AI Changes Everything

Alan Dix

Cardiff Metropolitan University

Draft 26th May 2025

Notes on talk given at Cardiff Metropolitan University, 29th April 2025 as part of the final event of Data Fusion Dynamics: A Collaborative UK-Saudi Initiative in Cybersecurity and Artificial Intelligence.

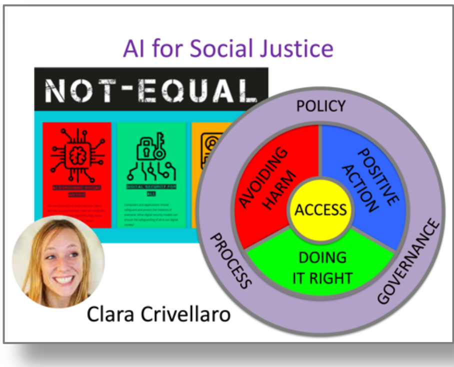
Is AI just another technology, or does it fundamentally change the way we live think?

Every technology has a direct impact with micro-ethical consequences, some good, some bad. However more profound are the ways in which some technologies reshape the very fabric of society with macro-ethical impacts. The invention of the stirrup revolutionised mounted combat, but as a side effect gave rise to the feudal system, which still shapes politics today. The internal combustion engine offers personal freedom and creates pollution but has also transformed the nature of urban planning and international trade. When we look at AI the micro-ethical issues, such as bias, are most obvious, but the macro-ethical challenges may be greater.

At a micro-ethical level AI has the potential to deepen social, ethnic and gender bias, issues I have warned about since the early 1990s! It is also being used increasingly on the battlefield. However, it also offers amazing opportunities in health and educations, as the recent Nobel prizes for the developers of AlphaFold illustrate. More radically, the need to encode ethics acts as a mirror to surface essential ethical problems and conflicts.

At the macro-ethical level, by the early 2000s digital technology had already begun to undermine sovereignty (e.g. gambling), market economics (through network effects and emergent monopolies), and the very meaning of money. Modern AI is the child of big data, big computation and ultimately big business, intensifying the inherent tendency of digital technology to concentrate power. AI is already unravelling the fundamentals of the social, political and economic world around us, but this is a world that needs radical reimagining to overcome the global environmental and human challenges that confront us. Our challenge is whether to let the threads fall as they may, or to use them to weave a better future.

1 Background



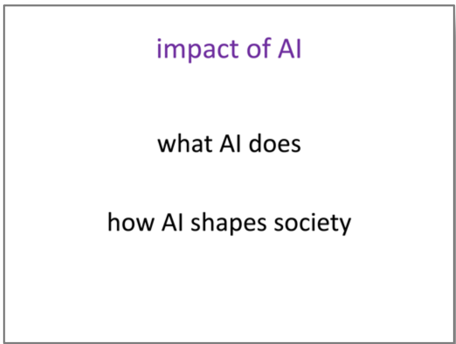
These are notes of a talk given at Cardiff Metropolitan University, 29th April 2025 as part of the final event of Data Fusion Dynamics: A Collaborative UK-Saudi Initiative in Cybersecurity and Artificial Intelligence funded by the British Council UK-Saudi Challenge Fund 2024 [Ta25].

The talk builds on a variety of projects, collaborations and books. It draws particularly on work with colleagues on the UKRI funded NotEqual NetworkPlus (https://not-equal.tech/) and the TANGO Horizon project on synergistic human-machine decision making (https://tango-horizon.eu/).

Further information about topics discussed below can be found in part IV of the book “Artificial Intelligence – Humans at the Heart of Algorithms” [Dx25] and a forthcoming volume “AI for Social Justice” to be co-authored with Clara Crivellaro [CD26].

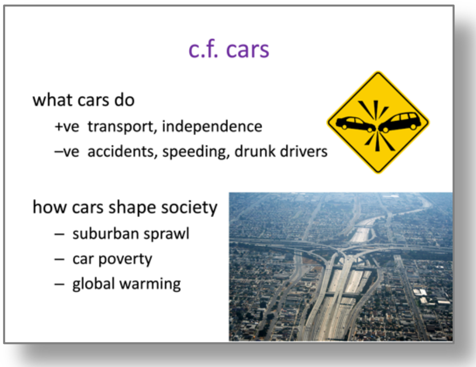
Web versions of Query by Browsing (mentioned later) and other research prototypes can be found in my ‘Alan Labs’ microsite (https://alandix.com/labs/).

2 The Impact of AI



AI has two impacts: *what AI does* directly and *how AI shapes society*

AI, like any technology, has a direct effect, the things you do with it: create an image of a skateboarding cat, suggest the next movies to watch, or medical research. However, some technologies, do more than that, they change the very nature of the world in which we live.

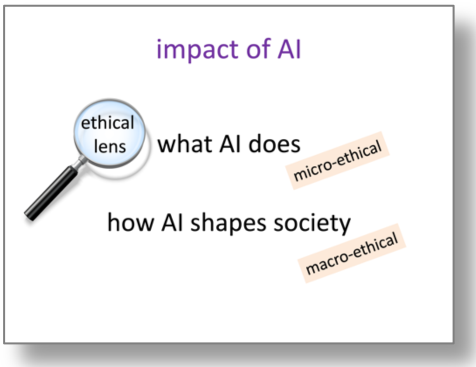


(image: By Remi Jouan - Photo taken by Remi Jouan, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=7245143)

To see this more clearly, let's consider cars.

First there is *what cars do*. Some of this direct impact is good: offering transport, and feelings of independence – think of the great road movies. Some is bad: accidents, speeding and drunk driving.

But more profound is *how cars shape society*. The car has fundamentally changed both our conceptions of distance and determined the fundamental development of the urban environment – enabling suburban sprawl, the out-of-town mall, and consequentially creating car poverty. Furthermore, road transport accounts for 1/3 of greenhouse gas emissions in Europe [EP19] (the only sector where these are growing) and a significant proportion of particulate emissions so impacting the health of our environment and our bodies … even if you are one of the 84% of people in the world who do not own one [MS24].

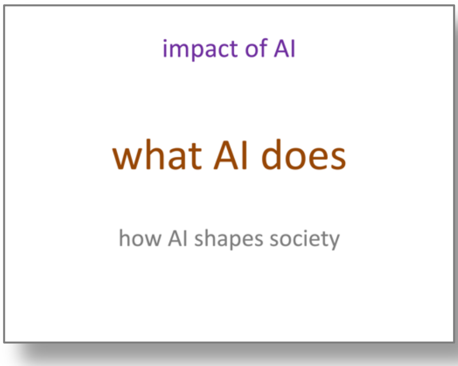


(image: AlphaZeta - Own work, CC0, https://commons.wikimedia.org/w/index.php?curid=16256700 https://simple.wikipedia.org/wiki/File:Magnifying_glass_icon_mgx2.svg)

In the rest of this paper we'll consider these impacts: both what AI does and how AI shapes society, exploring each under an ethical lens. We can think of these as *micro-ethical* and *macro-ethical* consequences of AI, although some of the 'micro'-ethical impacts are themselves pretty large.

We'll look first at examples of the direct effect of AI: including the way it can be used for good, for example in the medical domain, but also the dangers of bias. This will also lead to seeing how AI often acts as a mirror allowing us to see more clearly existing personal and ethical issues. We will then move on to the impact of AI on society, and see how the new AI has arisen out of commercial web businesses through the great 'AI winter', and the way that digital technology, accelerated by AI, fundamentally undermines the premises of market economics.

3 What AI Does



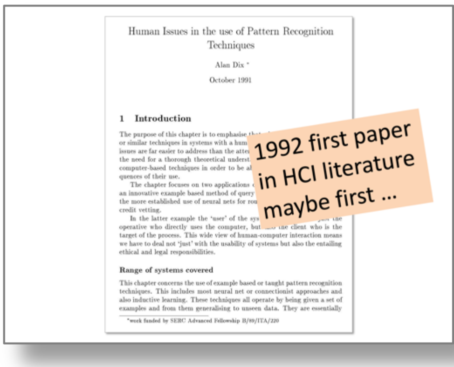
We start with the direct effects of AI, using several case studies and examples to explore both the good and more worrying aspects of the use of AI.

