How to be an Astute User of Artificial Intelligence

Fundamentals for Decision-Makers

There are many situations in which being able to process very large quantities of images, text or other data is extremely useful. Examples range from checking legal documents to improving the accuracy of medical diagnoses. AI can potentially provide us with very powerful tools for performing these tasks much more rapidly and precisely than humans.

This raises the question as to *when, how or even whether AI should be used*. The aim of this short guide is to provide potential users of AI working in policymaking with a plain English summary of the fundamentals and a set of questions to ask when thinking about applying AI to different problems.

What is AI?

AI - Artificial Intelligence is an umbrella term for a diverse set of computer programs (algorithms) that apply statistical analysis to process text, images or other data.

e.g. by recording the purchases of particular products in a supermarket, it is possible to identify groups of items frequently purchased together and so design advertising strategies to automatically offer individual customers products they are more likely to buy.

Image credit: Yutong Liu & Kingston School of Art / Better Images of AI / Talking to AI / CC-BY 4.0

How does it work?

Underpinning AI algorithms is a set of mathematical equations for calculating different characteristics of data. AI algorithms perform two basic tasks:

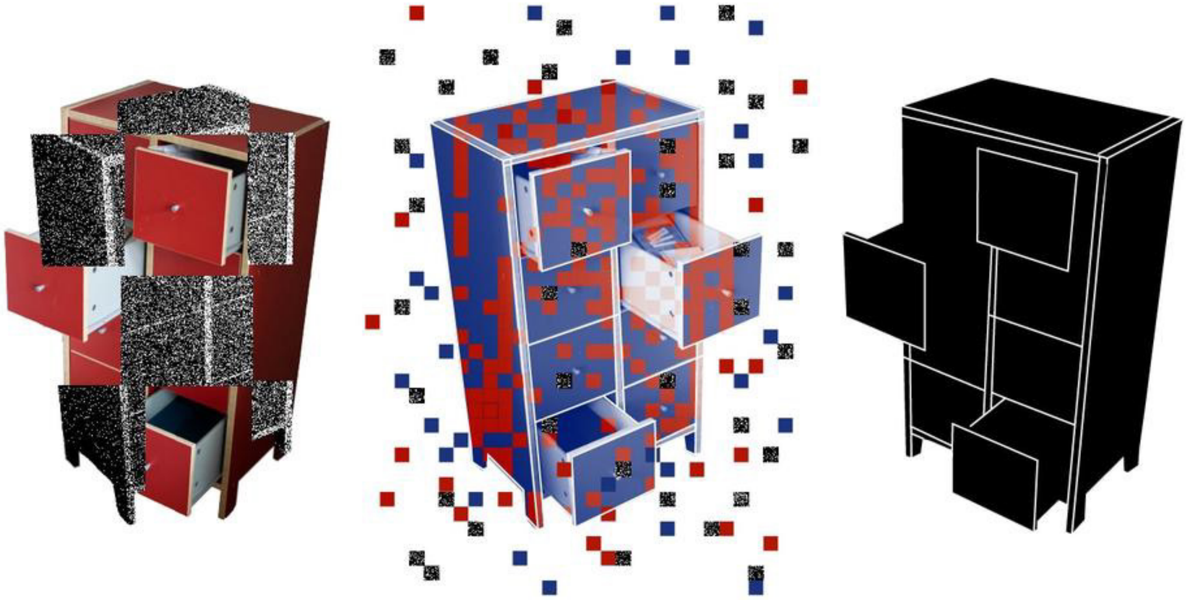

Image credit: Anton Grabolle / Better Images of AI / Classification Cupboard / CC-BY 4.0

Classifying data
Identifying features that
allow data to be grouped.

Predicting trends
Using data to identify patterns
and thus predict future behaviour.

