tsg insights

September 2021

An overview of artificial intelligence (AI) and machine learning (ML)

This whitepaper is the first in a series which act as companion pieces:

- An overview of artificial intelligence (AI) and machine learning (ML).
- Introduction to the testing of AI and ML.
- Testing of AI (artificial intelligence) and ML (machine learning) – supervised learning.
- Testing of AI (artificial intelligence) and ML (machine learning) – unsupervised learning.
- Testing of AI (artificial intelligence) and ML (machine learning) – reinforcement learning.
- AI and machine learning – algorithmic bias – the cruel mirror AI and ML reflects back at us.

What's the difference between AI and ML?

The two terms are often used interchangeably, but actually there's a crucial difference:

- Artificial intelligence (AI) is the capacity of machines to carry out tasks in a way people would consider to be intelligent. Essentially, they are programmed to think and act like humans.
- Machine learning (ML) is basically a sub-set of AI. It's the idea that machines should be given access to data and allowed to learn for themselves.

A brief history of AI and ML

Artificial intelligence as we understand it today originated in the mid-20th century.

- The concept of the programmable computer originated during World War Two. The world's first electronic digital programmable computer was the Colossus, first used at Bletchley Park in 1944 to help decipher German military codes.
- Starting in 1956, a workshop at Dartmouth College coined the phrase "Artificial Intelligence" with initial efforts focusing on basic challenges such as playing draughts. Research then moved to understanding natural human language. This was heavily funded by the US Department of Defence and the UK Ministry of Defence; their interest at that time was in translating Russian and other Eastern European language documents into English.
- In 1974, a scathing review by Sir James Lighthill highlighted the limitations of AI when it comes to understanding multiple meanings of different words. Funding ceased and the first "AI winter" began.
- AI experienced a resurgence of interest and funding in the early 1980s due to the commercial success of "expert systems". However, changing computer markets and loss of funding led to the second AI winter.

- We are currently experiencing a third “AI spring” due to several factors. Computing power has grown dramatically, as vastly more data can be stored on computer chips and computer parts have become more standardised. For AI and ML, specialist computing devices such as Google’s TPUs (tensor processing units) allowing significantly more processing power and parallel processing capabilities than were available even ten years ago.
  multiple meanings of different words. Funding ceased and the first “AI winter” began.
- AI experienced a resurgence of interest and funding in the early 1980s due to the commercial success of “expert systems”. However, changing computer markets and loss of funding led to the second AI winter.

However, previous AI winters followed a pattern: hype about the capabilities of AI and ML, followed by significant investment, then initially steep but plateauing progress which resulted in the reduction of interest (and funding). So it’s possible that a third AI winter may yet occur.

The three different types of machine learning

There are broadly three different approaches to training machines to learn, each with its advantages and disadvantages.

1. **Supervised learning**
   This is based on existing examples, with expected results. It uses statistical methods such as regression and classification.

   Advantages:
   - In general, the fastest and most precise training.

   Disadvantages:
   - Backwards-looking training.
   - Sufficient data with good coverage needs to be available in the first place.
   - Large data sets need to be boiled by down using mathematical “normalisation”.
   - There’s a strong risk of learned data bias, which can result in skewed outcomes, low accuracy and errors. This bias can become normalised over time.

2. **Unsupervised learning**
   This involves simply exposing the machine to the data and letting it draw its own conclusions, based on statistical methods such as clustering and modelling.

   Advantages:
   - While being fundamentally backwards-looking, it can result in new and unexpected conclusions in the data.
   - It requires less data input.

   Disadvantages:
   - It generally results in less precise and slower training compared with supervised learning.
   - There’s a risk of learned data bias, which again can become normalised over time.

3. **Reinforcement learning**
   This is about exposure to a simulator and letting the machine experiment there. There are points or rewards for progress towards expected behaviour.

   Advantages:
   - Can come up with unexpected solutions.

   Disadvantages:
   - Training time can be lengthy and expensive.
   - Learning is only as good as the simulator permits.

AI and ML strengths

TIn limited, defined systems like playing strategy board games such as chess or video games such as StarCraft II, AI and ML are outperforming humans by a large margin.

They are also superior in other defined areas. These include: image and face recognition, video surveillance and tracking, natural speech recognition and text-to-speech generation, plus first-line support chatbots.
here are broadly three different approaches to training machines to learn, each with its advantages and disadvantages.

AI and ML challenges

There are many everyday challenges to the adoption of effective AI and ML, which we will look at in more detail in other papers.

These include:

- Differences in cultures, locations and language
- Qualification and definition issues of words or word families
- Word-sense ambiguity (for instance, the word 'set' has 430 different meanings in English)
- Learning choices (for example, less data input can speed up learning, but more may improve accuracy)

In addition, there are several broader issues to consider, outlined below.

"Cruel mirror " bias

Training data can be unintentionally biased, with the results only noticed when the ML results come in. In that respect, it can be said to be a "cruel mirror", reflecting and indeed exaggerating the effects of the faulty original data. Bad decision making is not just a risk, but a guarantee in these circumstances.

For example:

- Amazon's recruitment system started in 2014 with the declared aim of eliminating recruitment bias. By 2015, Amazon realised that the AI was heavily biased against women. Multiple interventions failed and in 2018 Amazon gave up and reverted to a human-resourced solution.

- Microsoft chatbots on Twitter were introduced but later decommissioned – the first, Tay, was racist and inflammatory, while its successor Zo was unbearably politically correct.

Adjustment to paradigm shifts

Since the turn of the century, there have been various developments leading to fundamental changes in assumptions. These paradigm shifts are not just related to major events such as 9/11, Brexit and the COVID-19 pandemic, but issues such as Me Too and Black Lives Matter.

Human staff members can be called up for training and new instructions, but there is no known precedent where AI or ML has been radically reshaped, virtually overnight.

Taking the issues with bias (stemming from historical data bias) into consideration, retraining AI or ML is highly challenging. In the Amazon recruitment example above, Amazon tried for four years before giving up.

Algorithmic accountability (algorithmic decision systems) legislation

Law in this area is still evolving, but has brought social media giants under the spotlight already. If every AI and ML operator is required to explain and justify any algorithmic decision, will they be able to do so?

Especially when (deep) neural networks are involved, this could in practice be very difficult or even impossible to achieve. Under ADS legislation, it could expose the operator to legal and regulatory action.

TSG provides expert guidance on AI and ML, as well as assurance and testing services. We make change happen, safely and predictably. If you have any question about issues covered in this whitepaper or would like to know more about how we can help you, please contact us now.

Call: +44 (0) 207 469 1500 Email: info@tsgconsulting.co.uk www.tsgconsulting.co.uk