SMU Classification: Restricted

Emerging Technologies and Law

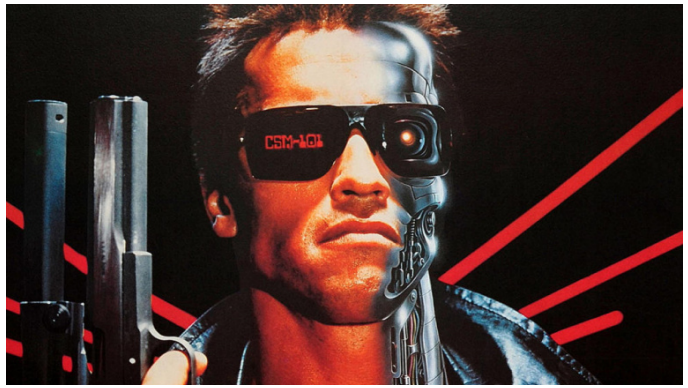
Artificial Intelligence

SMU Classification: Restricted

In this Video…

- What is AI really?
- How “machine learning” works (but too briefly)
- How to think about AI systems in law

What it sounds like…



Artificial intelligence 'judge' developed by UCL computer scientists

Software program can weigh up legal evidence and moral questions of right and wrong to predict the outcome of trials

▲ The algorithm examined English language data sets for 584 cases relating to torture and degrading treatment, fair trials and privacy. Photograph: Cultura/Rex/Shutterstock

Home / Innovation / AI & Robotics

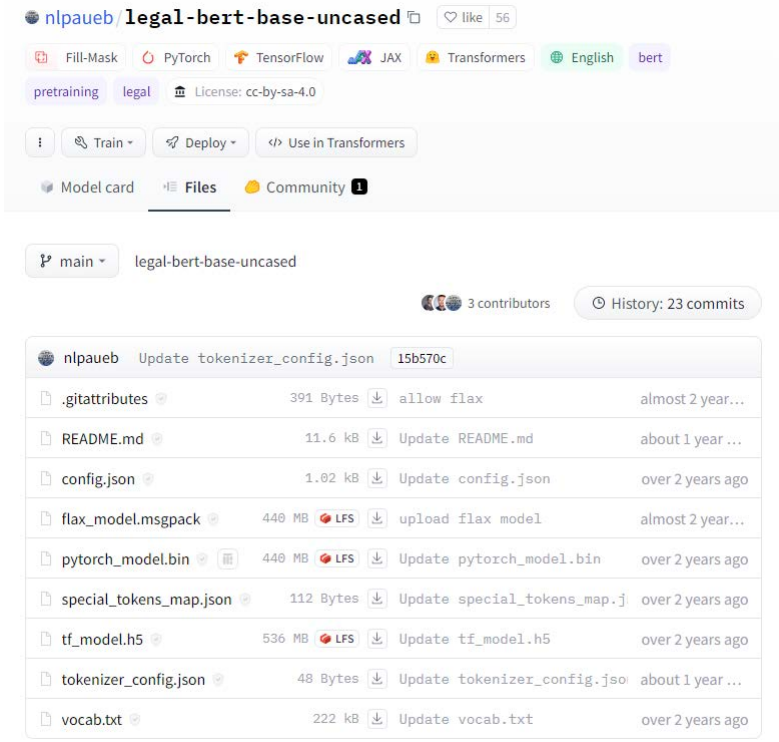
Evil AI: These are the 20 most dangerous crimes that artificial intelligence will create

A new report tells us which criminal applications of AI we should really worry about.

Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian

Tempe police said car was in autonomous mode at the time of the crash and that the vehicle hit a woman who later died at a hospital

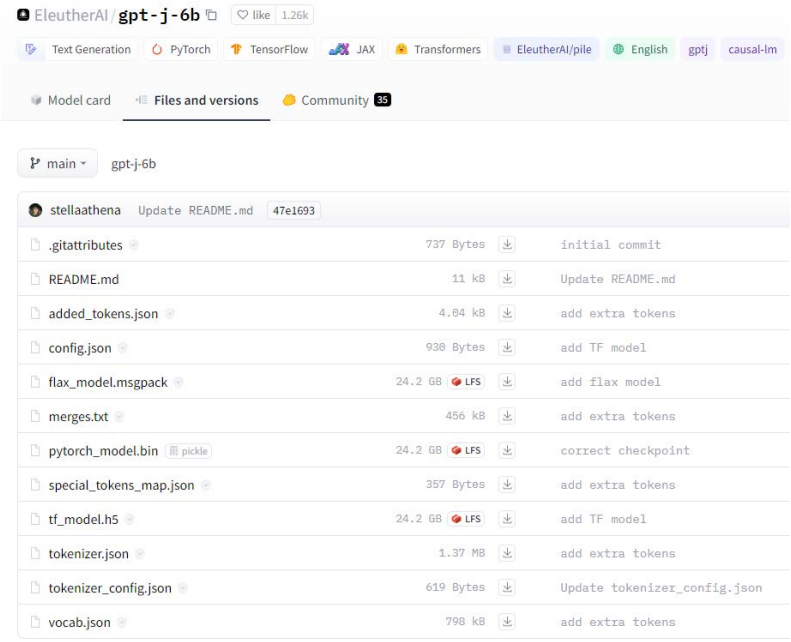
What It Actually Is



A 110M network ("LEGALBERT") trained on a 12gb corpus of legal documents by Chalkidis et al, 2019

```
>>> from sklearn.svm import LinearSVC
>>> from sklearn.pipeline import make_pipeline
>>> from sklearn.preprocessing import StandardScaler
>>> from sklearn.datasets import make_classification
>>> X, y = make_classification(n_features=4, random_state=0)
>>> clf = make_pipeline(StandardScaler(),
...                     LinearSVC(random_state=0, tol=1e-5))
>>> clf.fit(X, y)
Pipeline(steps=[('standardscaler', StandardScaler()),
                ('linearsvc', LinearSVC(random_state=0, tol=1e-05))])
```

The **Support Vector Machine** (Vapnik & Chervonenkis, 1963) in Python library code.
https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html#sklearn.svm.LinearSVC