By predicting a range of possible future behaviours and the probability of each occurring, some types of algorithm (known as generative AI) can then use this information to create new data e.g. text or images. To create a useful piece of software for a particular task, a set of data (the “training data”) is used to define the relevant features or patterns. The trained algorithm can then be used as an **AI tool** to process new data.

e.g. the autocomplete function in email and messaging apps is based on analysis of large quantities of text (the training data) to identify groups of words that commonly appear together. This allows the app to predict what you’re likely to type next. Large language models (LLMs) such as ChatGPT work on a similar principle, just on a much larger scale, allowing them to compose longer and more complex pieces.

Algorithms may be designed for specific tasks from the outset or can be more general purpose (foundation models) and then “fine-tuned” for specific tasks by the user, e.g. LLMs can be trained using text from an extremely wide range of sources and then tuned using a much narrower set of documents in a particular field. It is extremely important, however, that both the **algorithm** and the **training data** are suitable for the specific task.

When are AI tools useful?

AI tools work well when there is a very large number of similar images/text or other types of data to analyse ***and*** a limited number of features to pick out.

In the supermarket advertising example, the supermarket will hold a very large database of customer purchases and have a finite number of products that it offers. Thus, it is relatively simple to accurately predict shopping trends, especially if customers provide additional information through a loyalty card scheme, e.g. age, address, number of children they have etc.

Another important example of where AI tools can be extremely powerful is in the diagnosis of cancer. Pathology laboratories receive 1000s of images of potentially cancerous tumours daily. These images are reviewed by experts who look for features that indicate cancer. Algorithms can be designed to identify those same features and perform the analysis automatically. This greatly increases the speed of the process, reduces the likelihood of a cancerous tumour being missed because a human reviewer is tired or distracted, and frees up the expert for other work that cannot be automated.

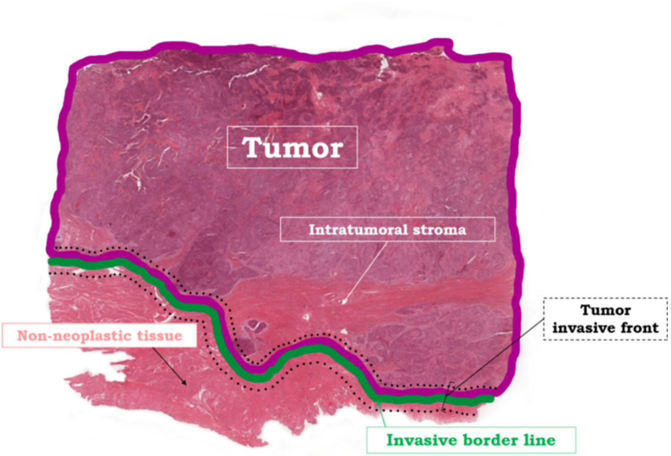

Image credit: Dr. Aleksandra Żuraw, Digital Pathology Place

Other similar examples include drafting and checking of routine legal documents or quality checking of images for publication.

There are some very important points to note about the cancer diagnosis example.

Sufficient training data — Pathology labs have very large numbers of historical images that have already been classified by experts and can be used to train the algorithm to correctly identify those features that indicate cancers.

Constrained task — The number of possible features that can be identified is relatively small.

Well-defined question — There is a clear question to be answered, i.e. is a tumour cancerous or not?

Representative data — It is reasonable to expect that new tumour images are going to be similar to historical tumour images, i.e. ***historical trends will be good predictors of future ones.***

Limited autonomy — There will still be oversight from an experienced pathologist before any clinical decisions are made.

When is it potentially problematic to apply AI tools?

Unfortunately, there are many situations in which:

- much smaller quantities of training data are available;
- there is a much larger number of variables;
- there are multiple questions to be answered;
- historical data are not useful.

e.g. decisions about maintaining or withdrawing life support from critically ill patients is a much more complex decision-making process, involving multiple factors and extremely sensitive ethical questions. There will be a far smaller number of prior cases than for a cancer diagnosis and no guarantee that those cases represent appropriate precedents. Similar concerns would arise in the case of legal decisions involving child custody or granting of probation.

