

# Reinforcement Learning

Aldo Faisal with contributions  
by Ed Johns and Paul Bilokon

Imperial College London

October 2022

# Contents

Our first half of the course is an introduction to Reinforcement Learning (RL)

- 1 Mathematical Foundations
- 2 Markov Decision Process (much more than a Markov Process)
- 3 Dynamic Programming (not the way you know it)
- 4 Monte-Carlo Learning
- 5 Temporal Difference Learning
- 6 Function Approximation Methods (Deep RL, Policy Gradients)
- 7 Reinforcement Learning Outlook

# Acknowledgements

The lecture notes for this course are based on material from colleagues' books or lecture notes by Peter Dayan, Rich Sutton, Andrew Barto, David Silver, Zoubin Ghahramani, Sergey Levine, and many others, as well as older versions of my own courses (MLNC 2009-2015, CO424H, 2015-2018).

We especially would like to thank the following people for their input and suggestions

- Manon Flageat
- Athanasios Vrontzos
- Paul Festor
- Luke Dickens
- Marc Deisenroth
- Luchen Li

- Ali Shafati

and the feedback of many excellent students at Imperial College who took or take this course.

## Admin: Assessment

- Course structured in Lectures, Q&A and Computer Labs. Content and learning occurs in all of them, they may be complementary to each other (i.e you will learn things in one that you do not learn in the other explicitly).
- Entire module: 25% Coursework 1, 25% Coursework 2, 50% exam.
- In Computer Labs you can work on Lab Assignments and Coursework.
- Initially more weight on lectures to get us up to speed, then less lectures more labs.
- We want to use Teams for the lab work and EdStem for all Q&A and support where possible.

# Admin: Teaching

Course has two parts taught by

① Dr Paul Bilokon (Part 1)

- Students watch 1 hour of pre-recorded lecture on their own
- Q&A (over Teams)
- Interactive computer labs

② Prof Aldo Faisal (Part 2)

- Students watch 1 hour of pre-recorded lecture on their own
- Q&A (over Teams)
- Interactive computer labs

## Admin stuff

- We want to use Teams for the lab work and EdStem for all Q&A and support where possible.
- The college expects that students invest time outside the course and invest about 1 hour per hour of course in their own time, plan your calendar accordingly.
- Labs and Courseworks may have programming tasks – we expect you to know how to program (i.e. you should be able to pick up by now on your own an unknown programming language as you go along).

## Ed (Ed Stem)

- We are using **Ed Discussion** for class Q&A.
- This is the best place to ask questions about the course, whether curricular or administrative.
- You will get faster answers here from staff and peers than through email.
- Here are some tips:
  - Search before you post;
  - Heart questions and answers you find useful;
  - Answer questions you feel confident answering;
  - Share interesting course related content with staff and peers.
- For more information on Ed Discussion, you can refer to the Quick Start Guide:  
<https://edstem.org/quickstart/ed-discussion.pdf>

## How we will use Ed I

- To encourage student engagement in Ed we will leave questions for 2 working days so that other students have the opportunity to answer.
- Coursework related questions are fielded on Ed up to two days before the coursework deadline, so that the answer will become available in time for submission.
- We cannot provide answers to question that directly solve part of the coursework.
- We will accept exam related questions up to 9am the day before the exam, and answer them by 6pm on the day before the exam.
- We cannot provide answers to questions that directly solve exam-style or coursework-style question (so as to systematically explore all possible questions)



## How we will use Ed II

- General questions will be fielded by GTAs from Monday to Friday during normal working hours.

## Text books

- Richard Sutton and Andrew Barto (2018, new edition) "Reinforcement Learning: An Introduction", MIT Press. Available online and in the library. There is also a great new version 2.0 you can find on their homepage with a slightly different notation).
- Csaba Szepesvari (2010) "Algorithms for Reinforcement Learning", Morgan Claypool. Available online.
- Mathematics background: Marc P. Deisenroth, A. Aldo Faisal & Cheng Soon Ong (2020) "Mathematics for Machine Learning", Cambridge University Press, available freely on the web <https://mml-book.com>

# Outline I

- 1 Motivation
- 2 Reinforcement Learning 101
- 3 Lets go Markov
- 4 Markov Decision Process
- 5 Dynamic Programming
- 6 Model-Free Learning
- 7 Model-Free Control

# Section overview

## Motivation

- 1 Motivation
  - Artificial Intelligence
- 2 Reinforcement Learning 101
- 3 Lets go Markov
- 4 Markov Decision Process
- 5 Dynamic Programming
- 6 Model-Free Learning
- 7 Model-Free Control

# Artificial Intelligence I

## Definition (Artificial Intelligence)

**Artificial Intelligence** is a question: How do we build systems that solve tasks for which humans need intelligence?

## Definition (Machine Learning)

**Machine Learning** is the contemporary answer to the AI question: Methods, algorithms and data structures that **learn** to solve such tasks from data.

# Artificial Intelligence II

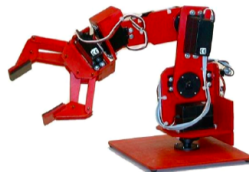
## Definition (Big Data)

**Big Data** encompasses methods to handle and understand large data sets, and is composed of **Data Science** (how to analyse the world in a data-driven manner) and **Data Engineering** (how to collect, store, process, clean, maintain large data).

- AI means often an intelligent system that embodies an entire practical solution (e.g. a self-driving car), while machine learning is more focussed on the algorithm (e.g. python code).
- Machine Learning puts the engineers efforts on building a system that learns to solve a problem, instead of engineering a system that solves a problem – for the first time in human history.