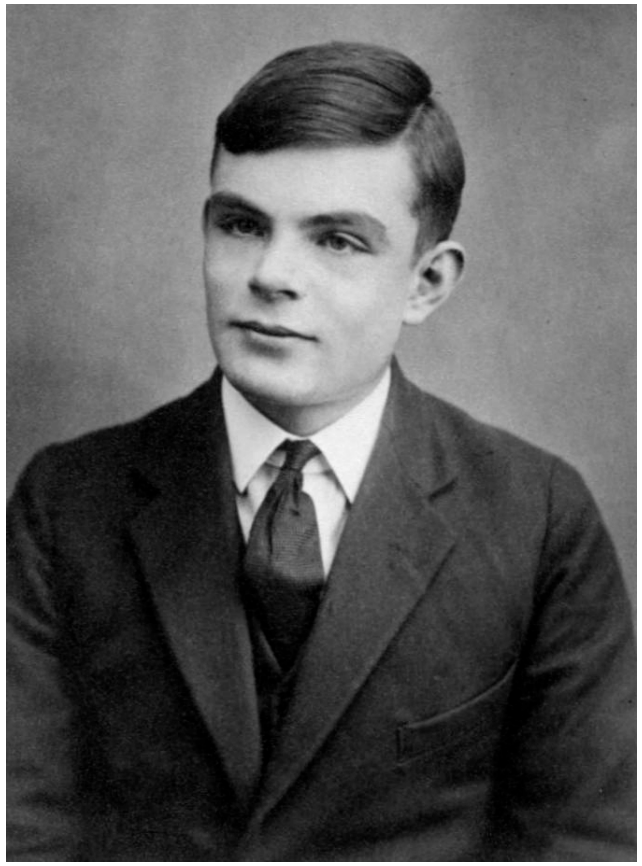


An open-sourced adaptation of GPT2
https://huggingface.co/EleutherAI/gpt-j-6b/tree/main

Modern AI Theory

- Turing, 1950: "AI is about building machines that think"
- But "machine" and "think" are undefinable, not worth asking this question

- Instead: imitation game
- I know it when I (can't) see it
- **So what if AI passes the Turing test?**



The Textbook Definition(s) of AI

Thinking Humanly	**Thinking Rationally**
"The exciting new effort to make computers think . . . *machines with minds*, in the full and literal sense." (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . ." (Bellman, 1978)	"The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)
Acting Humanly	**Acting Rationally**
"The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)	"Computational Intelligence is the study of the design of intelligent agents." (Poole *et al.*, 1998) "AI . . . is concerned with intelligent behavior in artifacts." (Nilsson, 1998)

Figure 1 Some definitions of artificial intelligence, organized into four categories.

Russell and Norvig, Artificial Intelligence: A Modern Approach (2014)

SMU Classification: Restricted

Modern AI Systems

Rules-based or **"symbolic"** AI

Machine learning or **"statistical"** AI

SMU Classification: Restricted

EXPERT SYSTEMS IN LAW

Out of the Research Laboratory and into the Marketplace

DR. RICHARD E. SUSSKIND

Manager, Expert Systems Group, ERNST & WHINNEY, UK.

The major goal of workers in the field of expert systems in law can be summarised as follows: through the use of computer technology, to make scarce, human legal knowledge and expertise more widely available and easily accessible. This paper is directly concerned with that goal. It is divided into two parts. The first discusses a three year interdisciplinary research project in expert systems in law that involved the Law Faculty and the Programming Research Group of the University of Oxford. Some of the jurisprudential aspects of the project are discussed and the prototype system that was developed is described. The second part of the paper addresses various considerations that are crucial for bringing expert systems in law out of the research laboratory and into the marketplace. And in this latter connection, the author's latest work - with Ernst & Whinney - is introduced.

THE RESEARCH LABORATORY

The Oxford Project - Background

In 1983, a collaborative research project concerning expert systems in law was initiated at the University of Oxford. The project was

The Oxford system is discussed later in this paper. The second goal of the project was to examine the use of certain mathematical techniques for the formal specification or description of expert systems in general, and of expert systems in law in particular. (That work is extensively documented in Gold's doctoral thesis, 'Specification and Implementation of an Expert System in Law' (1987)). The third goal of the project was to make an inquiry into the process of building expert systems in law from the point of view of jurisprudence (legal theory). Although the value of jurisprudence has often been questioned by practitioners, academics, and students alike, one branch of the discipline - analytical jurisprudence - is of the utmost relevance to expert systems in law. For that part of legal theory seeks to promote a theoretical, systematic, and general understanding of such issues as the nature of legal reasoning, the structure of legal rules and legal systems, the interpretation of legislation and case law, and the role of logic in the law. And each of these topics is central to building expert systems that reason with substantive law. The following section of this paper discusses some of the jurisprudential aspects of the Oxford project.

Richard Susskind, Proceedings of the 1st Intl Conf AI & Law, 1987

The program holds **knowledge of the Divorce (Scotland) Act 1976, of relevant judicial precedents, and also contains some legal heuristics**. It is designed, once the facts have been elicited from the user, to **inform him whether there are grounds for divorce...**

The knowledge of the legal domain was, in the first instance, culled largely from primary and secondary written legal sources, that is, from legislation, case law, and standard legal textbooks. Initial tentative representations of the field of law were refined in light of consultation with experienced legal practitioners.

The knowledge is represented in the knowledge base as a network of interrelated rules that can be altered with little fuss: it is a flexible, rule-based system. The corpus of knowledge can be regarded as a set of possibilities - a search space - which the system must explore with the guidance and direction of the user. At the beginning of the interaction, and periodically thereafter, **the user is required to enter some basic data, such as the names of parties, relevant dates, and so forth. However, the principal ways in which the user apprises the system of the facts of a case are through "yes", "no", or "don't know" responses to questions asked of him and through selections from menus.**

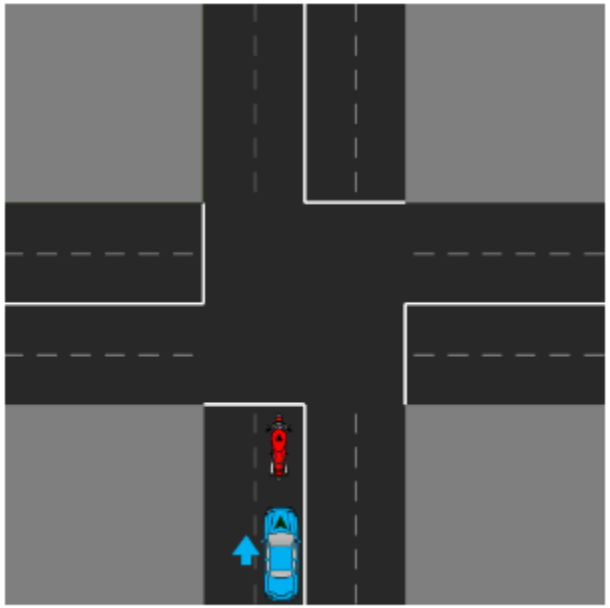
The order in which the system explores possible solutions to problems is conditioned by the system's inference engine, which can best be described as facilitating a "user-controlled backward-chaining" reasoning mechanism. Very broadly speaking, an expert system that backward-chains starts its reasoning process at the conclusion it is trying to reach, and moves backwards through its body of rules in search of premises that will justify that conclusion.

