CS4618: Artificial Intelligence I

Introduction

Derek Bridge
School of Computer Science and Information Technology
University College Cork

Initialization

In [1]:

```
%reload_ext autoreload
%autoreload 2
%matplotlib inline
```

In [2]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

About CS4618 & CS4619

Lecturer:	Derek Bridge, Room G-61, Western Gateway Building d stop bridge amphora cs plip ucc plop ie www.cs.ucc.ie/dbridge.html (http://www.cs.ucc.ie/dbridge.html)
Credit weighting:	Both modules are 5-credit electives
Prerequisites:	CS2515, CS2516 (Algorithms & Data Structures) Lots of notation and formulae, especially vectors and matrices Python programming
Course web site:	www.cs.ucc.ie/~dgb/courses/ai1.html (http://www.cs.ucc.ie/~dgb/courses/ai1.html)

About CS4618 & CS4619

Lectures:	2 × 1 hr per week
Labs:	1 × 2 hr per week
Private study:	At least 2 hrs per week

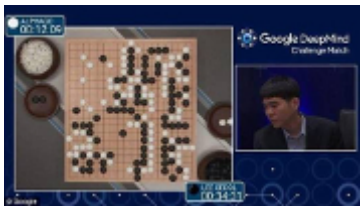
About CS4618 & CS4619

Examination:	1.5 hr written exam (80% of the marks)
Continuous assessment:	Short programming assignment(s) (20% of the marks)
How to fail:	Skip lectures & labs; avoid private study; cram just before; expect the exam to be a memory test
How to pass:	Attend lectures & labs; summarize the notes; tackle the lab activities properly; expect a problem-solving exam

Plagiarism

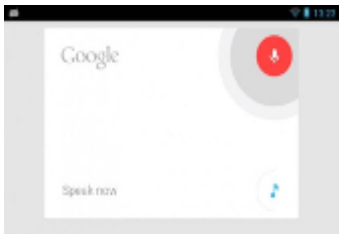
1. Plagiarism is presenting someone else's work as your own. It is a violation of UCC Policy and there are strict and severe penalties.
2. You must read and comply with the UCC Policy on Plagiarism www.ucc.ie/en/exams/procedures-regulations/ (http://www.ucc.ie/en/exams/procedures-regulations/)
3. The Policy applies to all work submitted, including software.
4. You can expect that your work will be checked for evidence of plagiarism or collusion.
5. In some circumstances it may be acceptable to reuse a small amount of work by others, but only if you provide explicit acknowledgement and justification.
6. If in doubt ask your module lecturer prior to submission. Better safe than sorry!

AI in the news

1997		2016	
2011		2017	

- What makes chess difficult?
- Quiz shows? Go? Poker?

But AI is much more pervasive

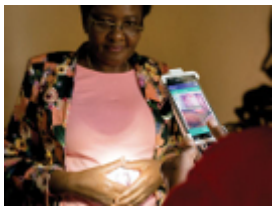



Why do we want to build intelligent systems?

- The main goal of AI is to build smarter tools
- Some rough-and-ready definitions:
  - A **tool** is an object that a creature uses intentionally to transform another object (possibly including the creature itself)
  - A tool is only **useful** if it:
    - **increases throughput**: more output for the same or less input (better use of scarce resources, including time); and/or
    - **reduces externalities**: fewer undesirable side-effects (e.g. less pollution, lower risks to life and limb, etc.)

Examples

- So, based on the previous definitions, why do we want to build…

…self-driving vehicles?	…intelligent news filters?	…medical diagnosis tools?	…bomb disposal robots?
			

Is AI even possible?

No: there's a special and essential ingredient that can't be replicated, e.g. soul, spirit, consciousness, free will, creativity, humour, …	Kind of, yes: we can **simulate** intelligence • Outwardly, systems may *behave as if* intelligent • But the way they achieve this behaviour (the internal process) doesn't qualify as true thinking	Yes, we can build **true human-like** intelligence	Yes, we can build true intelligences but they won't necessarily be like us AI = **alien intelligence**

- Where do you sit in this table? Or, do you have a different view?

