AI – someone needs to know what's going on!

When developing AI models, we must consider the cause and effect, whether the input-output relationship is meaningful, and is the relationship useful in predicting outcomes based on new data

Data mining seeks to find new knowledge previously 'hidden' in data sets, and a subject matter expert can make sense of what we find while mining data

You may recall, a few years back, a deep internet discussion concerning the lethality of cheese [1]. Maybe not. It was an interesting discussion of a case from a website devoted to finding interesting correlations in disparate data. Fig. 1 shows the chart which generated a lot of interaction: the annual *per capita* consumption of cheese in the USA and the number of annual fatalities from people being entangled in their bedsheets. Does eating more cheese put us at increased risk of bedsheet fatality?

Some of the discussion was purely statistical – is there a third factor which is a common cause? Does the data merit a statistical correlation analysis at all [2]? I happened to present the chart at a couple of events just to promote data analysis discussion and got a generally amused but dismissive response from the audience: "it's just a coincidence". So, I posed a question to the audience: "How do you know what your cheese is doing right at this moment in your refrigerator?" A question I feel is especially reasonable as the refrigerator light goes out once you shut the door. And I didn't stop there… "How do you know cheese isn't plotting your demise *right now*? Plotting with the mayonnaise?" A brief discussion of monitoring/surveillance occasionally follows. The point of the questions is to make us think about statistics and correlations and consider the cause and effect and whether the relationship is *meaningful*. Is the relationship useful in predicting outcomes based on new data? Who can tell? With the cheese, we all likely have some relevant experience and some *a priori* knowledge. In other cases, we may need someone who knows what is going on… a subject matter expert (SME) who can make sense of what we find while mining data.

Data mining seeks to find new knowledge previously 'hidden' in data sets; a CIGRE Technical Paper [3] and subsequent Technical Brochure [4] looked at various types and applications of data mining – what might, these days, be called Machine Learning, ML, or Deep-ML. A transformer dissolved gas analysis (DGA) system allowed for multi-parameter / dimensional dissolved gas data to be analyzed for clusters and then represented the clustered data in a two-dimensional chart while preserving the multi-dimensional distances between those clusters. Interpretation of the meaning of those data clusters requires some knowledge – an SME's input – or application of experiential knowledge as may be captured in standards and guidelines from, say, CIGRE, IEEE or Duval analyses: diagnoses such as overheating, arcing, partial discharge and so on. The trajectory of results for a particular transformer could be plotted over time, yielding an easily interpretable graphical representation of where a transformer may be heading and how soon it may get there (useful for the asset management types who don't necessarily know so much about real transformers).

Using data mining to find clusters in data is one thing, just as finding correlations in data can be achieved with remarkably disparate data – what is needed is an interpretation to give the meaning and usefulness of the findings. And if we look to find new knowledge, we should be aware that the same techniques should also be able to find what we already *think* we know – correlations between hydrogen and acetylene for some transformers, correlations between ethane and carbon monoxide for others, say.

Expectations in applications of artificial intelligence, AI, often exceed the actual performance. Some well-known cases include bias, poorly representative training data, and, in the final case, some basic common sense:

- Amazon's automated resume analysis rejected those which mentioned women or featured certain all-women colleges [5],
- Volvo's self-driving car was good at identifying moose but struggled with kangaroos [6],
- an automated traffic surveillance monitor sent a penalty notice when it mistook a pedestrian in a sweater for a car with a specific license plate (it didn't check for the license plate actually being on a car…) [7].

But should we be surprised? Cognitive biases afflict all *humans* meaning that we look at the world through the filters

