BCS is a charity with a Royal Charter. Its mission is to make IT better for society. It does this through leadership on societal and professional issues, working with communities and promoting excellence.

BCS brings together industry, academics, practitioners, educators and government to share knowledge, promote new thinking, educate, shape public policy and inform the public. This is achieved through and with a network of 75,000 members across the UK and internationally. BCS is funded through membership fees, through the delivery of a range of professional development tools for practitioners and employers, and as a leading IT qualification body, through a range of widely recognised professional and end-user qualifications. www.bcs.org

Background description of Algorithms

1. The Oxford English Dictionary defines “algorithm” as “A procedure or set of rules used in calculation and problem-solving”. There are various ways of classifying algorithms and their uses.

2. Many “traditional” algorithms, such as optimisation, linear programming, Monte-Carlo Simulation, Dijkstra’s route-finding algorithms, are very powerful, as is covered in the Institute of Mathematics and its Applications submission (especially paragraphs 2, 7, 18, 33). We repeat their call for greater use of these algorithms in Government, and note that searching for a mathematically optimal solution reduces or eliminates personal or political bias, as in “Gerry./mandering” [Klarreich2017].

3. BCS, The Chartered Institute for IT, focuses its submission on those “algorithms” which are actually the result of applying an algorithm in the sense above to an amount of data in order to deduce rules (often described as “an algorithm”) to apply to new data. Hence the rule being applied has been inferred from the existing (or “background”) data. These derived algorithms go by various names: “Machine Learning”, “Deep Learning”, “Neural Nets” and sometimes “Artificial Intelligence” (though many Artificial Intelligence practitioners would object to this usage). We also note that some standard statistical algorithms such as Regression are sometimes described as “Machine Learning” to confuse the picture.

Algorithms can be published or secretly understood, or not understood

1. Traditionally, algorithms, either prescribed like Income tax (published in the Finance Act), or inferred, like the wind chill algorithm, are published algorithms. Many companies operating in regulated industries, generally in finance, have secret algorithms for credit scoring, loan approval etc. These algorithms are part of their competitive edge. Yet, they need to be able to explain them to regulators, and to justify decisions if required. An example might be a car insurance company, whose algorithm might include the step “To the base premium, we add a sum depending on the insured’s occupation, from this table”, where the table is computed based on past claims data. The table, the precise definition of occupations, and indeed whether it's
adding a sum, multiplying by a ratio, or both, are part of the insurer’s trade secrets. Such companies may, indeed do, wish to use more powerful machine-learning techniques, but the regulatory (at least in the USA, and to be imposed by GDPR in Europe) requirement prevents them [Knight2017b].

2. Conversely, many of the algorithms produced by machine learning, notably those based on “deep learning” or neural networks, are not understood at all: no human being can say “why” the algorithm does what it does, nor can predict what it will do on data which are not the training data. [Knight2017a]. Even strong advocates of these admit this major weakness. “Although Deep Neural Networks have demonstrated tremendous effectiveness at a wide range of tasks, when they fail, they often fail spectacularly, producing unexplainable and incoherent results that can leave one to wonder what caused the DNN to make such decisions. The lack of transparency in the decision-making process of DNNs is a significant bottleneck in their widespread adoption in industry, such as healthcare, defence, cybersecurity, etc., where the error tolerance is very low and the ability to interpret, understand, and trust decisions is critical.” [Kumar et al., 2017]. We should also note that the research in that paper, though good, is only looking at specific mistakes, and answering questions like “why did the DNN misclassify this Chihuahua as a Shih-Tzu”, and not “how does the DNN recognise Chihuahuas”.

Algorithms can be Advisory or Determinative

1. The income tax algorithm determines the amount payable: it is not a suggestion to the tax inspector how much to charge. University degree classifications have become more determinative over the years. A recent adoption of a determinative algorithm is the Duckworth-Lewis algorithm in cricket. The algorithm is public, and understood by experts, but not by the general public, who just accept it.

2. Advisory algorithms produce a piece of advice to a human being who makes an ultimate decision. This advice may consist of evaluating several different scenarios, or may just be a simple “answer”. These may or may not have some measure of confidence attached. Of course, an algorithm may in principle only be advisory but the human beings using it may in practice just rubber-stamp its “advice”, so in practice it’s determinative. In principle, the GDPR (Article 22.1) talks about a “decision based solely on automated processing”, i.e. a decisive algorithm. One could expect a lot of litigation in this area.

