

Slide 7 AI had finally come of age, yielding 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” (Meyer, 2011).

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

Slide 8 At this point, many techniques which had traditionally not been part of AI,

e.g. data science, optimization, and control theory

began to be included under the umbrella of AI, contributing significantly to its success.

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

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."

Slide 9 This expansion of AI represents a departure from its original focus on strong AI

which is human-centred, in the sense that the basis for its decisions is understandable by and communicable to humans,

and human-level, in the sense that it displays the same versatility and flexibility that humans do

Slide 10 In contrast, the highly-successful AI that derives from the behaviorist tradition, including

connectionism,
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.

Slide 11 Consequently, much effort today is being expended to make this approach to AI

more explainable (often shortened to XAI (eXplainable AI) and more trustworthy

This is not a trivial problem.

Slide 12 While deep learning AI may achieve exceptional performance, by virtue of its statistical nature,

it is susceptible to errors when faced with outliers

or data that are not drawn from the distribution on which the system is trained

Outliers are data in the training set that isn't representative of majority of the data.

In the second case, the data that aren't even 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 to that data, i.e., it would infer that the data never occurs.

This, of course, is exactly 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.

Slide 13 Such outliers may be very common, as is the case when systems are trained on data sets that exhibit implicit or explicit bias.

The bias is modelled during training

The systems then inevitably operate in a biased manner when subsequently used in applications,

even if the bias in the original data set was not intended.

Slide 14 We return to the issue of trustworthy AI in Module 4, Lecture 3

In the next lecture, 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.

To summarize:

- Breakthrough developments in neural networks laid the foundation for the end of the AI winter and the success of AI in the 2000s.
- 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.
- AI absorbed many techniques from other engineering disciplines including statistical data science, machine learning, control theory, and optimization.
- 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.

Here is some recommended reading. It will help you understand the basis of the current success of AI, the challenges that remain in our efforts to achieve the versatility and flexibility of human-level AI, and the need to develop more explainable and trustworthy AI, especially for AI based on statistical machine learning and deep neural networks.

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

Here are the references cited to support the main points in what we covered today.

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