3.1 AI For Good

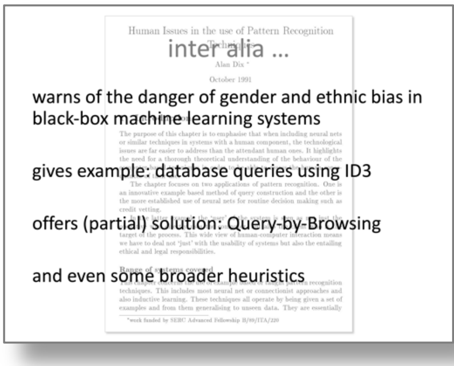
(images: [NF24], CSBIOPASSION, CC BY-SA 4.0, via Wikimedia Commons. https://commons.wikimedia.org/wiki/File:C12orf29_AlphaFold.png)

When focusing on the ethical issues, it is easy to focus on problems, so it is worth starting with the positive, the way AI can be used for good in society. The 2024 Nobel Prize in Chemistry was jointly awarded to three scientists [NF24]. One, David Baker, is a chemist who used computational methods to develop new proteins, the others, Demis Hassabis and John Jumper, are Google engineers who created AlphaFold2, which will aid the future work of researchers such as Baker and enable new advances, especially in pharmaceutical development. As well as at this chemical level, AI is already making major contributions to the analysis of large medical datasets, and many other areas of life.

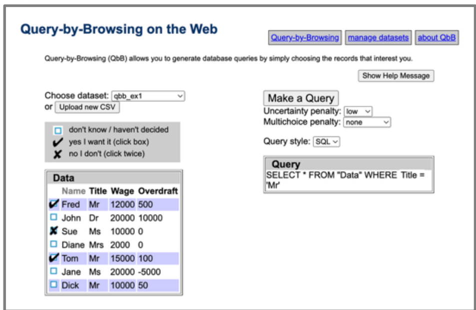
3.2 Bias and Discrimination



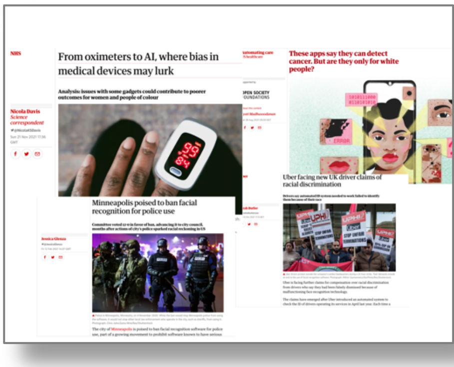
In 1992 I published what I believe to be the first paper to highlight the potential dangers of ethnic, gender and socio-economic bias in black-box machine learning a systems [Dx92]. The paper was developed from a talk at a workshop on pattern recognition and human–computer interaction in early 1991 organised by Janet Finlay and Russell Beale. This was at the height of the 1990 AI summer, which had been driven by the re-vitalisation of neural-network research based on backpropagation [RH86].



As well as warning of the dangers of AI systems creating bias, it digs into some of the underlying mechanisms, such as the dangers of proxy variables, which effectively reveal protected characteristics. In addition, it used a putative system 'Query by Browsing' (QbB), to show how this might inadvertently arise and to exemplify how this might be tackled in part by heuristics such as increased transparency and explainability. A web-based implementation of QbB is available for you to experiment with [Dx25b].



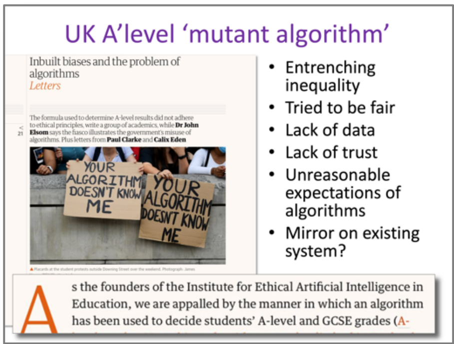
Quqery-by-Browsing on the Web (https://alandix.com/labs/qbb/)



(images: [Da21, Gl21, Ma21, Bu21])

At the time I believed this was going to be an issue within the next five years as there were already proposals to use neural networks in recruitment, credit checking and even nuclear fusion [Gr91]. In fact, it took twenty-five years but now reports of new cases of AI bias are ever present in the media (e.g. [Da21, Gl21, Ma21, Bu21]).

3.3 A Case Study – The A' level 'Mutant Algorithm'



(image: [Gu00])

On 18th March 2020, in the face of rising Covid-19 cases, the UK Government announced that there would be a nationwide lockdown including shutting schools and the cancelation of GCSE and A' level exams [WA00]. The latter, normally taken at 18 years of age, are not only a school leaving exam, but also used to determine entrance to university courses. They are standardised exams run by a small number of exam boards, but are usually preceded by teachers' predicted grades, which are used by universities as the basis of offers. However, teachers' predicted grades were widely believed to be inaccurate, in particular underpredicting top grades for those in lower socio-economic groups. Ofqual (The Office of Qualifications and Examinations Regulation), the arms-length organisation that monitors educational standards, was given the task of determining a way to award grades, without exams, but which maintained farness between years and between different demographic groups.

Ofqual produced an algorithm, but when the results were announced in the summer of 2020, they caused uproar. In some cases, teachers' predicted grades (formally called CAGs – Centre Assessed Grades) were raised, in others lowered, and many felt this had unfairly affected them.

In particular, there was a focus on the fact that the algorithm seemed to entrench existing gaps in society with a greater rise in the number of top A' level grades for independent fee-paying schools compared with state schools [DM20]. Eventually this led to the government denouncing the algorithm and using the higher of CAGs and the algorithm result, and only the CAG the following year [Co20,Fx20].

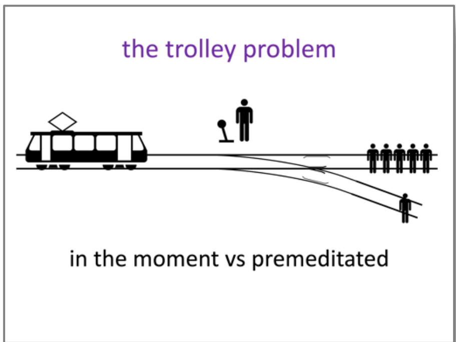
Although the algorithm was criticised for not applying ethical algorithm design principles [Gu00], Ofqual had attempted to be fair. As soon as the decision to scrap exams was made, they commissioned a literature review to determine the evidence for human bias in CAGs. The review struggled to find solid data, but did conclude that there was evidence of socio-economic differences in particular, and this was confirmed by a more extensive review the following year [LW20,LN21]. Furthermore, later analysis showed that the situation after the algorithm was scrapped was worse from the point of view of socio-economic fairness [AM21,AM21b].

3.4 An Ethical Mirror

(image: https://stablediffusionweb.com/image/20424059-reflection-of-a-robot-in-a-mirror)

Shannon Vallor [Va24a, Va24b] talks about AI as a mirror in the sense that it reproduces our existing patterns of thought including biases and this recapitulates and reinforces these, often without the nuances and depth of the original.

However, a mirror also *reveals*, by looking in a mirror we can see ourselves in a way that we cannot when looking outwards. For a physical mirror this is about surface appearance, but AI can force us to examine ethical issues we might otherwise avoid and to see societal issues that were previously buried



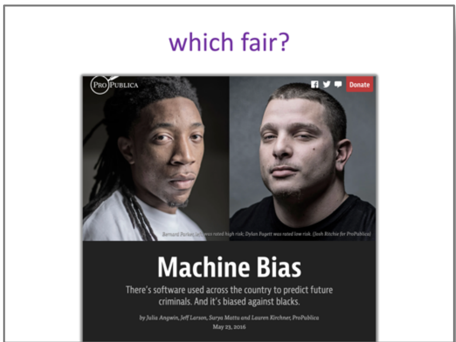
The ‘trolley problem’ has been discussed by moral philosophers for at least fifty years, with many variants [Th76]. A trolley (a US English term for a small train or metro) is heading down the tracks and will hit five people unless you divert it to an alternative track where it would kill just one – would you pull the lever? What about situations where the single person was a trespasser on the line, but the five were supposed to be there?mWhat about five million people versus one? The protagonists in films are sometimes placed in similar situations, albeit created by psychopathic villains rather than philosophers.

