Carnegie Mellon University Africa
Certificate I: Understanding AI and Machine Learning in Africa

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

## Module 4: Future Challenges
## Lecture 4: Intelligence, Brains, and Consciousness

Welcome to Lecture 4 of Module 4, the last in this course, in which we will study the links between intelligence, brains, and consciousness.

In this lecture, we will answer three questions:

1. What is intelligence,
2. Why does intelligence matter, and
3. Why do we have brains?

We highlight the predictive nature of brains. We will consider the challenge of understanding the nature of consciousness. We consider what the future might hold for AI and what we need to do to make sure it serves all of us well. We will finish up by summarizing what e have covered and identifying the articles that you should read to consolidate what you have learned

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

1. Explain what it means to be intelligent and what brains allow us to do

2. Discuss what the future might hold as we pursue the challenge of designing and building artificial intelligence systems, possibly ones that are autonomous or conscious.

Slide 1 Welcome to Lecture 4 of Module 4, the last in this course, in which we will study the links between intelligence, brains, and consciousness.

Slide 2 Why does intelligence matter?

Indeed, what is intelligence?

You should be surprised by the fact that we haven't yet answered this question.

We avoided it by defining AI as the endeavour to mimic and augment human intelligence and we avoided saying what human intelligence is.

This has not caused us any problems because we all take human intelligence for granted.

But let's pause here to consider what we mean by intelligence and why it is so important to be intelligent.

The answers will reveal why we have brains — the seat of intelligence — and what brains do.

This will lead to other questions that are harder to answer, such as how consciousness fits into the picture.

Slide 3 Let's start by answering the question: what is intelligence?

There are many possible answers.

Slide 4 But the one that has the most appeal derives from the answer to a different question: why do we have brains?

The neuroscientist Daniel Wolpert provides an unexpected but compelling answer.

He argues that we have brains to allow us to control movement (Wolpert, 2011).

Slide 5 This mirrors what Francisco Varela and Umberto Maturana say about knowing:

"Knowing is effective action" (Maturana and Varela, 1987).

Slide 6 From this perspective, we view intelligence as the way to be effective in our control of our movements and in the way we act in the world.

Slide 7 The key to understanding why intelligence is so important — and so difficult to achieve — is to see that the number of possible ways we can move and act,

and the number of possible outcomes of these movements and actions,

is infeasibly large if we are to consider all the possibilities when choosing the best one, or even a good one.

Slide 8 This is what Allen Newell and Herbert Simon pointed out in their Turing Award (a sort of Nobel Prize for computer scientists) lecture:

"The task of intelligence, then, is to avert the ever-present threat of the exponential explosion of search" (Newell and Simon, 1976)

That is, the search for good ways to act.

Newell and Simon were referring to the search for the solution to a problem, but it amounts to the same thing.

This is a satisfyingly straightforward and very practical way of understanding intelligence and the brains that give rise to intelligence.

Slide 9 However, brains are even better than that.

They also predict the need to act and the outcome of those actions, and they do so all the time, at every instant, as we act and as we anticipate the future, milliseconds ahead, seconds ahead, hours, days, years.

Slide 10 It has been argued that brains are, in effect, probabilistic

(meaning they can deal effectively with uncertainty)

prediction machines.

Slide 11 But what then of consciousness?

Can intelligent machines also be conscious?

Can we take AI even further and build machines with artificial consciousness?

Slide 12 Many people think this is a distinct possibility.

Indeed, according to Paul Vershure,

“understanding the nature of consciousness is one of the grand outstanding scientific challenges”

Slide 13 and he proposes a scientifically-grounded approach to addressing the challenge of answering the question of what consciousness is and how physical systems can give rise to it.

Slide 14 He maintains that this challenge stands at the centre of knowing what it is to be human.

In this, he echoes the original motivation for studying AI at its inception over sixty-five years ago:

To understand human intelligence and use AI to support the way humans live, work, and relate to one another

Slide 15 We have reached the end of the course.

We hope you enjoyed it.

We will wrap up with a summary of the story of AI so far

and brief insight into what might lie ahead

We present this as a narrative

We hope you will take the time to read it and reflect.

Slide 16 Artificial intelligence impacts on all aspects of human activity:

it automates tasks,

it assists with decision-making,

it augments and extends our cognitive capabilities,

and it can even operate autonomously, if we allow it,

without recourse to human oversight.

Slide 17 AI began as an attempt to understand and replicate human intelligence, initially taking two routes to that goal,

one via connectionism,

and one via symbolic computationalism,

reflecting their inspiration by behaviorist and constructivist psychology, respectively.

Slide 18 These two approaches waxed and waned in their own respective ways over the decades,

to be joined in the 1980s by machine learning and in the 1990s by statistical machine learning,

probabilistic inference networks,

and other established disciplines in computer science.

Slide 19 Breakthroughs in deep neural network learning and deep neural network topologies,

aided by very large data sets and equally large increases in processing power,

yielded great success in many application domains.

Slide 20 The symbolic knowledge representation and reasoning approach also developed rapidly,

especially in cognitive architectures.