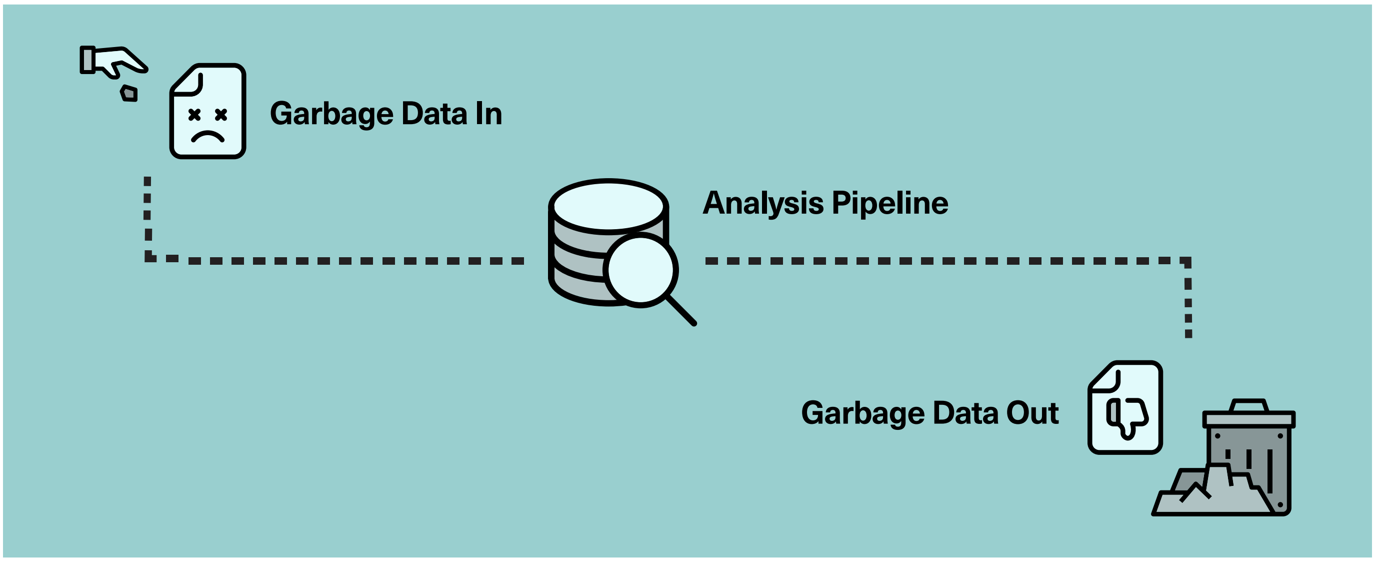

Image credit: Rendition created with credit to William F. Kezele.

What are the risks of applying AI tools to the wrong problem?

Very simply, if we use inappropriate data and/or algorithms as the basis for decision-making, we are likely to make bad decisions.

Both the quantity **AND** the quality of the data used to train an algorithm are critically important.

e.g. sticking with the example of the supermarket, if data were only available for an area with a majority of people over 65 years old, any predictions probably wouldn't be very useful in an area with a majority of young families. The data would be biased, i.e. not be representative of the population it is being applied to.

What are the risks of applying AI tools to the wrong problem? (Cont.)

In other applications, the risks associated with data bias can be much more serious than unwanted product recommendations, **e.g.** if medical records were only available from patients from one age group or gender but applied to another, it increases the risk of inappropriate clinical decisions being made. The recent controversy over the UK Biobank data has highlighted the risks associated with data bias, in this case due to the fact that volunteers donating tissue to the bank for scientific studies were not representative of the overall UK population.

For many problems it is very difficult, if not impossible, to get sufficient data of appropriate quality, **e.g.** a potentially very attractive application of AI tools is for summarising long documents and/or analysing sentiment; but if the training data are drawn from documents in a different field and/or a different language, they are unlikely to provide reliable results.