There are real-life situations like this, one of the most famous is when the British government received intelligence about the bombing of Coventry, but did not issue any warnings, as to do so would have alerted the Nazi’s to the fact that their communications were compromised. A less graphic, but equally life affecting decision, is when governments make choices between taxation and health services.

Most of us are not faced with such decisions, but imagine instead you are driving down a road and suddenly a child (or what might be a child) rushes into the road ahead. You swerve and crash into a bus queue killing five people. It would be hard to deal with the natural guilt that you would feel after this, but no one would *blame* you. You were reacting in the moment, without the time to fully assess the situation, and ponder an appropriate ethical principle.

Premeditation changes ethics. Autonomous cars are able to amass more information, more rapidly than a human, assessing not just what is in the road ahead, but also who may lie in an alternative path, and moreover apply higher-order rules rapidly and consistently. The developer or manufacturer needs to configure rules ahead of time possibly an utilitarian ‘least harm to least people’ principle, or traffic rules based one ‘favour pedestrians not on the road’. Indeed, these rules may need to shift from country to country depending on local laws, or, heaven forbid, be enabled as options in a personal ethical preferences settings panel. These decisions up front have a different ethical force to the in-the-moment reaction when faced with an emergency.

3.5 Case Study – COMPAS: Fairness in Conflict



(image: [AL16])

The many stories in the media about bias, sadly foreseen by my 1992 paper, are often explained in terms of bad data: if the training data is based on prior human decisions, then the system trained using it will embody all the bias and prejudice within those decisions: garbage in – garbage out. It is tempting to believe that if one can expunge the human bias from the training data, all will be well. There are techniques to achieve this, and it is certainly worth making the best efforts to clean data and to ensure that a wide range of groups of people are included, so that the data truly is representative. Furthermore, analysis of past and current data, for example wage gap figures, is a powerful mirror on human decisions and behaviour.

However, it is dangerous to assume that it is sufficient to 'de-bias' training data. There are a number of reasons for this, but one of the most common is connected with *base rates*, the underlying proportions of people with different characteristics that exhibit some trait. This may vary for intrinsic reasons, for example the rate of pregnancy is higher in the female population than among males. It is also often a reflection of past societal injustice, for example, the crime rates (or to be precise the *recorded* crime rates) in the US have marked racial differences, but this directly reflects socio-economic differences (poverty does engender crime), which are themselves an ongoing symptom of past and current racial inequity.

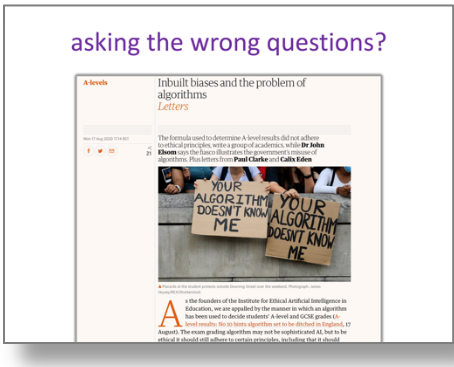
One of the most widely reported (and debated) systems in AI is the COMPAS system, which is used in the US to aid parole decisions. It takes in a wide range of data concerning past criminal history as well as other indicators as input, and creates a low/medium/high risk indicator of the likelihood of reoffending. The judge at a parole hearing can consult this and use it as an aid in their deliberations.

ProPublica magazine performed their own analysis of COMPAS based on court data from a particular area. They concluded that the system was "biased against blacks" [AL16] as apparently similar profiles were more likely to give rise to a 'high risk' indication if the person seeking parole was black. Given the system will not explicitly use race, this will be due to proxy variables, for example, perhaps higher rates of juvenile offences. Despite the massive controversy around this, in the UK there has been recent government project to create an 'Minority Report'-style AI 'murder prediction' system that could be used, inter alia, to inform parole. [Do25]

In their detailed analysis ProPublica also note that a man given a 'high risk' indicator is more likely to reoffend than a woman given the same rating, demonstrating an apparent bias against women [LM16]. This is in part due to the higher base-rate for male reoffending, and in fact the male:female data and black:white data follow very similar patterns; if one used the fairness metric used for the original article (equal results for equivalent profiles –*individual fairness*) on the male:female data it would conclude that COMPAS was biased against men, and if one used the male:female fairness metric (similar overall demographic profile – equalised odds), then the black:white results would also reverse. The two fairness measures run counter to one another, making one better makes the other worse – because *society is itself unfair*.

For the judge using such a system, they cannot reverse several centuries of injustice in one parole decisions, but the information could be used in different ways. If someone is scored as being of high risk of re-offending, this could be (and is being) used as an indication to deny them parole and continue to incarcerate them. Alternatively, it could be used to direct post-release support using the AI to identify which particular aspects of their profile are problematic, so creating a tailored programme of rehabilitation.

3.6 Recapitulation – A'levels: Asking the Wrong Question?



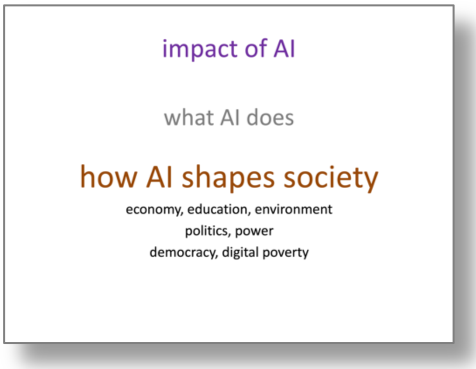
(image: [Gu00])

Turning back to the UK A' level fiasco. Even during the furore about the results in the summer of 2020, it was evident that the fact that fee-paying schools had a greater increase in top grades (4.7% vs 2%). was a direct reflection of the fact that they already were gaining more than twice as many top grades as comprehensive (state funded) schools (48.6% vs 21.8%) [DM20,Of20]. That is, the key issue was not the algorithm, but the massive existing educational divide within British education system.

At a deeper level, it is worth reflecting that the reason that this was such an important issue, is that A' level results are used for university admissions and more broadly make a huge difference to the rest of a person's life: A' levels add about 13% to earning potential, and a degree adds other third, with even greater gains for a 'top tier' university degree and STEM subjects [CP15,SM23,SM24]. Ignoring the effect of the Covid non-exam period, the ability to pay for a school doubles your chances of getting top grades, and even within the state system those from lower socio-economic backgrounds achieve less during their school years, differences that persist through life.

Looking at the 'mutant' algorithm, one can only wonder whether the real issue was the grades given or an educational system that makes attainment at age 18, based on choices made at puberty, highly variable home environments and whether you have a cold during the exam period, is really the appropriate way to fix people's future life chances?

4 How AI Shapes Society

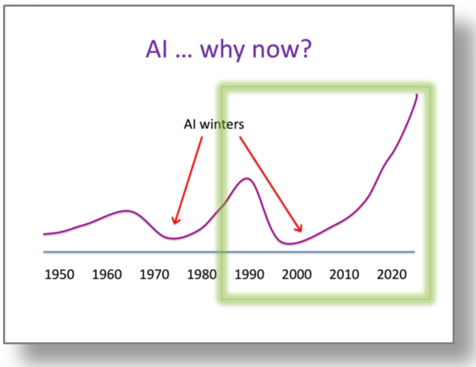


The issues we have been discussing so far are pretty far ranging, but still 'micro'-ethics' in the sense that they concern the direct effects of specific AI technologies. We now turn to macro-ethical issues that arise because of the way AI is reshaping the whole of society.

Some of these effects are described well elsewhere. At an individual level, the costs of AI are already deepening the digital divide, both in terms of access to [BN24] and inclusion in AI [St24], and threatening jobs [Me25]. At a social level there is evidence[1] that recommendation algorithms of social media and search engines may create echo chambers that exacerbate growing political rifts [BB24,TT24]. At a corporate level high tech influence and lobbying is potentially affecting policy in areas from intellectual property to nuclear power [Mi25,Ja25]

Here we'll look at how the 'New AI' emerged from 2000s big tech, the feedback loops this generates and how this changes the very underpinnings of market economies as well as concomitant environmental and democratic impacts.