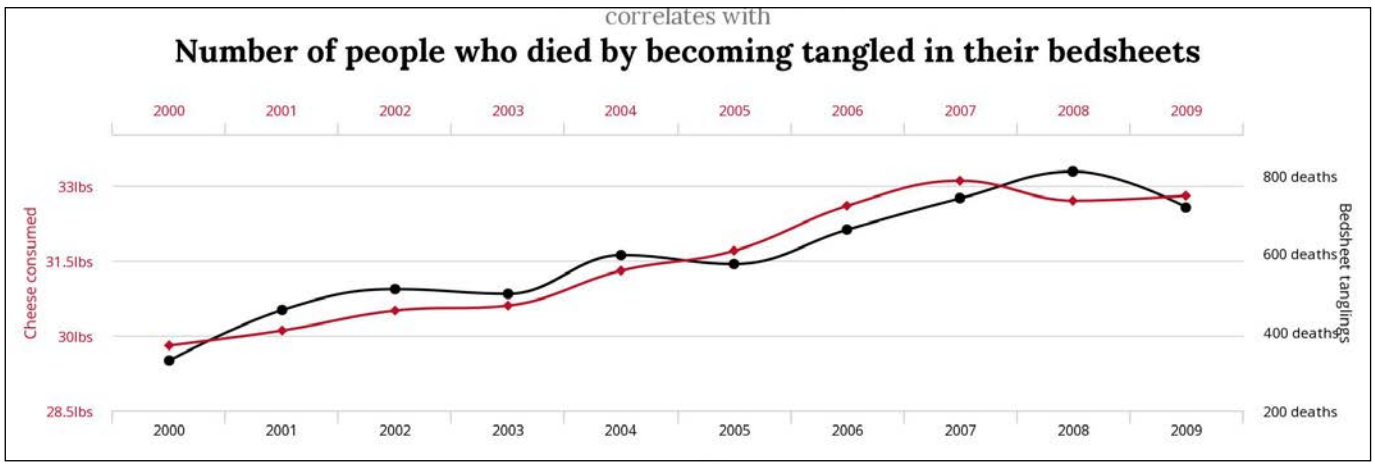

Figure 1. A correlation between Cheese Consumption and Bedsheet Fatalities

of our experience, memories, and our own brain's efforts to simplify the world around us: confirmation bias, hindsight bias, inattentional blindness and many more are well known and well documented [8]. Why, then, are we surprised to see similar biases affecting AI systems? And ultimately, if we make a decision based on AI, we need to know why that decision is made and that it is not subject to bias – what leads us from the data through analysis to a decision is an "explainable AI": a "glass box" not a "black box"!

Chomsky quote [9], "There is a notion of success ... which I think is novel in the history of science. It interprets success as approximating unanalyzed data."

In a previous article, we discussed the partial discharge (PD) classification system and its capability not only to identify individual PD sources from phase-resolved (PRPD) data but also to identify data which doesn't look like the initial training data: an "out-of-distribution" (OOD) analysis [10]. As a parallel, consider an ML tool which is trained to distinguish between cats and dogs based on a set of pictures supplied for the purpose. When we show the tool a picture it hasn't seen before, it can classify it as a cat or a dog… but first, we need to check that it looks at least *somewhat* like the training data. Show the tool a picture of an alligator, and we would hope it would say that this is not at all familiar: an OOD calculation helps identify the 'degree of membership' in a quantifiable way before we deem it a cat or dog.

We look at the world through the filters of our experiences, memories, and our brain's efforts to simplify the world around us, so we shouldn't be surprised to see similar biases affecting AI systems

A subsequent technical paper looks at real-world situations where there may be multiple PD sources operating simultaneously [11]. Multiple PD sources overlap in the PRPD patterns and may cause confusion in classification: imagine using the ML tool developed for cat/dog classification to look at a picture with multiple animals, some of which are a weird hybrid of cats and dogs… there could easily be some confusion. But the classification of multiple sources simultaneously is a common problem in the "real world". Standard statistical approaches may be useful for separating out the data for each source but still need an SME to identify the sources individually. There are standard data sets to test ML tools where, for example, ML tools to identify numerals are given examples of each type, but also overlaid data where two numerals are in one image [12]; this is a useful 'test' as the data is standard and the results from many trials have been published.

AI systems seem to perform best when they are targeted and have specific functions, but we do need to check for OOD, and we may need to have SME and other inputs to give the results meaning and validity and confirm that they are "explainable". Some years ago, as a high school teacher, I came across a relevant educational concept: don't just teach students "facts" but teach them how to *question* what they have learned. That would be a laudable achievement not just for human education but also for any AI system, leading to students becoming what the authors called "crap detectors" rather than just passive recipients [13].