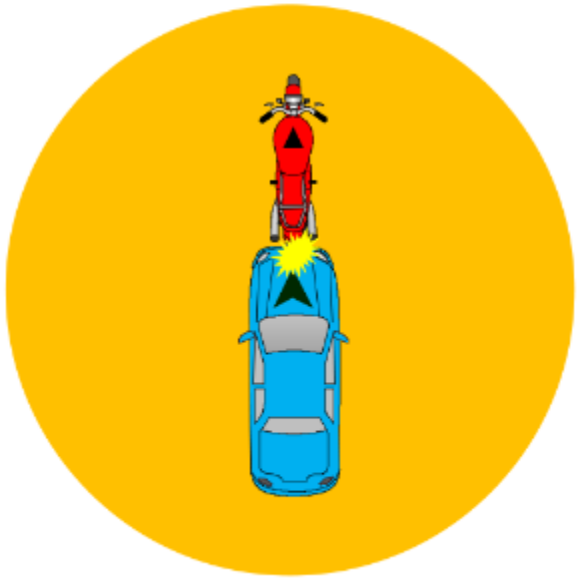
Based on information that you have provided,
The other party would be wholly at fault (100%) for the accident.

0% Your fault

100% Other party's fault

ILLUSTRATION: ACCIDENT SCENARIO





REASON(S)

Party who reversed is responsible for the accident

Problems with Rule-based AI

- **Costly to specify every rule**
- How to build AI for chess?
- For Go?
- For law?

- **Strategy:** invest resources to gather **examples rather than experts**; code the algorithm by example

"Machine Learning"

- **Branch** of AI research that uses **statistical methods** to get computers to perform tasks without explicit instructions (Arthur Samuel)
- **Consider:** How would you design a system that identifies email spam?
- **"Explicit instructions":** rule-based AI system
- **No explicit instructions:** statistical/ML AI system

Congrats! You've won our recent lottery…

I am writing bcos king in my country died n left large inheritnce..

Your vaccination appointment has been successfully scheduled…

Dear Friend, it's been a while. I need your help. Can you buy me this Apple Pay card?

We are pleased to inform you that your application for…

URGENT: YOUR MORTGAGE IS OVERDUE

ML-based Systems

Email	Lottery	Inheritance	Typos	Source	Truth Label
1	1	0	0	Unfamiliar	Spam
2	0	0	2	Unfamiliar	Spam
3	0	0	0	Familiar	Not Spam
4	0	0	0	Familiar	Spam
5	0	0	0	Familiar	Not Spam
6	0	0	0	Familiar	Not Spam

Congrats! You've won our recent lottery… 1

I am writing bcos king in my country died n left large inheritnce.. 2

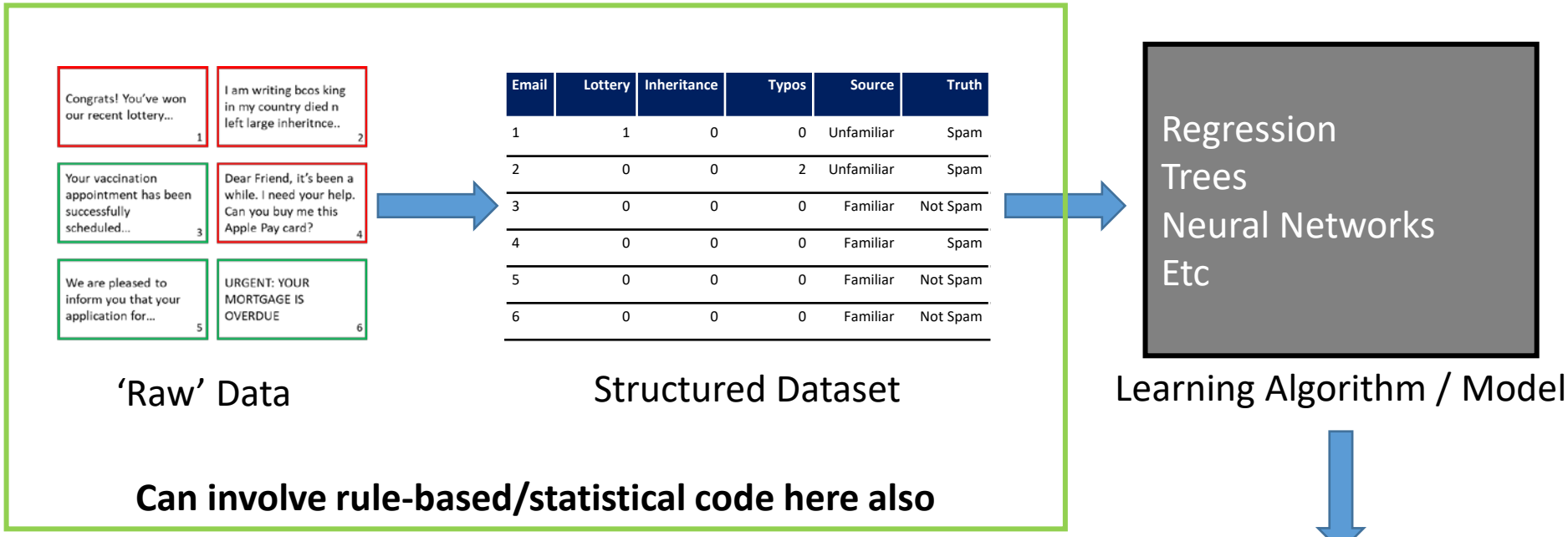
Your vaccination appointment has been successfully scheduled… 3