4.1 The New AI

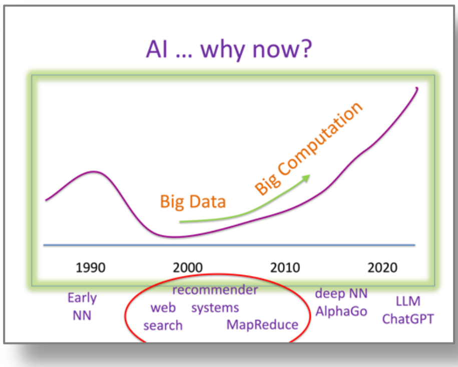


The history of AI is often seen as a roller-coaster journey with peaks of excitement punctuated by long AI winters when some of the over-hyped promises and expectations proved harder to achieve

[1] There are some mixed results from systematic reviews, which seems to be related to research sources, with results based on self-reporting showing less impact that those using behavioural data, suggesting that echo chambers are having an impact, but that those affected underestimate it [TB21, RR22, HW25].

than promised. The peak around 1990, when I first write about AI bias [Dx92], was followed by the long AI winter after neural networks failed to deliver, which only ended with the coming of deep neural networks and AlphaGo [Me16, SH16] in the 2010s.

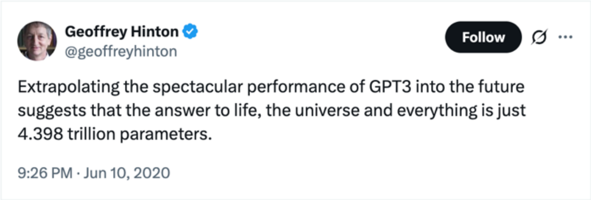
This 'New AI' may well be raising expectations too high but is unlikely to give rise to another winter as DNNs moved from games (no matter how hard) to more practical applications such as AlphaFold [JE21] and of course large language models (LLMs) such as ChatGPT [OA22], and AI is now embedded into an increasing proportion of software.



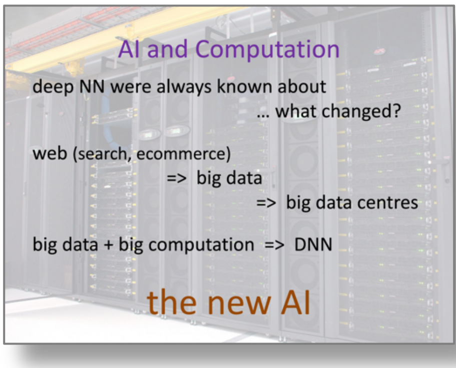
The long AI winter is portrayed as a period when AI scientists retreated into their labs weathering the storms. However, AI did not simply reappear from nowhere in the 2010s, but the groundwork of the revival was being laid in day-to-day systems such as web search and recommender systems [RV97, Ag16]. These were not labelled 'AI' as they used combinations of statistical techniques and large-matrix algebra applied to vast datasets. However, a common definition of AI is when machines do things that we would thing smart if a human did them. By this measure these web systems were 'intelligent', or even – in the case of Google's iconic, "I'm feeling lucky", almost oracular.

Google's PageRank algorithm [BP98] can be regarded as finding an eigenvalue of a vast (but sparse) matrix with a column and row for every web page (plain matrix algebra) or, equivalently, as the stable probabilities of a Markov model (probability theory). However, the Markov model is a 'drunkards walk' around the web a form of swarm computing (usually regarded as AI). In addition, PageRank can be seen as a form of spreading activation over the network of pages similar to connectionist cognitive models(again classic AI). That is while AI did not accept these practical algorithms at the time, they were 'AI'.

In addition, these big data techniques drove the development of large-scale cloud algorithms, not least MapReduce [DG08], which made large-scale computation easier and hence fuelling yet more data collection. Perhaps more crucially, this process demonstrated the "unreasonable effectiveness of data" [HN09], the way in which relatively simple algorithms, such as n-grams, when applied to vast quantities of data, are able to achieve tasks, such as language translation, that had previously been assumed to need knowledge-rich AI. All you need is *scale* [Hi20].



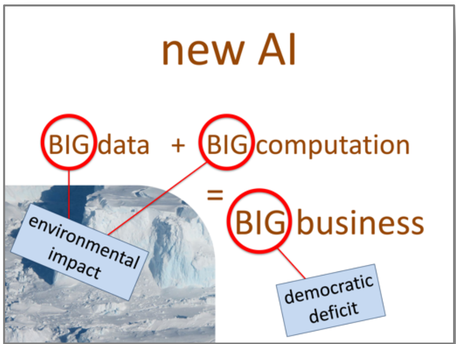
Source Geoffery Hinton, Twitter [Hi20]



Turning back to the AI spring, the idea of deep neural networks was not new. From the onset of multi-layer neural networks and backpropagation, many layers were theoretically possible. So what changed?

The internal layers in a deep network are always poorly constrained, and thus training needed lots of data and would be intolerably slow given the computation available in the late 1980s and early 1990s. By the 2010s, this had all changed. The web and recommender systems for e-commerce and social media had given rise to vast datasets meant that there was now both big data and also cloud computing infrastructure for large-scale computation.

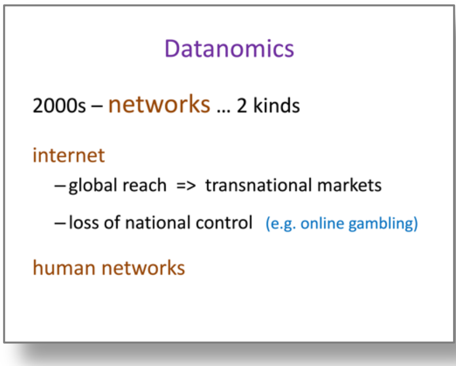
It is this combination of big data and big computation that enabled DNNs and the rise of new AI.



Of course, this big data and big computation did not arise in academic labs, nor even government research programmes, but from big business. Some of this was then available more widely through platforms such as AWS (Amazon web Services) or Microsoft Azure, and also at university computer clusters but, on the whole, the largest computations, including the development of LLMs, has meant that only big business can afford it.

Even with the most benign business, if only big tech companies in a few countries can afford to create cutting-edge AI, this clearly creates a democratic deficit. But the problems cut deeper, undermining the very foundations of market economies.

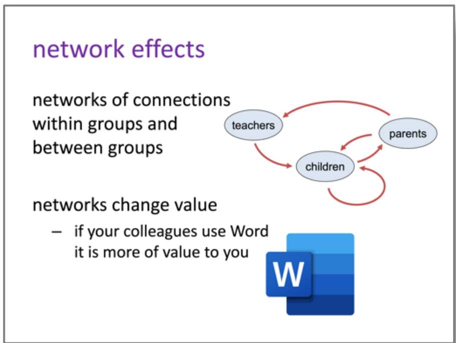
4.2 Datanomics – Network Effects and Emergent Monopolies



Two kinds of networks have defined digital growth in the twentieth century. Both have fundamentally reshaped the nature of social, national and transnational life.

The most obvious is the digital communication and information network of the internet and web. E-commerce has enabled a global reach of companies and transnational markets, without the immediate need for physical presence, think, for example, of the growth of Temu. Of course, physical logistics is still necessary, as is evident by Amazon's vast warehouses, but while it is now a major global logistics player, Amazon itself started with the web and the post. Amongst other things this has led to a loss of national control. The nascent web gambling industry of the early 2000s was an early example of this, creating an uneven playing field in the UK between regulated (and taxed!) physical betting shops and the unregulated web and ultimately to the relaxation of regulation in the 2005 Gambling Act, which even its architects later regretted due to the ensuring massive rise in gambling addiction.

However, while these obvious digital networks have clear impacts on society, perhaps more profound is the way in which they interact with human networks. This is obvious in social network applications, but is also critical for less clearly social application, due to, what are called in economics, *network effects*.

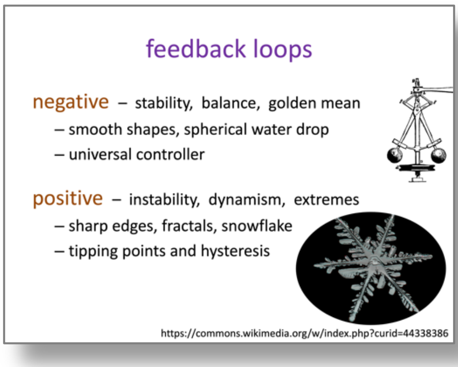


We all live within networks of interactions with other people: family, friends, work colleagues, some physical, some online. Some are highly individual, some related to various groups to which we belong, for example teachers have daily connections with the pupils they teach, who themselves have parents, who in turn may occasionally interact with teachers at parent evenings or at the school gate.