3. Consider for example the US case of Paul Zilly [reported in Angwin et al. 2016]: defence and prosecution had agreed a plea bargain of a year in prison, but the judge looked at a recidivism score produced by a proprietary (and at least secret, probably not understood), overturned the plea deal, and imposed a two-year sentence. In theory that was an advisory algorithm, but in practice it was being used to overrule the agreement which would have been rubber-stamped by the judge.

Continuous or Discrete?
1. Many algorithms are used that yield numerical results varying continuously with their input. So, if you earn £10 more, you pay £2 (or £4 or …) in income tax, and £20 more is twice that (unless you cross a threshold) and so on. If the wind speed increases by 1 km/h, then the wind chill changes by a certain amount. Other algorithms give one of a small number of discrete results, e.g. degree classification. In many cases the answer is binary: bail or no bail; melanoma (or at least referral for melanoma) or no referral. We use the shorthand “continuous algorithm” for “an algorithm whose results depend continuously on their input”.

2. For a continuous algorithm, a small error in the inputs should result in a small error in the answer, whereas for a discrete algorithm, even the smallest error in the inputs may result in a different answer. If we understand the algorithm, we can consider the question “how close are we to a boundary”, and possibly act on this information. For example, a university department might adopt (and some do) a rule that if a student fails by 1% or less, the scripts should be checked again. But if all we have is a black box that outputs decisions, we cannot ask this question.

Fixed or Evolving

1. Normally, an algorithm is fixed, like conventional software, or may undergo periodic well-controlled updates, as when the current wind chill algorithm replaced the previous one. There is an additional challenge with so-called “online algorithms” (algorithms which learn from the new data being fed in, also called “active learning”), where today’s foreground data becomes tomorrow’s background data. These can have a disturbing effect of amplifying bias [GoodmanFlaxman2016]. Furthermore, there is no guarantee that an online algorithm will remain valid.

Effect on people

1. It is a truism that every decision has consequences. The main concern is determinative (either de facto or de jure) algorithms making decisions that significantly affect people’s lives. Examples would include medical diagnosis, mortgage approvals and [in the USA] granting or otherwise of bail, and indeed [Zilly’s case] sentencing.

2. However, there are also effects on people’s finances, e.g. gender discrimination in insurance, which in theory is illegal, but can be perpetuated through use of gender-correlated data. There are also effects on people’s careers: if an algorithm decides that a job offering isn’t appropriate for you [Datta et al. 2015] so that you don’t even see it, or if an algorithm decides not to short-list you.

Data

1. An inferred algorithm is based on data. In the case of a scientific experiment, where the algorithm is often called a Law, as in Boyle’s Law, one expects the experiment to be reproduced, and these repeat experiments to confirm the algorithms. In the case of other fields, the nature of the data is more variable. The term “ground truth” (borrowed from meteorology) is used to refer to absolute facts, such as “this skin...
lesion is a melanoma”, or “this person did not offend while on bail”, whereas we use “proxy data” to refer to what is actually used as the training data, “did a dermatologist diagnose skin cancer” or “did a bail judge refuse bail [on the grounds of likelihood of re-offending]”.

2. Clearly ground truth is better, but it may be too difficult, or even impossible, to obtain. The study in [Grin et al. 1990] (which deduces that dermatologists have a 64% accuracy rate, but a sensitivity of over 80%) was very rare in actually conducting biopsies on patients with a negative diagnosis, so that we have ground truth. In many circumstances, it may be impossible, so we can’t obtain “did this person [who was bailed] offend while on bail”, merely “was this person [who was bailed] caught offending while on bail”. This may seem pedantic, but note that we can’t obtain the opposite datum “would this person [who was not bailed] have been caught offending while on bail had he been released”.

3. This use of background data to derive an algorithm has two important consequences. The first is that any biases in the background data will be perpetuated in the algorithm. The second is that we have no guarantee at all that the algorithm produced will remain valid outside the range of those data. This means that every such algorithm needs to be traceable back to the background data that produced it.