Dear Friend, it's been a while. I need your help. Can you buy me this Apple Pay card? 4

We are pleased to inform you that your application for… 5

URGENT: YOUR MORTGAGE IS OVERDUE 6

ML-based Systems

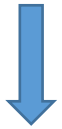


$$P(spam) = 0.2 \times Lottery + 0.00 \times Inheritance + 0.05 \times typos + 0.2 \times source$$

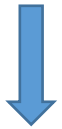
Classification Algorithm / **Trained Model** /"Hypothesis"

ML-based Systems

Hi Jerrold, I need some legal help with lottery winnings that my parents left me as part of my inheritance…



$$P(spam) = 0.2 \times Lottery + 0 \times Inheritance + 0.05 \times typos + 0.2 \times source$$



Score = 0.2

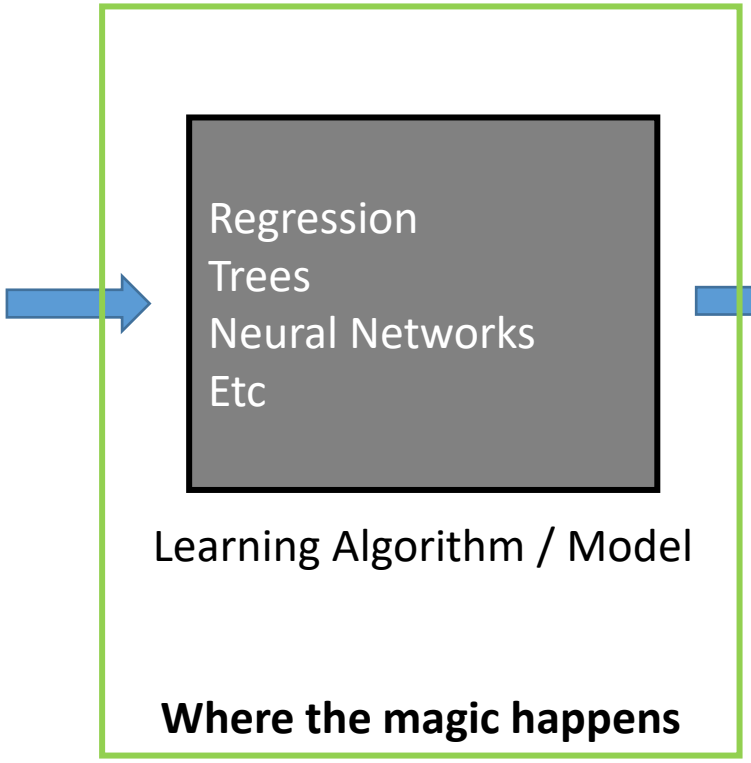
Prediction: **Not spam (assuming >0.5 cutoff)**

Notice: this simple 'AI' is very **explainable.**

How does the learning work?

Email	Lottery	Inheritance	Typos	Source	Truth
1	1	0	0	Unfamiliar	Spam
2	0	0	2	Unfamiliar	Spam
3	0	0	0	Familiar	Not Spam
4	0	0	0	Familiar	Spam
5	0	0	0	Familiar	Not Spam
6	0	0	0	Familiar	Not Spam

Data



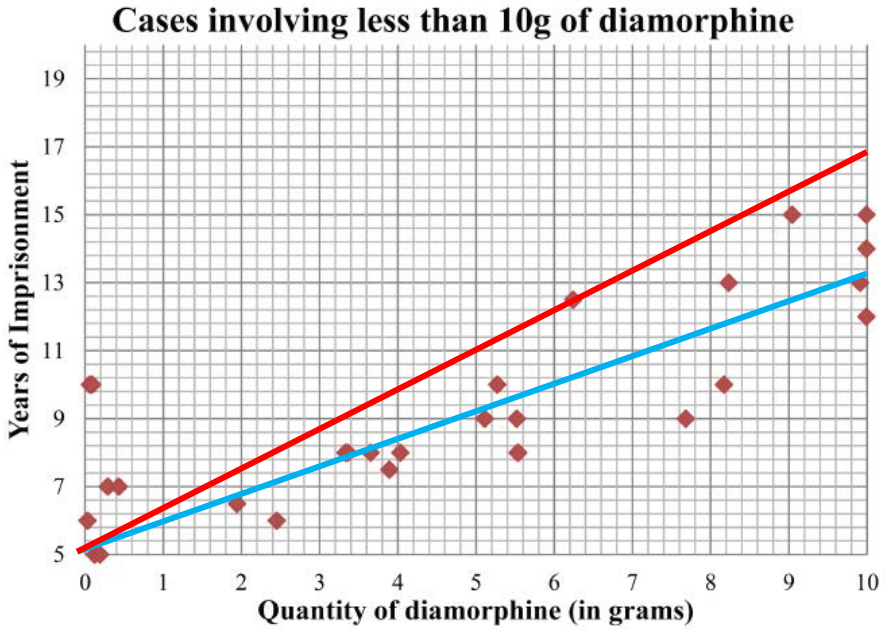
$$P(spam) = 0.1 \times Lottery + \cdots$$

Prediction Algorithm

"Learning" from the data

- On the right we have a scatterplot found in a Singapore judgment
- Can we use it to learn a prediction algorithm for drug trafficking sentences (Y) given quantity trafficked (X)?
- That means not just for a specific gram range, but for **all** gram ranges, even those gaps in between

- **Supervised learning tries to find a "best fit" pattern for E[Y|X]**
- Makes certain assumptions
- **Linear** regression is when we assume a line can fit this



Vasentha d/o Joseph v Public Prosecutor [2015] SGHC 197 at [28]