What are the risks?

 "the most serious threat to the survival of the human race"	 "The development of full artificial intelligence could spell the end of the human race…It would take off on its own, and re-design itself at an ever increasing rate"

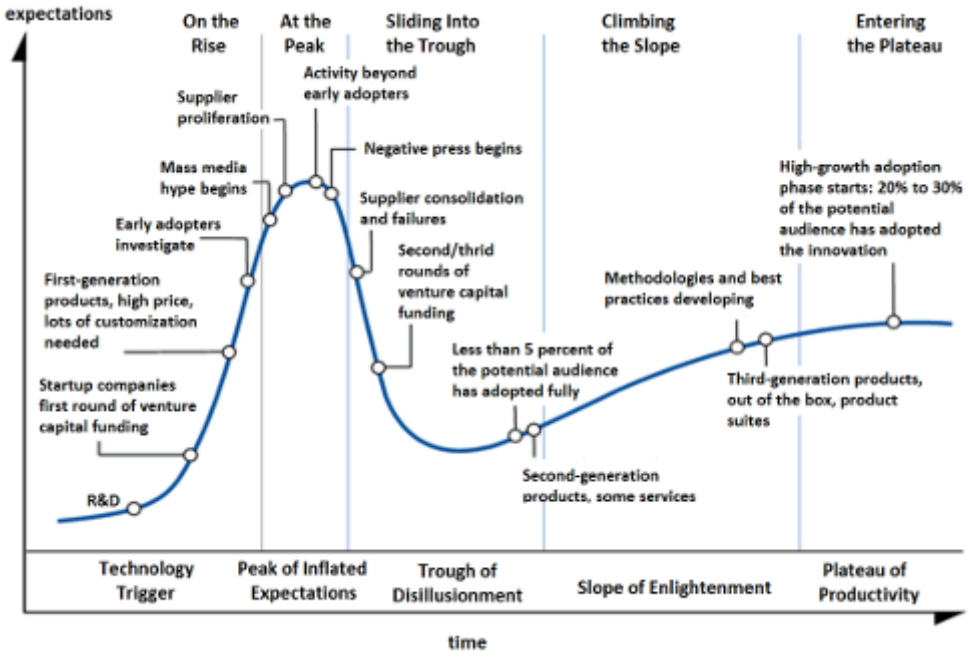
A dose of realism

"I don’t work on not turning AI evil today for the same reason I don't worry about the problem of overpopulation on the planet Mars."

Some advice for journalists writing about AI (http://togelius.blogspot.ie/2017/07/some-advice-for-journalists-writing.html)

"AI is a collection of methods … that can do something impressive, such as playing a game or drawing pictures of cats. However, you can safely assume that the same system cannot both play games and draw pictures of cats. … [Journalists can] make it seem like there are machines with general intelligence out there. There are not."

Is AI currently overhyped?



So what are the real risks?

- We can analyse the dangers in terms of:
  - malevolent goals, and
  - destructive methods for achieving benevolent or malevolent goals (e.g. methods that have unacceptable externalities)
- In the near to medium term, we should worry much less about super-intelligences that develop their own malevolent goals (e.g. to kill, enslave or displace us)
- Rather, we should worry about governments, corporations and individuals intentionally or unintentionally building AI systems that try to achieve their goals using destructive methods
  E.g.
  - so-called 'collateral damage' from autonomous weapons
  - displacement of employment
  - reduction in the economic, military or social value of some classes of human beings
  - invasions of privacy
  - 'filter bubbles' or 'echo chambers'
  - adoption or perpetuation of bias and prejudice
  - data-intensive AI restricted to a handful of hardware-rich and data-rich corporations

But what is intelligence?

- Some people define it in terms of outward behaviours, e.g. the *Turing Test*
- Some people define it in terms of various skills, e.g. reasoning, planning, learning, … or playing chess, composing poetry, …
- Think about my far-from-perfect definition before the next lecture:
  A system's degree of intelligence is defined in terms of its capacity to act autonomously and rationally when faced with disorder, uncertainty, imprecision and intractability.