Extent

1. The IMA submission (paragraphs 18-20, 33) talks about use in Government.

Intel Corporation.

1. This huge (capitalisation $169 thousand million, but a single factory costs $8 thousand million) technology corporation employs a Chief Mathematician (essentially Chief Algorithms Scientist) at a very senior level. He gave a lecture (reported at http://staff.bath.ac.uk/masjhd/Meetings/AAG-ICIAM15.pdf#chapter.10.1) in which he made several important comments about his use of all sorts of algorithms, conventional and machine learning, which were advisory (but compelling) to the company board.

2. “I am confident telling any group at Intel that using our tools will halve your decision time (I normally get a factor of 5 or 10) and get a 5% better solution (I expect 10%–15%).”

3. “Using my algorithms, we typically end up with a set of equipment in a factory that costs less but actually produces more.”

4. “Executives are irrationally overconfident in their decisions.”

5. Intel is far from alone in its use of these techniques: merely their use is better documented.

Commerce in general

1. Linear Programming, and its generalisation Mixed-Integer Programming, are used, generally as determinative algorithms, in practically every sector [Bixby2015, slide
The specification of the algorithm is public, even though the details of what makes it fast (not what the answer is, which is defined by the specification) are trade secrets. It is also an area in which the algorithmic advances are even greater than hardware advances, to the point where you would be better off running today’s software on 1991 computers than vice versa [Bixby2015, slide 37]. For example, a supermarket chain would use Mixed-Integer Programming to decide how many lorries to send from each warehouse to each (delivery run of) supermarket(s), carrying which goods, to meet the order plan. If it used an independent haulage firm, that firm would use Mixed-Integer Programming to schedule the drivers and vehicles, and so on, and again this would be a determinative algorithm.

2. Equally, advanced machine-learning algorithms will use the weather forecast, along with past shopping data and much else (precisely what is a trade secret of the various supermarkets), to decide what the mix of goods in the supermarket should be, and probably the placement of those goods in the supermarket, along with any special offers. As pointed out in [Grindrod2016], this is the real value of loyalty card data to the supermarkets. These algorithms are sometimes determinative, sometimes advisory.

3. A casual trawl through the Internet will reveal companies (ideal.com was the first authors encountered) boasting about using Artificial Intelligence (and almost certainly Machine Learning) with claims “it’s easy to use artificial intelligence to make faster and better-informed recruiting decisions” and “allows you to eliminate all biases in your recruiting and selection processes”. See also “Amplify” below.

4. Nowhere is commerce using opaque machine-learning algorithms more than in the Internet itself: algorithms determine what search engines return, what advertisements are shown to human beings, what advertisements are shown next to which content (which has caused controversy recently: [BirminghamMail2017 etc.]) and so on. It is probable that none of these activities are intentionally (on the part of the designers of the algorithms) biased, but in practice they are [Datta et al. 2015]. These algorithms are determinative.

Good practice in Algorithmic Decision-Making

1. Firstly, it must be noted that ‘biases or discrimination’ are human perceptions of a process, and what is, or is not, discrimination depends on society, and changes over time, as in the case of gender-based insurance pricing.

2. “Eliminate”. Since the meta-algorithms are generally well-understood, they can be used to eliminate bias, as is proposed for U.S. gerrymandering [Klarreich2017]. They can eliminate human bias in the limited sense that two identical sets of data should produce the same results when fed to an algorithm.

3. “Introduce”. It is certainly possible to introduce biases through an algorithmic process, just as it is possible to do so through a manual process.
a. However, society seems to accept, probably wrongly, feeding more data to an algorithmic process than it would accept in a manual process. In the Zilly case [Angwin et al. 2016], one of the answers considered by the risk scoring algorithm was that to “was one of your parents ever sent to jail or prison”, whereas it’s hard to see a judge accepting a prosecutor’s argument “the defendant deserves a longer prison sentence because one of his parents went to prison”.

b. As pointed out [Kay et al. 2015], online image searches by occupation tend to reflect stereotypes about occupations, as well as a systematic bias against women, so that “CEO” shows up 11% in image searches versus 27% in (U.S.) reality, etc. They also showed that these image search results affect people’s perceptions of reality. Unfortunately, the authors have no explanation for how this happens.

c. Even if direct bias/discrimination (e.g. asking for gender on car insurance) is not allowed, it is perfectly possible to achieve much the same effect (known in the U.S. literature as “indirect discrimination”) through other means, such as asking for occupation as in [McDonald2015]. In the USA, which has a long history of worrying about racial discrimination, such indirect discrimination would be as illegal as the corresponding direct discrimination. There is a large literature, mostly but not exclusively American, on this, e.g. [Miller2015] and [Zliobaite2015]. This material is not well-known, nor is the need for it generally understood, in the nascent UK “Data Science” community, and purchasers need to write it in explicitly.