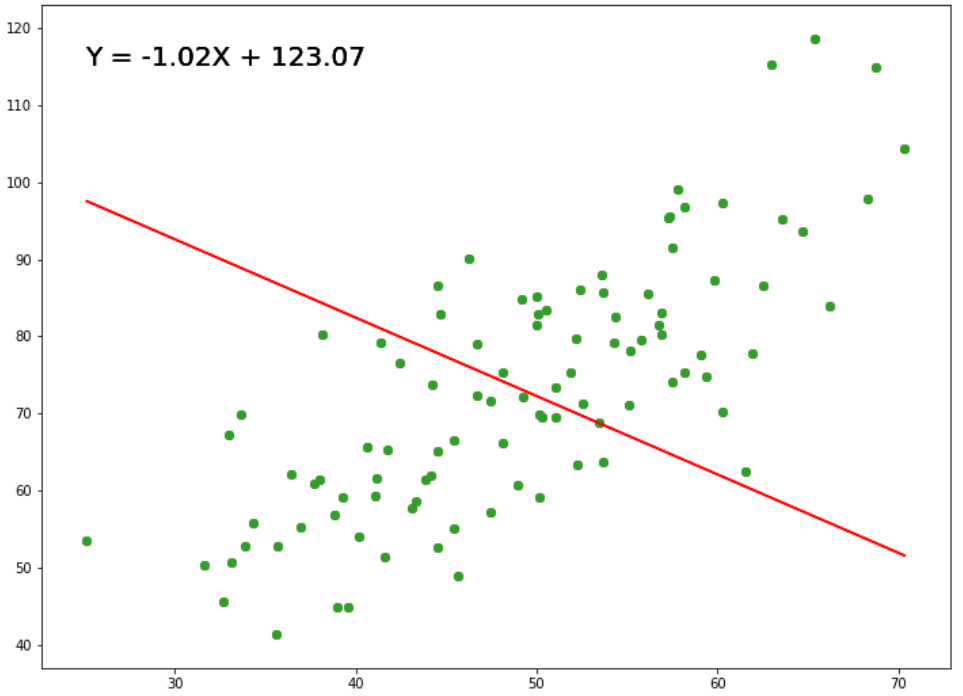
Machine Learning on Data

While humans can easily "eyeball" the line, we can't do it for more than 2 (3?) dimensions.

But math can:

- **Define a measure of "fit"**
- E.g. "least squares" distance between points on the line to data points
- Specify line equation – **determined by "coefficients"** (the changing numbers >>)
- See how well it fits
- See if changing the equation improves the fit
- If so, move to new equation
- Repeat until stable

Most ML algorithms operate on the same principle



https://dphi.tech/blog/tutorial-on-linear-regression-using-least-squares/

Regression Analysis

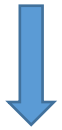
- The linear regression line implied by the scatterplot data was:
  - Sentence = 6.1597 + 0.6691 x Grams

- Therefore, E[Sentence|8.98 grams] = 6.1597 + 0.6691 x 8.98 = 12.168 years
- **This is the ML prediction *if* we use a single variable linear regression model**

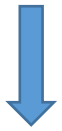
- "77 Since the present case involves 8.98g of diamorphine, the indicative starting point would be within the range of 10–13 years' imprisonment. As the quantity involved in this case is at the high end of the range, **I take as an indicative starting point a sentence of 12 years' and nine months' imprisonment.**"

Legal "Predictions"

Plaintiff is a 4-year-old girl. On 27 July, 2020, she visited the defendant's drive-thru with her mother…