These connections change the value of both physical and digital goods. If I use Microsoft Word and you want to work with me, then the value of Word for you becomes higher than if I used a different word processor. Each of our individual choices affects not only our own personal benefits and costs, but also, because of the need to share or collaborate on documents, those of others.

Economists talk about '*externalities'* the costs and benefits that lie outside a particular economic transaction; for example, the way pollution causes diffuse but widespread environmental and health problems that are not costed in the original transaction.

Environmental externalities do not clearly feed back to the original industries and customers; indeed this has been part of their problem. In contrast *network externalities* have very rapid feedback loops. If you use Word, then its value is almost immediately greater for your colleagues, who are then more likely to also use Word, but of course this means that Word is even more useful to you, so you are less likely to switch to another product … even if it were intrinsically better. This is a *positive feedback loop*.

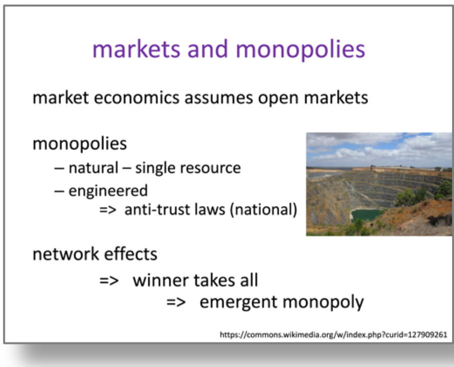


(image: Charles Schmitt – Own work, CC BY-SA 4.0,
https://commons.wikimedia.org/w/index.php?curid=44338386)

Some feedback loops may be negative, where as something gets larger it creates effects which counter its growth. This is exploited in engineered systems such as the universal governor in steam engines, and also within our bodies. On the whole systems with negative feedback settle into stable states and often give rise to smooth shapes such as a spherical water drop.

In contrast positive feedback loops lead to unstable systems that rapidly shift between extremes and often exponential growth, as we saw with Covid-19. This can be used to give benefit. In our own bodies some forms of T-cells have positive feedback loops to enable hair-trigger response to infections, but these are themselves normally regulated by higher-level negative feedback loops, so as to only use the rapid growth when needed.

Physically positive feedback often leads to jagged, pointy or fractal shapes such as the snowflake. In dynamic systems it leads to tipping points and exponential growth or hysteresis.



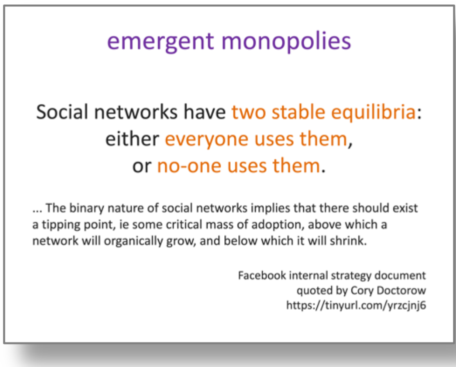
(image: Calistemon - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=127909261)

Our society in virtually every part of the world is underpinned by market economics. The success of markets is, it is argued, due to the way they self-regulate to encourage efficiency. If demand for some product persistently exceeds supply, then the price rises, encouraging new producers to enter the market, eventually balancing supply and demand. That is a negative feedback loop leading to stability and (within limits) efficient production.

Monopolies break this model. If one person or company controls the production of a product, for example owning the only mine for a mineral, then the balance of power between consumer and producer is lost and there is no self-correction within the market. In the case of the mine, this is a *natural* monopoly, based on a single resource. However, monopolies can also be *engineered*, for example if all the manufacturers of a certain product get together to fix a price. Most countries have forms of 'anti-trust' legislation to prevent this form of 'unfair' pricing (note this definition of 'fair' is an economic one, not necessarily 'fair' ethically), preventing, inter alia, this form of market collusion or one company taking over all of its competitors.

However, network effects create a 'winner takes all' effect whereby one brand of a product dominates – an *emergent monopoly*. Note this is not due to any maleficence, simply the natural workings of positive feedback. This is not to say that companies are powerless, they will often seek legal 'barriers to entry', to ensure that they retain market share, or they can do the opposite through adoption of open standards. But, of course, companies can also make illegitimate use of this 'accidental' power; indeed the economic notion of 'network effects' had been laregly a theoretical academic abstraction until it rose to prominence the late 1990s during the US Department of Justice anti-trust case against Microsoft, who were accused of using their near-monopoly position in personal operating systems to unfairly favour Internet Explorer [LM01].

It is also possible for small players to use positive feedback to grow their markets. Indeed, that was my own introduction to these issues in a dot-com era internet startup, before the term 'viral marketing' had become current, where we tried to build an understanding of what we called the 'lattice of value' [Dx99, DF04] and 'marketplace ecology' [Dx99b].

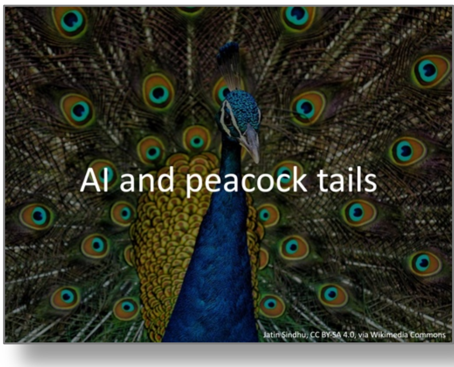


Tech Companies are certainly not unaware of the power of these emergent monopolies. Meta is facing anti-trust charges relating to whether their acquisition of Instagram was intended to stifle competition. In a Medium post, Cory Doctorow quites a Facebook internal strategy document that says: “social networks have two stable equilibria: either everyone uses them, or no-one uses them” [Do25].

These emergent networks mean, that even without bad actors exploiting them, the very nature of digital technology *breaks market economics*. In a world where the rhetoric of the efficiency of free markets drives large swathes of policy as well as global trade itself, this is profound.

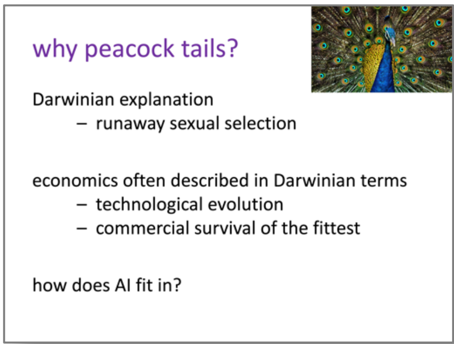
Furthermore, while on its own digital breaks market economics, AI makes this worse. This is partly due to the new AI’s dependence on vast computational resources leading to ever greater concentrations of market power. However, AI creates other feedback effects, that exacerbate this further.

4.3 AI and Peacock Tails



(image: Jatin Sindhu, CC BY-SA 4.0, via Wikimedia Commons
https://commons.wikimedia.org/wiki/File:Peacock_Plumage.jpg)

The peacock's tail is a glorious thing, but odd. Surely such an exuberant display must use valuable sustenance, impede flight and make it hard for peacocks to avoid predators?



The Darwinian explanation for this is runaway sexual selection. Whereas Darwinian selection often leads to species that ideally fit their ecological niches, it can sometimes go awry.

The argument goes as follows:

1. Successful and well-fed males are healthier and therefore have better plumage
2. Females want (instinctively through breeding not consciously) a healthy mate to have better offspring
3. Females therefore choose mates with better (natural) plumage
4. Males develop even better plumage (even at the expense of health) in order to attract mates
5. The level of decoration necessary to determine male good health is raised, therefore females become more selective about good plumage
6. Yet more cycles of (4) and (5)

The peacock's plumage is not well suited to the survival of the species, but the positive feedback process of sexual section has led inexorably towards this sub-optimal state.