Slide 21 The hybrid paradigm, combining symbolic approaches and sub-symbolic connectionist approaches, became prevalent,

e.g., in cognitive architectures such as Soar
ACT-R
CLARION
among others.

Slide 22 While the success of statistical machine learning in narrow targeted applications yielded great success,

it did so at the expense of losing focus on AI's original goal of understanding and replicating human-level intelligence.

Slide 23 There has been a resurgence of interest in what is now known as Artificial General Intelligence (AGI) in cognitive science and cognitive systems.

Still, the ultimate goal of replicating the versatility of human cognition remains elusive and it is unclear when it will be achieved.

Slide 24 What is certain is that the AI quest will continue and AI in its many guises will continue to permeate our lives, change them, hopefully for the better.

Slide 25 In seeking to steer the path to the future, it is likely that other strands of thinking will be woven into the fabric of AI,

especially concerning the trustworthiness of AI in autonomous systems.

Slide 26 AI will play an important role in serving the bigger agenda of creating self-maintaining systems

that can operate robustly and prospectively in the face of uncertainty

and that can continually develop through self-programming as they interact with and learn from the world and the people in it.

Slide 27 While there is much important work yet to be done to promote the development of

democratized,
trustworthy,
ethical AI

in the developed and developing worlds,

an equal challenge will be how to control the role of AI in autonomous systems,

possibly conscious ones, where the relationship with humans is no longer symbiotic.

Slide 27 We are far from that point at present, but it is likely we will reach it, and everything will change quickly when we do.

Slide 29 In Ernest Hemingway's novel *The Sun Also Rises* there is a dialog between two characters which goes as follows.

Slide 30 "How did you go bankrupt?" Bill asked.
"Two ways," Mike said.
"Gradually and then suddenly."

Slide 31 And so too it will be with autonomous AI:
it will happen gradually and then suddenly.

Our collective responsibility is to work together in a directed manner during the present gradual phase so that,

when the full impact of AI is suddenly felt,

it will be for the greater good of all humankind.

To summarize:

1. The brain allows agents to control their movements.

2. Intelligence is the way to be effective in controlling our movement and the way we act in the world.

3. Intelligence allows us to consider how we can act most effectively, without having to search through the intractably-large number of choices.

4. Brains can be viewed as probabilistic prediction machines.

5. Understanding the nature of consciousness is one of the grand challenges of science and is closely linked to the challenge.

Here is some recommended viewing.

Wolpert D (2011) The Real Reason for Brains.
https://www.youtube.com/watch?v=7s0CpRfyYp8

Here is some recommended reading. It is the same as the first lecture, AIML01-01-01, but these articles will be much more meaningful, now that you have finished the course.

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

Delmus Alupo C, Omeiza D, and Vernon D (2022) "Realizing the Potential of AI in Africa", in *Towards Trustworthy Artificial Intelligence Systems*, M. I. Aldinhas Ferreira, O. Tokhi (Eds.), Intelligent Systems, Control and Automation: Science and Engineering. Springer.
http://vernon.eu/publications/2023_Delmus_Alupo_et_al.pdf

Vernon D (2019). "Robotics and Artificial Intelligence in Africa", IEEE Robotics & Automation Magazine, Vol. 26, No. 4, pp. 131-135.
http://vernon.eu/publications/19_Vernon_RAM.pdf

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


Anderson JR, Bothell D, Byrne MD, Douglass S, Lebiere C, Qin Y (2004) An integrated theory of the mind. Psychological Review 111(4):1036–1060. http://act-r.psy.cmu.edu/wordpress/wp-content/uploads/2012/12/526FSQUERY.pdf

Downing K (2009) Predictive models in the brain. Connection Science 21:39–74. https://www.tandfonline.com/doi/full/10.1080/09540090802610666

Friston KJ (2010) The free-energy principle: A unified brain theory? Nature Reviews Neuroscience 11(2):127–138. https://www.nature.com/articles/nrn2787

Laird JE (2012) The Soar Cognitive Architecture. MIT Press, Cambridge, MA. https://direct.mit.edu/books/book/2938/The-Soar-Cognitive-Architecture

Maturana H, Varela F (1987) The Tree of Knowledge — The Biological Roots of Human Understanding. New Science Library, Boston & London. https://www.shambhala.com/the-tree-of-knowledge.html

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search (ACM 1975 Turing award lecture), Communications of the ACM, 19, 3, 113-126. https://dl.acm.org/doi/10.1145/360018.360022

Seth AK (2015) The cybernetic Bayesian brain — from interoceptive inference to sensorimotor contingencies. In: Metzinger T, Windt JM (Eds.) Open MIND, vol 35, Frankfurt am Main: MIND Group, pp 1–24. https://open-mind.net/papers/the-cybernetic-bayesian-brain/at_download/paperPDF

Sun R (2016) Anatomy of the Mind: Exploring Psychological Mechanisms and Processes with the Clarion Cognitive Architecture. Oxford University Press. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199794553.001.0001/acprof-9780199794553

Verschure PFMJ (2016) Synthetic consciousness: the distributed adaptive control perspective. Philosophical Transactions of the Royal Society of London, Series B 371. https://royalsocietypublishing.org/doi/full/10.1098/rstb.2015.0448