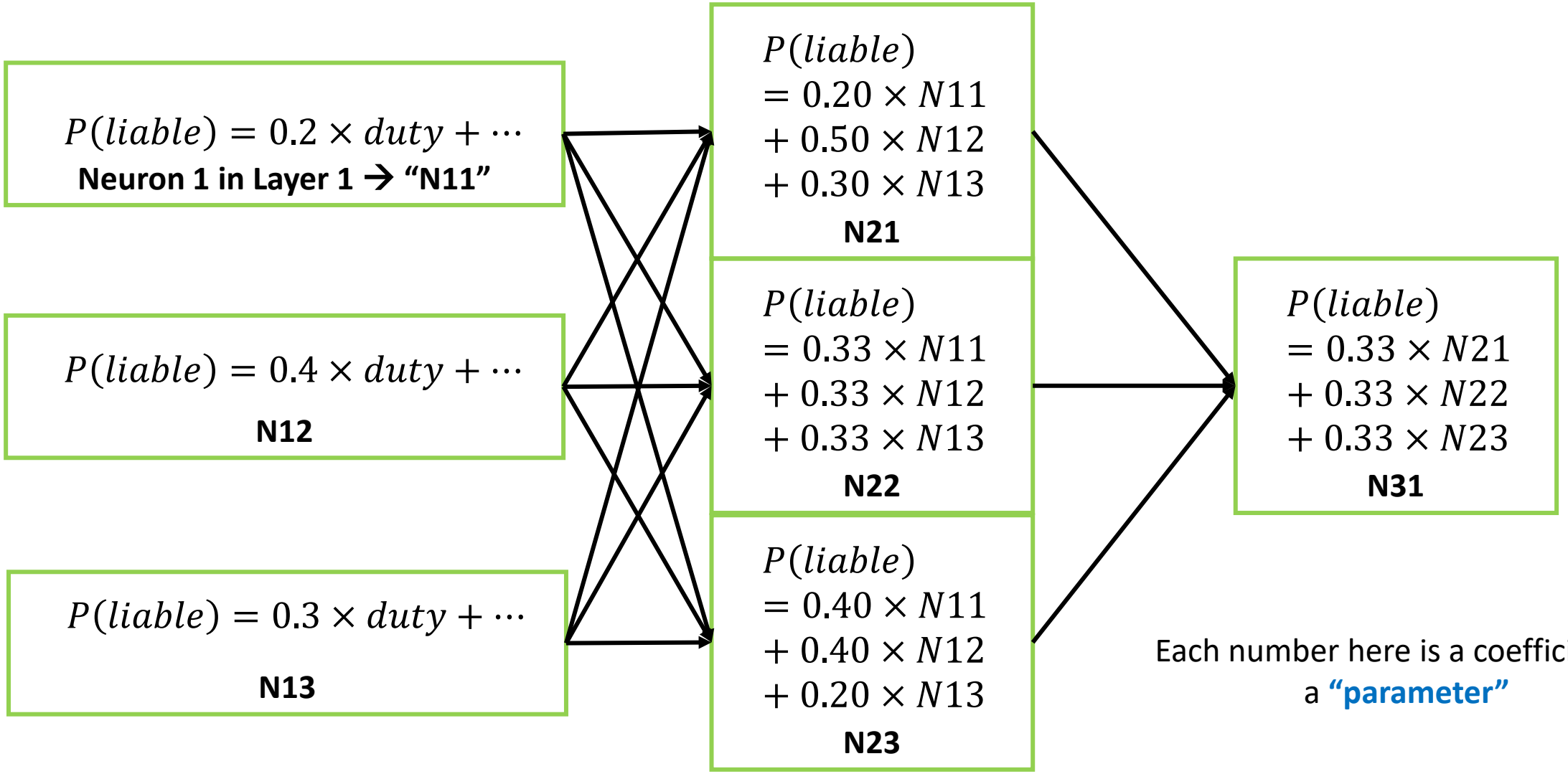


$$P(liable) = 0.2 \times duty + 0.2 \times burnt + 0.05 \times badly + \cdots$$



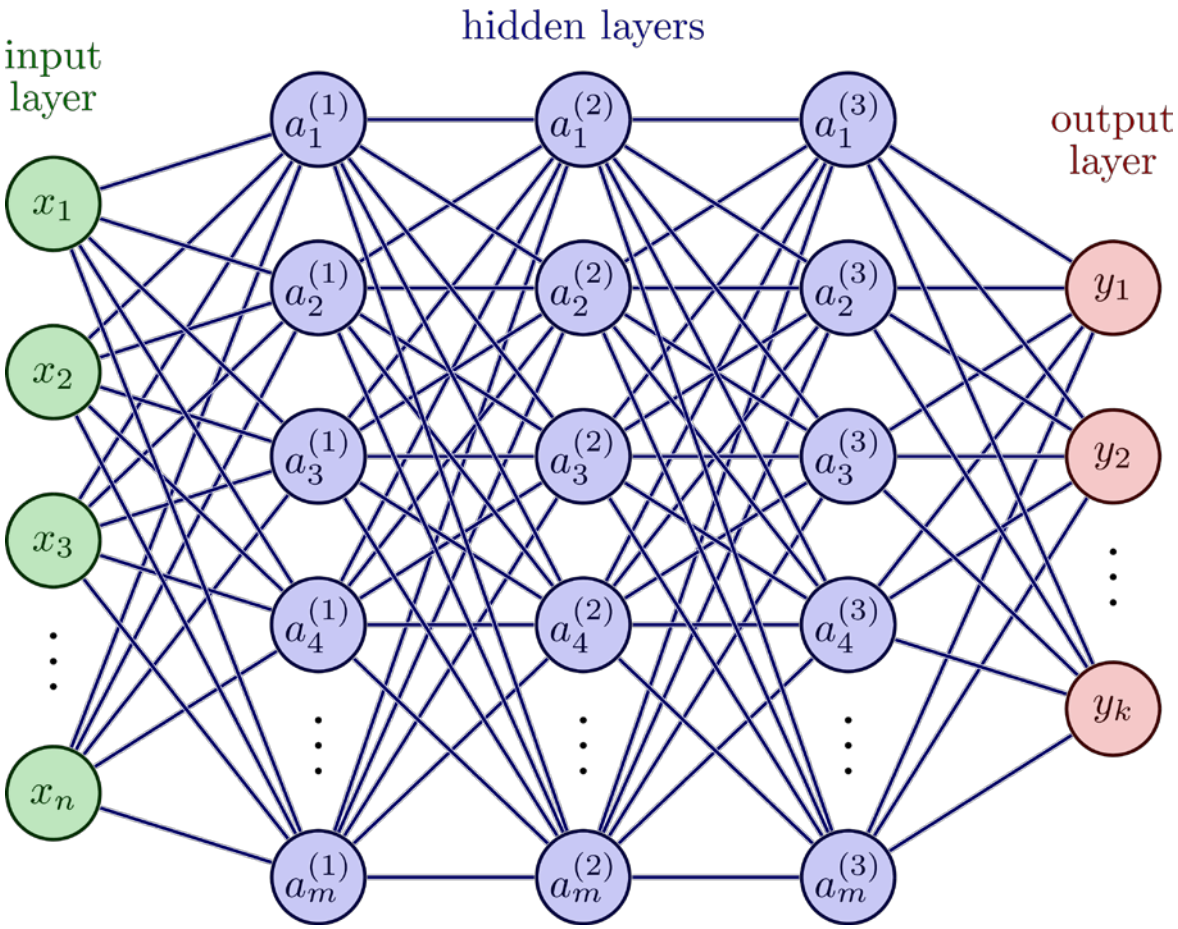
Score = 0.7. Prediction: **Liable.**

Neural Networks



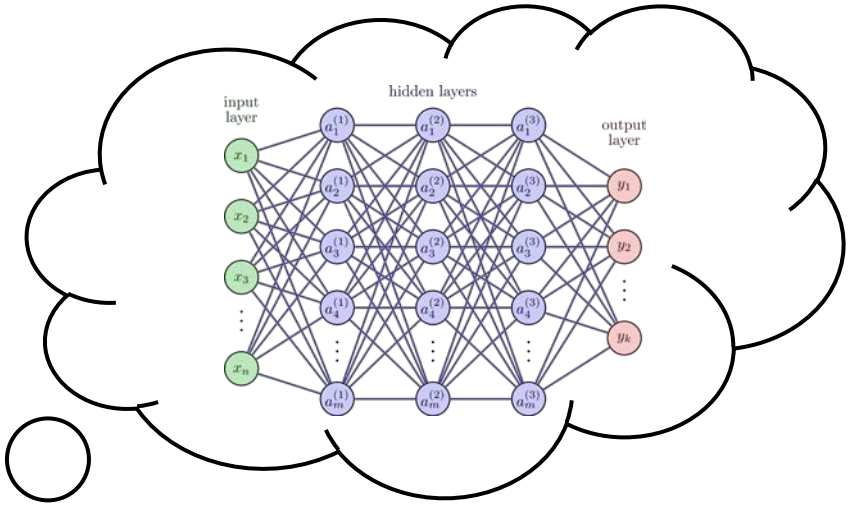
Each number here is a coefficient or a **"parameter"**