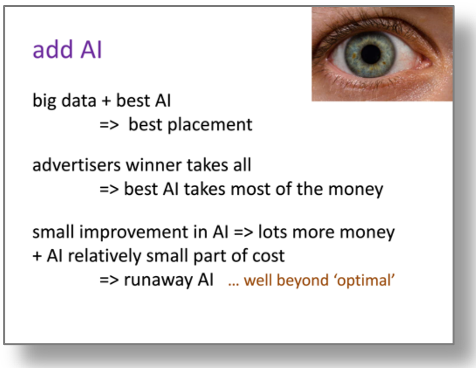
Economics is often also described in Darwinian terms including narratives of technological evolution and commercial survival of the fittest. The argument (monopolies withstanding) is that this normally leads to efficient processes and utilisation of resources. AI is part of this process, but can lead to peacock-tail outcomes.

(image: ROTFLOLEB, CC BY-SA 3.0, via Wikimedia Commons https://commons.wikimedia.org/wiki/File:Human_eye_with_blood_vessels.jpg)

Many of the web-based services we use are 'free' in the sense that we do not directly pay for them but of course pay indirectly through our data and attention. Indeed, as advertising ultimately pays for much of the 'free' web, it is often seen as an 'attention economy' where the secret to economic success is in engaging people's eyeballs.

Web sites including social media and search engines tune their offerings to attract and hold on to users (click bait), largely because this maximises the time that their users see adverts. Statistical algorithms and AI are used to tune the material presented, the adverts chosen and even where to place the adverts on a page or within a stream.

Advertisers of course want customers. They may seek coverage of different demographics and so spread advertising across platforms but overwhelmingly pay whoever can manage placement that maximises the likelihood of click-throughs and so advertisers themselves will use algorithms to help determine this. If the AI algorithms on one platform perform better than another then advertisers will switch – winner takes all!



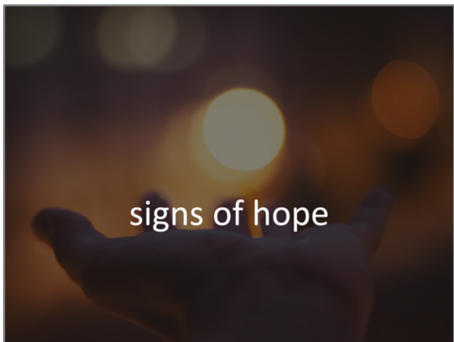
The whole industry is highly tuned, and, like an Olympic sprint, relatively small improvements in AI can change who is that winner. Maintaining a social media platform is expensive so as long as the per-page costs of AI are still relatively small there is almost inevitably a runaway growth in AI way beyond the 'optimal' point in terms of overall industry advertising spend – a bit like the prisoner's dilemma.

Ultimately consumers pay as unnecessary advertising spend adds costs to sellers and hence buyers. As a side effect more money goes to the small number of semi-monopoly platforms, reducing further the proportion who can afford the massive AI development spend, further entrenching those monopolies. And, of course, AI beyond the optimum unnecessarily consumes computational resources and energy.

4.4 Financial and Environmental Cost

(images: [Tr19], [SG20], NASA ICE – Public Domain https://commons.wikimedia.org/w/index.php?curid=24450299)

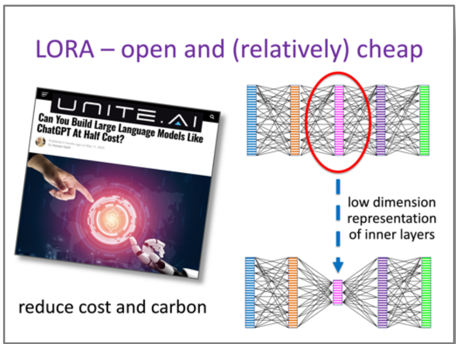
We are all aware of the increasing environmental costs of digital technology, this was evident well before the rise of cybercurrencies and LLMs, but has intensified due to AI [Tr19, SG20, Sa20]. Early Bitcoin mining was often powered by coal power stations (and often particularly dirty ones) and indeed continues to do so [We25]. Facing increasing energy costs of computation, the big AI players have been seeking cleaner power sources. This move to provide AI from low carbon sources is leading to a resurgence of nuclear power with Microsoft reopening Three Miles Island [Sh24] and Amazon lobbying for more nuclear power in the UK [Ja25]. Although there are debates about whether nuclear power has a place alongside true renewables as part of a low-carbon energy mix, the idea of a new nuclear proliferation driven by AI is disturbing. Paradoxically, many of the most beneficial uses of AI, for example, medical research applications, could be run on a ‘when power is abundant’ basis [Dx17], favouring a greater renewable proportion, but of course the commercial drivers are for instant access.



Happily there are signs of hope, from both academic research and commercial developments, with significant efforts aimed at reducing the computational demands of AI, which often, as a side effect, also serve to make the use of AI more democratic.

(images: [Di22], https://github.com/deepseek-ai/DeepSeek-V3)

On the commercial side, Meta's release of pretrained open-source foundation models based on Llama 2 was a major step to make LLMs more widely available [ZD22, Di22, TH23]. Perhaps more dramatic still, certainly for the general public, was the release of DeepSeek 2 [LF24], which had performance matching or outperforming the major LLMs engines, but at a fraction of the training and running costs. It is widely believed that DeepSeek's training includes a 'distillation' of OpenAI models [We25b], so the true training gains may be less. This said, the MoE (mixture of experts) architecture of DeepSeek 2 means that only a fraction of the network is activated for each query, substantially reducing execution costs, and its release as open source has created opportunities for research and innovation.

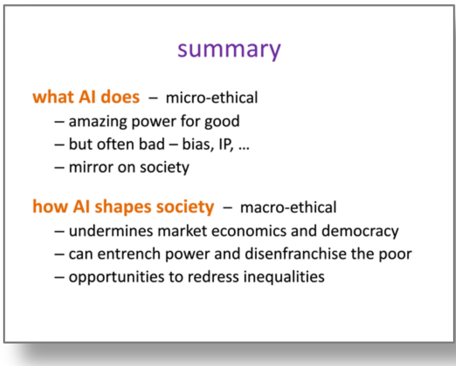


(images: [Sa23] [Dx25])

The development of DeepSeek was driven by the US export ban of NVIDIA chips to China, forcing a more computationally efficient approach. There is also a more general movement within the academic and commercial AI community to create more computationally 'lean' algorithms, both for commercial reasons (lower costs, running on phones) and societal benefit (reducing carbon footprint, widening digital access).

Some techniques seek to reduce the size of models, for example NPAS (neural parameter allocation search), automatically looks for opportunities to share parameters in large networks [PD20]; and LoRA (Low-Rank Adaptation) does this by creating lower-dimensional representations of layers within a trained large language model [HS22]. Others seek to reduce the training cost itself, notably LiGO starts with a small network and progressively increases its size but reusing the weights in the smaller model at each stage to kick-start the next larger iteration [WP23].

5 Summary



We have considered two aspects of the impact of AI and its ethical consequences:

- **what AI does** (*micro-ethical*) – AI offers amazing power for good, but often this is not realised and instead we see negative consequences including bias, IP infringement and job losses. In many cases, AI acts as a mirror on society highlighting ethical problems or societal injustice that were already present, just brought into focus by AI.
- **how AI shapes society** (*macro-ethical*) – We have seen how AI and digital technology more generally undermines both market economics and democracy and, if not curbed, has an intrinsic tendency to entrench power and disenfranchise the poor. However, there are signs of hope, ways in which smarter (rather than simply scaled) AI can offer opportunities to redress inequalities.

In short, AI Changes Everything! As researchers, designers, decision makers or consumers of AI, we can make decisions about where and how to use AI. What will *we* do?

For more about these topics see: https://alandix.com/academic/talks/CMet2025-AI-Changes-Everything/ and https://alandix.com/ai4sj/.

6 Acknowledgments

This work has been supported by the HORIZON Europe projects TANGO - Grant Agreement n.101120763 and SoBigData++ Grant Agreement n. 871042. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them. Parts of this research were also funded by UKRI EPSRC, 'Social justice through the Digital Economy' Not-Equal NetworkPlus, EP/R044929/1.

7 References

[Ag16] Aggarwal, C. C. (2016). *Recommender systems* (Vol. 1). Cham: Springer International Publishing.