Image credit: Rens Dimmendaal & Johann Siemens / Better Images of AI / Decision Tree reversed / CC-BY 4.0

There are several other problems that also need to be considered:

- The legality of using some training data. In conventional research, there are fees for using copyrighted material and/or requirements for crediting its authors. Using data for algorithm training does not remove these requirements.

- There is a risk of malicious “data poisoning” to deliberately introduce bias in order to influence results.

- Algorithms do not necessarily reduce the burden on humans. In many cases the human resource required to create a suitable dataset is as substantial as that required for analysis.

- The energy consumption involved in training some algorithms can be enormous. The benefits gained from AI need to be weighed against the corresponding carbon footprint.

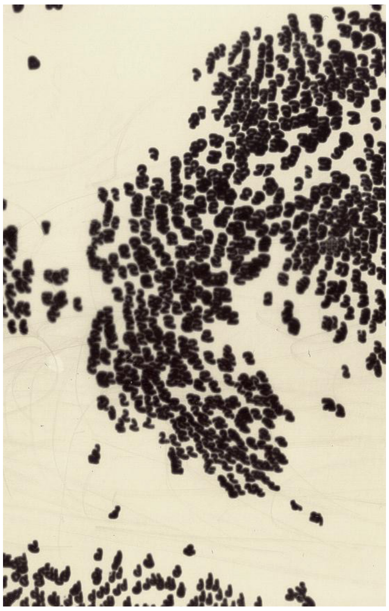

Image credit: Philipp Schmitt / Better Images of AI / Data flock (digits) / CC-BY 4.0

Summary & Suggested Actions

AI has huge potential to improve the speed and accuracy with which we handle data in numerous applications, but there are many possible pitfalls. We need decision-makers, algorithm authors and data scientists to collaborate in order to ensure AI-driven analysis is robust, unbiased and transparent.

The following is a checklist of questions to ask when considering if AI is going to be useful for a particular task or problem:

1. Am I looking at a problem with an answer that is likely to be predictable from previous data?
2. For what purpose has the AI tool been developed and is it a good fit to the problem I want to solve?
3. What features of the data is the AI tool I want to use actually looking at and are they the ones in which I am interested?
4. What data have been used to train the tool and are those appropriate in terms of quantity and quality for the problem I'm looking at? *Here it is strongly advisable to seek support from a data scientist.*
5. Has the AI tool been validated rigorously and appropriately using relevant data? Have the rates of false positives/negatives been quantified?
6. What are the consequences of an incorrect prediction and are those acceptable?
7. Is the AI tool adequately maintained to mitigate the risk of data bias or poisoning?
8. Is the AI tool compliant with a recognised regulatory framework ? e.g. https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_1683

If the answer to any of these questions is "no", or if the designers of the AI tool you are considering are unable to provide satisfactory answers, then it is unlikely the tool will provide a reliable result.

Contributors

Joe Bourne
Muffy Calder
Tom Coates
Andrew Duncan
Alison Etheridge
Nick Jennings
David Leslie
Eleanor Stride
Paul Taylor
Tim Watson
Mike Wilby

Further reading

https://fortune.com/2024/03/22/how-to-identify-ai-washing-products-services/

https://medium.com/@kalanabandaranayake/artificial-intelligence-is-it-too-good-to-be-true-e6d809b67622

https://vitalflux.com/wp-content/uploads/2020/12/machine_learning_mind_map_3.png

https://www.piie.com/blogs/realtime-economics/2024/ais-carbon-footprint-appears-likely-be-alarming

https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_1683

https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/guidelines-responsible-use-generative-ai-research-developed-european-research-area-forum-2024-03-20_en

https://www.iso.org/artificial-intelligence/responsible-ai-ethics

https://ai.google/responsibility/responsible-ai-practices/

https://senseaboutscience.org/responsible-handover-of-ai/