Large Neural Networks



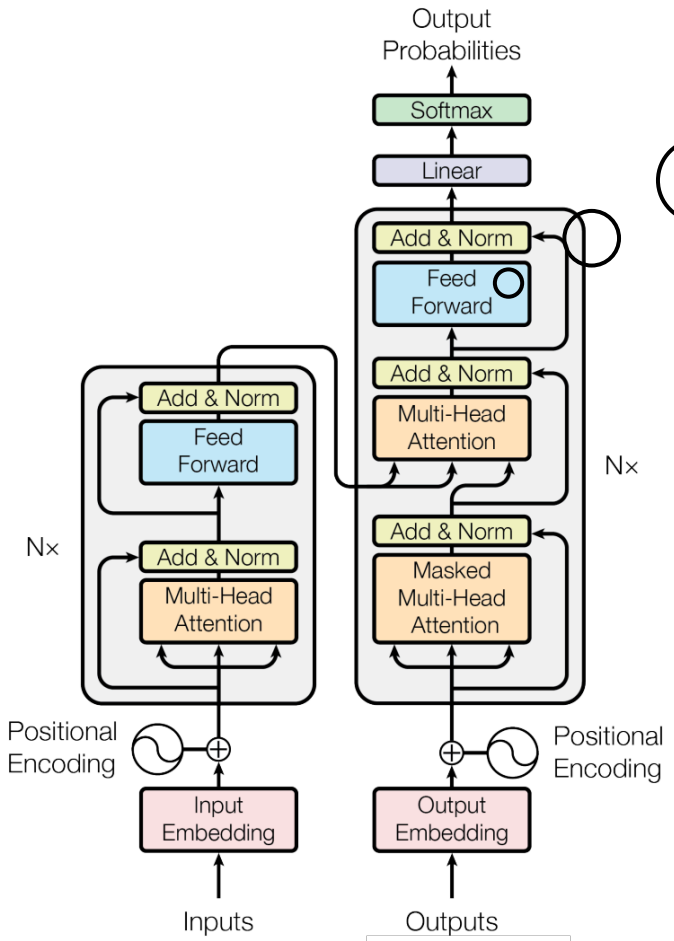
https://tikz.net/neural_networks/

Large Language Models



BERT

Encoder



GPT

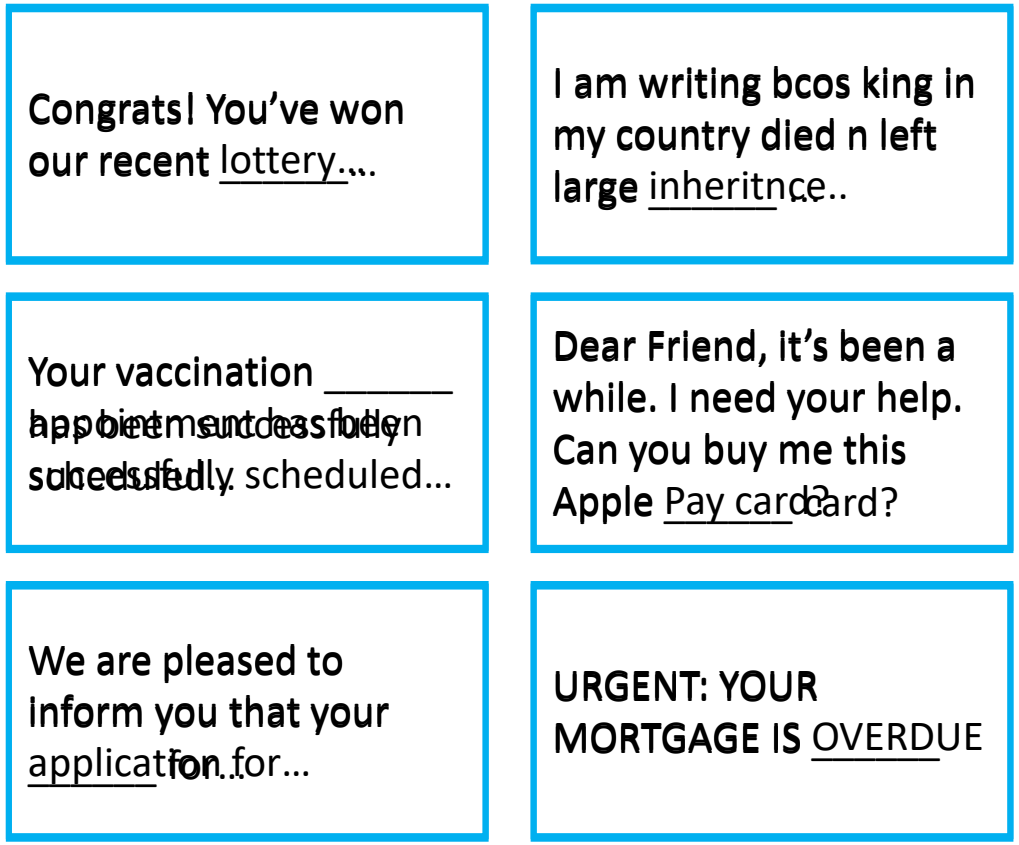
Decoder

https://heidloff.net/article/foundation-models-transformers-bert-and-gpt/

GPT3.5 has 175 **billion** parameters

Language Modelling and Generative AI

- Say we want to **write emails, not label them**
- **A "language model"** is created by taking a large corpus of text, deleting words, and training the model to **predict** missing words
- Generative AI is really a subset of predictive AI
- Given word 1, predict word 2
- This scales easily: Given sentence 1, predict sentence 2. Then now that you have predicted sentence 2, use that and predict sentence 3…
- If sentence 1 is a question, then sentence 2 is the predicted answer
- **Everything else we saw previously holds**



Getting to ChatGPT

- **1990s – "Bag-of-Words**": representing docs via simple word counts
- **2013 – Word Vectors:** a better way to represent docs using vectors trained on actual texts
- **2017 – "Attention"**: a new network module type that allows for variable rather than fixed parameters
- **2019 – Transformers**: NN archetype built on attention modules that better model long texts and allow parallel computation → the first LLM "BERT"
- **2020 – RLHF**: a way to fine-tune LLMs towards producing outputs rated highly by humans using a separate rater network
- **2021 – GPT3**;
- **2022 – InstructGPT/ChatGPT; late 2022 – GPT4**
- **Generative AI is not new**

https://towardsdatascience.com/all-you-need-to-know-about-attention-and-transformers-in-depth-understanding-part-1-552f0b41d021#4c16

Legal Implications

- All machine learners/NNs/LLMs are matrix multiplications
  - **Parameters are numbers computed from data**, will necessarily reflect what the data says (and does not say)
- **"Learning"** is a metaphor for updating parameters
- **"Machine"** is a metaphor for the mathematical matrices and algorithms
  - Or, ChatGPT is an Excel table writ (very) large
  - And by the way, LLMs are more than just ChatGPT
- The power behind LLMs is that next word/sentence estimation really encompasses a wide range of (legal) tasks
- But the fact that **it is really just math** does not mean it is nothing to worry about (quite the contrary)

The Executive's Guide to Getting AI Wrong

It's all math. Really.

by Jerrold Soh

Unless you've been living on a deserted island, you've probably been told that Artificial Intelligence (AI) will 'disrupt' or 'revolutionise' your industry in some way or other. In this Fourth Industrial Revolution, livelihoods will be up-ended, old ways of working will go the way of the dinosaur, old revenue streams will shrivel, and new ones will emerge. So your organisation had better start planning for AI's impact on you, and start building AI into its key business units, processes, and workflows. It comes highly recommended that you do this under a consultant's expert guidance.