[AM21] Anders, J., Macmillan, L., Sturgis, P. and Wyness, G. (2021). *Inequalities in young peoples' educational experiences and wellbeing during the Covid-19 pandemic* (CEPEO Working Paper No. 21-08). Centre for Education Policy and Equalising Opportunities, UCL. https://EconPapers.repec.org/RePEc:ucl:cepeow:21-08

[AM21b] Anders, J., Macmillan, L., Sturgis, P. and Wyness, G. (2021). *The 'graduate parent' advantage in teacher assessed grades*. (blog) UCL, Centre for Education Policy and Equalising Opportunities. 8 June 2021. https://blogs.ucl.ac.uk/cepeo/2021/06/08/thegraduate-parentadvantageinteacherassessedgrades/

[AL16] Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2016). Machine bias there's software used across the country to predict future criminals: And it's biased against blacks. *ProPublica* (23 May 2016). https://www.propublica.org/article/machine-bias-risk-assessments-incriminal-sentencing.

[BN24] Bentley, S.V., Naughtin, C.K., McGrath, M.J., Irons, J.L. and Cooper, P.S.. (2024). The digital divide in action: how experiences of digital technology shape future relationships with artificial intelligence. *AI Ethics* 4, 901–915. https://doi.org/10.1007/s43681-024-00452-3

[BB24] Bliuc, AM., Betts, J.M., Vergani, M., Bouguettaya, A. and Cristea, M. (2024). A theoretical framework for polarization as the gradual fragmentation of a divided society. *Commun Psychol* 2, 75. https://doi.org/10.1038/s44271-024-00125-1

[BP98] Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117.

[Bu21] Butler, S. (2021). Uber facing new UK driver claims of racial discrimination. The Guardian, 6 Oct 2021. https://www.theguardian.com/technology/2021/oct/06/uber-facing-new-uk-driver-claims-of-racial-discrimination

[CD26] Crivellaro, C. and Dix, A. (2026). *AI for Social Justice*. CRC Press (forthcoming).

[CP15] Conlon, G., and Patrignani, P. (2015). *The earnings and employment returns to A levels: A report to the Department for Education*. London Economics, London. https://londoneconomics.co.uk/wp-content/uploads/2015/03/London-Economics-Report-Returns-to-GCE-A-Levels-Final-12-02-2015.pdf

[Co20] Coughlan, S. (2020). A-levels and GCSEs: Boris Johnson blames 'mutant algorithm' for exam fiasco. *BBC News*. 26 August 2020. https://www.bbc.co.uk/news/education-53923279

[Da21] Davis, N. (2021). From oximeters to AI, where bias in medical devices may lurk. *The Guardian*, 21 Nov 2021. https://www.theguardian.com/society/2021/nov/21/from-oximeters-to-ai-where-bias-in-medical-devices-may-lurk

[DG08] Dean, J. and Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Commun: ACM*, 51(1):107–113.

[Di22] Dickson, B. (2022). Can large language models be democratized? *TechTalk*,-May 16, 2022. https://bdtechtalks.com/2022/05/16/opt-175b-large-language-models/

[Dx92] Dix, A. (1992). Human issues in the use of pattern recognition techniques. In *Neural Networks and Pattern Recognition in Human Computer Interaction*, R. Beale and J. Finlay (eds). Ellis Horwood. 429-451. https://alandix.com/academic/papers/neuro92/

[Dx99] Dix, A. (1999). *The Lattice of Value. principles for designing complementary products that foster their own spread*. Talis eBulletin, Sept 1999. https://magisoft.co.uk/alan/ebulletin/lattice-of-value/lattice-of-value.html

[Dx99b] Dix, A. (1999). *Marketplace Ecology – managing the interconnected market groups of the Internet. understanding the way feedback between different groups lead to market growth*. Talis eBulletin, Sept 1999. https://magisoft.co.uk/alan/ebulletin/market-ecology/marketplace-1999.html

[DF04] Dix, A., Finlay,, J. Abowd, G. and Beale, R. (2004). *Human-Computer Interaction*, third edition. Prentice Hall. https://hcibook.com/e3/

[Dx17] Dix, A. (2017). *Abundance a new approach to energy sustainability*. Tiree Tech Wave, 13 Feb. 2017. https://tireetechwave.org/docs/Abundance-white-paper-v0-2.pdf

[Dx25] Dix, A. (2025). *Artificial Intelligence – Humans at the Heart of Algorithms*, 2nd Edition, Chapman and Hall. https://alandix.com/aibook/

[Dx25b] Dix, A. (2025). *Query-by-Browsing: intelligent transparent database access*. Alan Labs. (accessed 17/5/2025). https://alandix.com/labs/qbb/

[Do25] Doctorow, C. (2025). *Mark Zuckerberg personally lost the Facebook antitrust case.* Medium, Apr 18, 2025. https://doctorow.medium.com/https-pluralistic-net-2025-04-18-chatty-zucky-is-you-taking-notes-on-a-criminal-fucking-conspiracy-a449f6aa35ad

[Do25] Dodd, V. (2025). UK creating 'murder prediction' tool to identify people most likely to kill. *The Guardian*, 8 Apr 2025 https://www.theguardian.com/uk-news/2025/apr/08/uk-creating-prediction-tool-to-identify-people-most-likely-to-kill

[DM20] Duncan, P., McIntyre, N., Storer, R. and Levett, C. (2020). Who won and who lost: when A-levels meet the algorithm. *The Guardian*, 13 Aug 2020. https://www.theguardian.com/education/2020/aug/13/who-won-and-who-lost-when-a-levels-meet-the-algorithm

[EP19] European Parliament (2019). *CO2 emissions from cars: facts and figures (infographics)*. European Parliament. https://www.europarl.europa.eu/news/en/headlines/society/20190313STO31218/co2-emissions-from-cars-facts-and-figures-infographics

[Fx20] Foxglove Legal (2020). *We put a stop to the A Level grading algorithm!* Foxglove Blog, 17th August 2020. https://www.foxglove.org.uk/2020/08/17/we-put-a-stop-to-the-a-level-grading-algorithm/

[Gl21] Glenza, J. (2021). Minneapolis poised to ban facial recognition for police use. *The Guardian*, 12 Feb 2021. https://www.theguardian.com/us-news/2021/feb/12/minneapolis-police-facial-recognition-software

[Gr91] Greake, E. (1991). Neural network keeps fusion plasma in shape. *New Scientist*, page 27, 12th October 1991.

[Gu00] The Guardian (2000). Inbuilt biases and the problem of algorithms. Letters, *The Guardian*, 17 Aug. 2000. https://www.theguardian.com/education/2020/aug/17/inbuilt-biases-and-the-problem-of-algorithms

[HN09] Halevy, A., Norvig, P., and Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24(2):8–12, 2009.

[HW25] Hartmann, D., Wang, S.M., Pohlmann, L. and Berendt B. (2025). A systematic review of echo chamber research: comparative analysis of conceptualizations, operationalizations, and varying outcomes. *J Comput Soc Sc* 8, 52. https://doi.org/10.1007/s42001-025-00381-z

[Hi20] Hinton, G. (2020). *Extrapolating the spectacular performance of GPT3 into the future suggests that the answer to life, the universe and everything is just 4.398 trillion parameters*. Twitter (now X), Jun 10, 2020. https://x.com/geoffreyhinton/status/1270814602931187715

[HS22] Hu, E. J., Shen, Y., et al. (2022). LoRa: Low-rank adaptation of large language models. *ICLR*, 1(2), 3. https://arxiv.org/abs/2106.09685

[Ja25] Jack, S. (2025). UK needs more nuclear to power AI, says Amazon boss. *BBC News*, 16 May 2025. https://www.bbc.co.uk/news/articles/cewd5014wpno

[JE21] Jumper, J., Evans, R., et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2

[LM16] Larson, J., Mattu, S., Kirchner, L. and Angwin, J. (2016). How we analyzed the COMPAS recidivism algorithm. *ProPublica*, 23 May 2016. https://www.propublica.org/article/how-weanalyzed-the-compas-recidivism-algorithm

[LW20] Lee, M.W., and Walter, M. (2020). *Equality impact assessment: literature review.* Office of Qualifications and Examinations Regulation (Ofqual) April 2020. https://assets.publishing.service.gov.uk/media/5e971f1de90e071a145ec51f/Equality_impact_assessment_literature_review_15_April_2020.pdf