4. “Amplify”. It’s a fine line between “introduce” and “amplify”, whereas Ba above is certainly introducing bias by referring to CEOs as anything other than 50:50, it is amplifying the bias by not even reflecting reality.

a. [Datta et al. 2015] note that Google searches for jobs are influenced by gender, with higher-paying jobs being shown more often to men. The opacity of the system means that it is hard to know why this is happening (and one assumes that, if there were a trivial fix, Google would have implemented it, but we can see no such rebuttal). It may well be that this is an indirect bias, introduced by the requirements stated by the advertisers, but nevertheless bias it is.

b. Since ideal.com (and its equivalents) claims it “uses AI on your existing resumé database”, it almost certainly breaches other parts of GDPR, but would breach 15.1(h), since the only explanation that could be offered is “our AI process doesn’t think your résumé looks like that of people we’ve successfully hired”. Since existing hiring practices are often biased [BertrandMullainathan2004], it might seem that this process would perpetuate bias. However, due to the statistical phenomenon of “uncertainty bias”, roughly speaking that, if we have a smaller sample, we are less certain of our results, it will actually amplify any existing bias [GoodmanFlaxman2016].
Methods for providing regulatory oversight

1. GDPR Article 15.1(h) provides that, even if “automated processing including profiling” is allowed, then the data subject, “shall have the right to obtain from the controller [...] meaningful information about the logic involved, as well as the significance and the envisaged consequences”. BCS is not qualified to offer an opinion on how regulatory oversight and an appeals process should be provided, but notes that it does need to be provided, presumably through the ICO, hence Recommendation 2.

Recommendations

1. (As in IMA Recommendation 1) The Government, and all its agencies and subcontractors, including recruitment agencies, need to review as a matter of urgency the use of all automatic processing to ensure it
   a. Is only based on legitimate personal data of the individual;
   b. Carries a precise description of any training data it was built from;
   c. Has been tested for indirect discrimination;
   d. Is capable of being explained in line with GDPR 15.1(h). In particular, buzz-phrases like “Artificial Intelligence”, “deep learning algorithm”, “data-based algorithm” should act as warning signs that the “algorithm” is in fact probably no more than unscientific reasoning by generalisation.

2. Strengthen the ICO’s office (numerically and in statistical skills) to handle 15.1(h) appeals.

3. (As in IMA Recommendation 2) The Government should make more use of advanced decision support tools, such as were used (as a novelty) for the QE carrier decision, but which have been used routinely in Intel (and many other corporations) for many years. These tools are not limited to capital purchases: [Li et al. 2015] show a 30% reduction in ambulance response times through use of optimisation algorithms based on actual traffic data. They have potential to improve “joined up thinking” [Smith et al. 2015].

4. No data should be fed to an algorithm, determinative or advisory, that would not be acceptable in an equivalent manual process: see the Zilly case above and [Angwin et al. 2016].

5. In machine learning algorithms, the background data contribute to the decision, so every such algorithm should be prominently labelled with the data that created it, both as a statement the lay person can understand (e.g. “based on London traffic data 1982-2002”) and question (“but that was before the congestion charge”), and ultimately such that an expert can analyse it. This is an algorithmic consequence of the 15.1(h) right. It would also prevent fiascos such as the Velodrome Ambulance case [Guardian2016].

6. Teachers should be aware that the Internet (largely driven by algorithmic processes) may well be biased in its choices, especially of images (see [Kay et al. 2015] above).

7. School and other career advisors should be aware that Internet services are liable to be gender-biased (see Ca above) and use an anonymous window with no gender preferences when conducting searches to minimise the effect.
References


Written evidence submitted by the BCS, The Chartered Institute for IT (ALG0053)