# Uses of Reinforcement Learning

Reinforcement learning solves problems of **control**, i.e. choosing the "best" / optimal action at the right time.



# Learning to control our body: Long history of high jump



Olympic control policies with  
Gold medal reward: High-jump



Ethel Catherwood  
(Canada), 1928  
gold medal



Cornelius Johnson  
(USA), 1936,  
gold medal



Dick Fosbury  
(USA), 1968,  
gold medal





# History of success in reinforcement learning:

- Backgammon (Tesauro, 1994)
- Inventory Management (Van Roy, Bertsekas, Lee & Tsitsiklis, 1996)
- RoboCup Soccer (e.g. Stone & Veloso, 1999)
- Helicopter drone control (e.g. Ng, 2003, Abbeel & Ng, 2006)
- Few-shot learning of pendulum swing up (Deisenroth et al, 2011)
- Playing Atari video games - from pixels to joystick command (DeepMind, 2015)
- Grand-master level Go playing (DeepMind, 2016)
- AI Clinician (Komorowski et Faisal, 2018)
- Solving Rubik's cube with a robot hand (OpenAI, 2019)

# Examples: Learning to play Go against human grand master

The screenshot shows the top portion of a web browser displaying a page from the journal Nature. The header is dark red with the 'nature' logo in white. Below the logo, a navigation bar contains links for Home, News & Comment, Research, Careers & Jobs, Current Issue, Archive, Audio & Video, and For Authors. A secondary navigation bar shows 'Archive', 'Volume 529', 'Issue 7567', 'Articles', and 'Article'. The main content area has a dark red background with the text 'NATURE | ARTICLE' and a Japanese link '日本語要約'. The article title 'Mastering the game of Go with deep neural networks and tree search' is prominently displayed. Below the title, the authors' names are listed in blue links. Further down, there are links for 'Affiliations', 'Contributions', and 'Corresponding authors'. The article's publication details, including volume, issue, date, and DOI, are provided. At the bottom of this section, there are five buttons: PDF, Citation, Reprints, Rights & permissions, and Article metrics.

**nature** International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive > Volume 529 > Issue 7567 > Articles > Article

NATURE | ARTICLE

[日本語要約](#)

## Mastering the game of Go with deep neural networks and tree search

[David Silver](#), [Aja Huang](#), [Chris J. Maddison](#), [Arthur Guez](#), [Laurent Sifre](#), [George van den Driessche](#), [Julian Schrittwieser](#), [Ioannis Antonoglou](#), [Veda Panneershelvam](#), [Marc Lanctot](#), [Sander Dieleman](#), [Dominik Grewe](#), [John Nham](#), [Nal Kalchbrenner](#), [Ilya Sutskever](#), [Timothy Lillicrap](#), [Madeleine Leach](#), [Koray Kavukcuoglu](#), [Thore Graepel](#) & [Demis Hassabis](#)

[Affiliations](#) | [Contributions](#) | [Corresponding authors](#)

Nature **529**, 484–489 (28 January 2016) | doi:10.1038/nature16961  
Received 11 November 2015 | Accepted 05 January 2016 | Published online 27 January 2016

[PDF](#) [Citation](#) [Reprints](#) [Rights & permissions](#) [Article metrics](#)

### Abstract

[Abstract](#) · [Introduction](#) · [Supervised learning of policy networks](#) · [Reinforcement learning of policy networks](#) · [Reinforcement learning of value networks](#) · [Searching with policy and value networks](#) · [Evaluating the playing strength of AlphaGo](#) · [Discussion](#) · [Methods](#) · [References](#) · [Acknowledgements](#) · [Author information](#) · [Extended data figures and tables](#) · [Supplementary information](#) · [Comments](#)

The game of Go has long been viewed as the most challenging of classic games for artificial intelligence owing to its enormous search space and the difficulty of evaluating board positions and moves. Here we introduce a new approach to computer Go that uses 'value networks' to evaluate board positions and 'policy networks' to select moves. These deep neural networks are trained by a novel combination of supervised learning from human expert games, and reinforcement learning from games of self-play. Without any lookahead search, the neural

# Reinforcement Learning Framework

- **Agent** interacts with **environment** to gain knowledge
- **Explores** and receives **rewards**
- **Actions** change the **state** of the environment
- Choose actions to **maximize long-term reward**

# AI & Machine Learning @ Imperial I

The Cross-Faculty Network in Artificial Intelligence (or simply @ImperialAI ) is the central hub for AI at Imperial College. We have over 200 Faculty members and over 800 PostDoc and PhD students members. The networks is open to all students at Imperial and we form London's and the UK's largest academic AI community.

- Follow us on Twitter @ImperialAI
- Visit our homepage <https://www.imperial.ac.uk/ai>
- Signup to our network announcements via email <https://mailman.ic.ac.uk/mailman/listinfo/ai-talks>
- Enquiries to AI Network Manager [ai-net-manager@imperial.ac.uk](mailto:ai-net-manager@imperial.ac.uk)
- Aldo is its current Speaker elected by its academics, i.e. that he answers if somebody 'calls' Imperial about AI.

# AI & Machine Learning @ Imperial II

- AI Talks Mailing List
  - If you are interested in receiving emails regarding talks, industry events, seminars and other activities of the network, please sign up here:  
<https://mailman.ic.ac.uk/mailman/listinfo/ai-talks>
  - Target audience: anybody interested in AI & Machine Learning
- Machine Learning Tutorials
  - Two hours lectures by leaders in the field from outside Imperial
  - Target audience: final-year students, PhD students, post-docs and academics with an interest in machine learning.
  - Announcements to go out over AI Talks.