[LN21] Lee, M.W., and Newton, P. (2021). *Systematic divergence between teacher and test-based assessment*. Office of Qualifications and Examinations Regulation (Ofqual), 17 May 2021. Ref. Ofqual/21/6781. https://www.gov.uk/government/publications/systematic-divergence-between-teacher-and-test-based-assessment

[LM01] Liebowitz, S., and Margolis, S. (2001). Network effects and the Microsoft case. Chapter 6 in *Dynamic competition and public policy: Technology, innovation, and antitrust issues*, J. Ellig (ed.), pp.160–192. https://personal.utdallas.edu/~liebowit/netwext/ellig%20paper/ellig.htm

[LF24] Liu, A., Feng, B., et al. (2024). *Deepseek-v3 technical report*. arXiv preprint arXiv:2412.19437. https://arxiv.org/abs/2412.19437

[Ma21] Madhusoodanan, J. (2021). These apps say they can detect cancer. But are they only for white people? *The Guardian*, 28 Aug 2021. https://www.theguardian.com/us-news/2021/aug/28/ai-apps-skin-cancer-algorithms-darker

[Me25] Merchant, B. (2025). *The AI jobs crisis is here, now*. Blood in the Machine, 2nd May 2025. https://www.bloodinthemachine.com/p/the-ai-jobs-crisis-is-here-now

[Me16] Metz, C. (2016). In two moves, AlphaGo and Lee Sedol redefined the future. *WIRED* (16 March 2016). https://www.wired.com/2016/03/two-moves-alphago-lee-sedol-redefined-future/

[Mi25] Milmo, D. (2025). UK copyright law consultation 'fixed' in favour of AI firms, peer says. *The Guardian*, 11 Feb 2025. https://www.theguardian.com/technology/2025/feb/11/uk-copyright-law-consultation-fixed-favour-ai-firms-peer-says

[MS24] Miner, P., Smith, B. M., Jani, A., McNeill, G., & Gathorne-Hardy, A. (2024). Car harm: A global review of automobility's harm to people and the environment. *Journal of Transport Geography*, 115, 103817. https://doi.org/10.1016/j.jtrangeo.2024.103817

[NF24] The Nobel Foundation (2024). *The Nobel Prize in Chemistry 2024*. NobelPrize.org. Nobel Prize Outreach 2025. Sat. 17 May 2025. https://www.nobelprize.org/prizes/chemistry/2024/summary/

[Of20] Ofqual. (2020). *Awarding GCSE, AS, A level, advanced extension awards and extended project qualifications in summer 2020: Interim report*. The Office of Qualifications and Examinations Regulation (Ofqual), Technical Report, 13 August 2020.

https://www.gov.uk/government/publications/awarding-gcse-as-a-levels-in-summer-2020-interim-report

[OA22] OpenAI (2022). *Introducing ChatGPT*. OpenAI, November 30, 2022. https://openai.com/index/chatgpt/

[PD20] Plummer, B. A., Dryden, N., Frost, J., Hoefler, T., & Saenko, K. (2020). *Neural parameter allocation search*. arXiv preprint arXiv:2006.10598. https://arxiv.org/abs/2006.10598

[RV97] Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3), 56-58. https://doi.org/10.1145/245108.245121

[RR22] Ross Arguedas, A., Robertson, C., Fletcher, R., & Nielsen, R. (2022). *Echo chambers, filter bubbles, and polarisation: a literature review* . Reuters Institute for the Study of Journalism. https://dx.doi.org/10.60625/risj-etxj-7k60

[RH86] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536. https://doi.org/10.1038/323533a

[Sa20] Saenko, K. (2020). It takes a lot of energy for machines to learn–here's why AI is so power-hungry. *The Conversation*. https://theconversation.com/it-takes-a-lot-of-energy-for-machines-to-learn-heres-why-ai-is-so-power-hungry-151825

[Sa23] Sajid, H. (2023). *Artificial Intelligence: Can You Build Large Language Models Like ChatGPT At Half Cost?* Unite.ai, May 11, 2023. https://www.unite.ai/can-you-build-large-language-models-like-chatgpt-at-half-cost/

[Sh24] Sherman, N. (2024). Microsoft chooses infamous nuclear site for AI power. *BBC News*, 20 September 2024. https://www.bbc.co.uk/news/articles/cx25v2d7zexo

[SH16] Silver, D., Huang, A., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 484–489. https://doi.org/10.1038/nature16961.

[SM23] Social Mobility Commission. (2023). *Labour market value of higher and further education qualifications: A summary report*. Social Mobility Commission, 9 February 2023. https://www.gov.uk/government/publications/labour-market-value-of-higher-and-further-education-qualifications-a-summary-report

[SM24] Social Mobility Commission (2024). *Income returns to education*. Social Mobility Commission, 11 September 2024 https://social-mobility.data.gov.uk/intermediate_outcomes/work_in_early_adulthood_(25_to_29_years)/income_returns_to_education/2.0

[St24] Stone, E. (2024). *AI shifts the goalposts of digital inclusion*. Joseph Rowntree Foundation, 8 February 2024. https://www.jrf.org.uk/ai-for-public-good/ai-shifts-the-goalposts-of-digital-inclusion

[SG20] Strubell, E., Ganesh, A., & McCallum, A. (2020, April). Energy and policy considerations for modern deep learning research. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 09, pp. 13693-13696). https://arxiv.org/abs/1906.02243

[Ta25] Tahir, S, (2025) Data Fusion Dynamics: A collaborative UK–Saudi initiative in cybersecurity and artificial intelligence. *Research Features*, 156. March 31, 2025. https://researchfeatures.com/data-fusion-dynamics-collaborative-uk-saudi-initiative-cybersecurity-ai/

[TB21] Terren, L., and Borge, R. (2021). Echo chambers on social media: A systematic review of the literature. *Review of Communication Research*, Vol.9. https://doi.org/10.12840/ISSN.2255-4165.028

[Th76] Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist*, 204-217. https://doi.org/10.5840/monist197659224

[TT24] Törnberg, P. and Törnberg, A. (2024). Inside a White Power echo chamber: Why fringe digital spaces are polarizing politics. *New Media & Society*, 26(8), 4511-4533. https://doi.org/10.1177/14614448221122915

[TH23] Touvron, H., Martin, L., et al. (2023). *Llama 2: Open foundation and fine-tuned chat models*. arXiv preprint arXiv:2307.09288. https://arxiv.org/abs/2307.09288

[Tr19] Trueman, C. (2019). Why data centres are the new frontier in the fight against climate change. *Computerworld*, Aug 9, 2019. https://www.computerworld.com/article/3431148/why-data-centres-are-the-new-frontier-in-the-fight-against-climate-change.html [Va24a] Vallor, S. (2024a). The AI mirror: How to reclaim our humanity in an age of machine thinking. Oxford University Press.

[Va24a] Vallor, S. (2024b). *The Turing Lectures: Can we live with AI?* The Alan Turing Institute. 9 Dec 2024. (section@21:27: The Mirror Metaphor for AI) https://www.youtube.com/live/7iX-wiKvYHs?si=W-QSferpGWZFcItI&t=1285

[WA00] Walker, P. and Adams, R. (2020). Coronavirus UK: schools, colleges and nurseries to close. *The Guardian*, 18 Mar 2020. https://www.theguardian.com/world/2020/mar/18/coronavirus-school-colleges-nurseries-england-close-uk-friday

[WP23] Wang, P., Panda, R., et al. (2023). *Learning to grow pretrained models for efficient transformer training*. arXiv preprint arXiv:2303.00980. https://arxiv.org/abs/2303.00980

[We25] Wendling, M. (2025). The loud hum sparking unrest in Trump's MAGA heartlands. *BBC News – InDepth*, 23 May 2025. https://www.bbc.co.uk/news/articles/cx2qg6e03l2o

[We25b] Werner, J. (2025). Did DeepSeek Copy Off Of OpenAI? And What Is Distillation? *Forbes*, Jan 30, 2025. https://www.forbes.com/sites/johnwerner/2025/01/30/did-deepseek-copy-off-of-openai-and-what-is-distillation/

[ZD22] Zhang, S., Diab, M. and Zettlemoyer, L. (2022). *Democratizing access to large-scale language models with OPT-175B*. Meta Research Blog, May 3, 2022. https://ai.meta.com/blog/democratizing-access-to-large-scale-language-models-with-opt-175b/