What does your mind's eye see when you hear this? Specifically, how does AI disrupting your industry *look* like? Unless you belong to the vast minority of decision-makers with specialised training in AI technology, your closest reference point is probably science fiction, especially of the Hollywood variety. Call this 'Hollywood-style AI': Marvel's J.A.R.V.I.S. (Just A Rather Very Intelligent System), Disney's Wall-E and, for sci-fi aficionados, HAL9000, Robocop, and Terminator. Perhaps you imagine one or all of these characters reporting to work one day, clad in metallic grey suits.

This article explores how we see AI and argues that we mostly get it wrong. In the process, it explains the reasons backed by social science research on why we tend to get AI wrong and illustrates the dangers of doing so from a managerial and law-making perspective. Some readers may also find the article useful as a guide on how and when to manipulate portrayals of AI in your favour.

GETTING AI WRONG

Hollywood-style AI systems are, almost without exception, instances of what philosopher John Searle classically termed 'strong AI': systems which think, act, and quack as humans do.[1] The only difference is that they are manufactured, not birthed. By contrast, 'weak AI' refers to systems programmed to do, and thus capable of doing, only specific tasks. Thus, they are also commonly known as 'narrow AI'. For example, you may be acquainted with basic statistical regression methods. The regression, you may be surprised to learn, is a kind of

https://cmp.smu.edu.sg/sites/cmp.smu.edu.sg/files/pdf/14_AMI17_GettingAIWrong.pdf

Why do we consistently get AI wrong?

- AI is *by definition* a human imitation game
- Market incentives to manipulate narratives
- Few have the technical training to properly; even fewer actually want to know

Legal Dispositionism and Artificially Intelligent Attributions (link)

- Dispositionism: A rational agent has internal reasons, motives, and intentions for acting (the "disposition")
  - So we fault them for 'bad' actions
  - And hold them to informed consent (e.g. contracts, data protection laws, etc)
- Calls for AI personality symptomatic of bias in law towards **dispositionism**
- When AI takes over, a **missing person problem**
- We also instinctively tend to dispositionise AI by giving it needs, wants, morals, thoughts, and a body

Wilson, *Cast Away*
Example from: Hanson, 2008

The Missing Person Problem

The law assumes that persons

{drive, contract, … , argue cases}

But, now, AI does it instead.

Therefore, the law

{is ill-equipped for AI, will be disrupted, … , needs reform}

And, in particular, we need to identify who else to hold liable/should consider AI personality.

The AI and the Situation

- **BUT psychologist**: actions result as much, if not more, from external circumstances (the "situation")
- Yet we are systematically biased towards attributing actions to disposition while missing the situation
  - "Fundamental Attribution Error" (Ross, 1977)

- Do we focus too much on 'the AI'?
- **AI's "situation"**: regulator, manufacturer, programmer, operator, users, etc

Case Study in the Article

- DABUS litigation filed by the "Artificial Inventor Project"
  - Argues that DABUS (an AI system should be registered as patent inventor)
  - More recently, trying similar arguments with copyright. See e.g. Thaler v Perlmutter (2023)
- DABUS "perceives like a person, thinks like a person, and subjectively feels like a person, abductively implicating it as a person"
  - From a rather hard to get article published by Thaler in the 'Journal of Artificial Intelligence and Consciousness'
- Invention apparently "autonomously generated" by DABUS. *Really?*
  - **In what sense is DABUS truly 'autonomous'?**
  - **Note**: not enough to show that AI could be autonomous in theory, we need to show this specific system to be autonomous in fact
  - Interesting how far the courts entertained these arguments (often, uncontested)
- Some courts bought the hype and went close to suggesting DABUS could be inventor
- Some (mostly senior) courts (thankfully) didn't; AIP failed in all jurisdictions so far

Lawyers framing AI dispositionally seldom seem to realise they may be personifying maths. **Reinforced by science fiction, our dispositionist tendencies lead us to conceive of AI systems as autonomous beings, seeing disposition when we should be seeing situation.** This tendency to personify AI has been identified by AI researchers as an 'anthropomorphic bias' and by legal scholars as an 'android fallacy'.

…

Legal narratives which dispositionise AI must therefore be scrutinised. Notwithstanding the imagery that wishful AI mnemonics conjure, they are **inexact metaphors for inevitably statistical computations.** To recall, 'neurons' are standalone statistical algorithms which compute numerical weights from data. 'Training' is the process of passing data through algebra to compute these weights. 'Attention' means increasing the numerical weights accorded to outputs from certain parts of the network. 'Memory' is particular type of neuron (i.e. computation) which feeds into itself such that previous computations influence subsequent ones more directly. **These metaphors make the maths appear as if it has its own mind but neither entail nor imply that it does. As Cardozo CJ famously held, '[m]etaphors in law are to be narrowly watched, for starting as devices to liberate thought, they end often by enslaving it'.** Likewise, Calo notes that judges' 'selection of a metaphor or analogy for a new technology can determine legal outcomes' surrounding AI.

Legal Dispositionism, part 3(b)

Meanwhile…

FORTUNE

Subscribe Now SIGN IN

Elon Musk and his archrival Sam Altman are racing to create a superintelligent A.I. to save humanity from extinction

BY **PAIGE HAGY**

July 19, 2023 at 2:41 AM GMT+8

Most Popular

FINANCE

This 51-year-old renter with stage-3 breast cancer thought she'd 'go to her grave' without owning a home. That was...

https://fortune.com/2023/07/18/elon-musk-xai-sam-altman-openai-artificial-superintelligence/

SMU Classification: Restricted

In this Video…

- What is AI really?
- How “machine learning” works (but too briefly)
- How to think about AI systems in law