Bibliography

[1] http://www.tylervigen.com/view_correlation.php?id=7

[2] https://errorstatistics.com/2015/05/04/spurious-correlations-death-by-getting-tangled-in-bedsheets-and-the-consumption-of-cheese-aris-spanos/

[3] T. McGrail et al., *Datamining techniques to assess the condition of HV electrical plant*, Paper 15–107 CIGRE, Paris, 2002

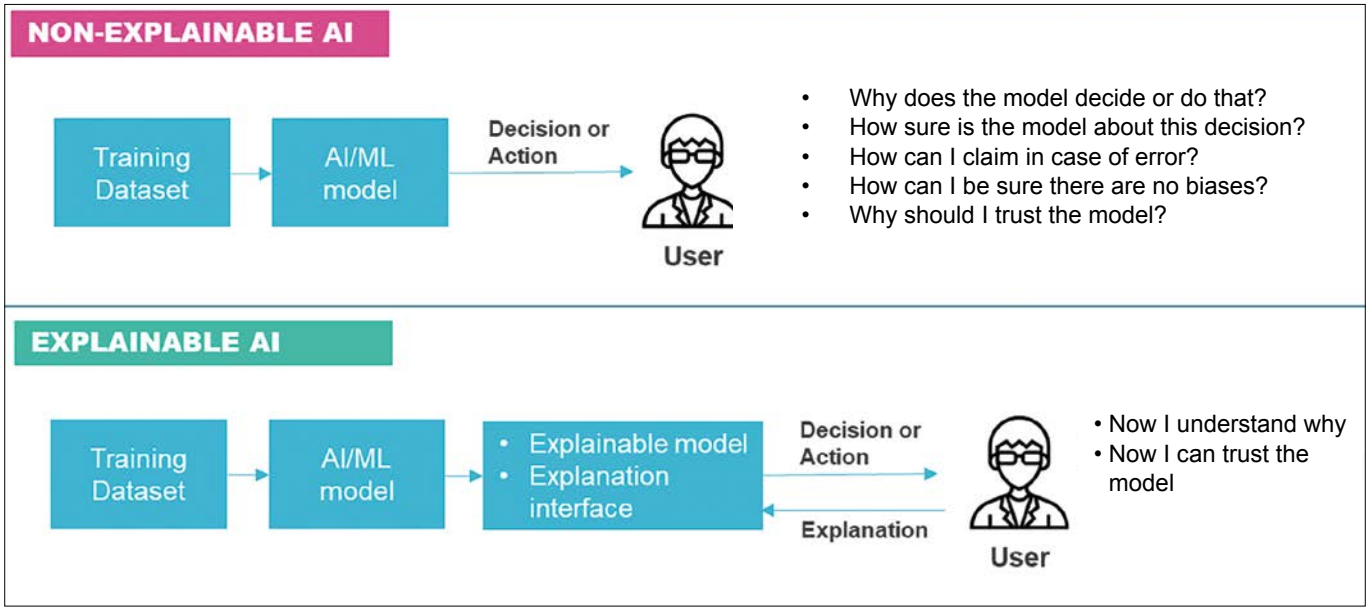

Figure 2. Taken from: https://worldline.com/en/home/knowledgehub/blog/2021/january/ever-heard-of-the-ai-black-box-problem.html

[4] CIGRE Technical Brochure 292, *Datamining techniques and applications in the power transmission field*

[5] https://www.geeksforgeeks.org/5-algorithms-that-demonstrate-artificial-intelligence-bias

[6] https://www.theguardian.com/technology/2017/jul/01/volvo-admits-its-self-driving-cars-are-confused-by-kangaroos?

[7] https://www.autoblog.com/2021/10/19/cctv-knitter-kn19ter-ticket-england/

[8] https://www.simplypsychology.org/cognitive-bias.html

[9] https://norvig.com/chomsky.html

[10] T. Rhodes, T. McGrail, *Successful application of AI techniques*, Transformers Magazine Special Edition – Digitalization, Nov 2020

[11] Dr. I. Mitiche et al., *Unsupervised source separation for multi-label classification*, EUSIPCO 2022

[12] *The MNIST database of handwritten digit images for machine learning research*, IEEE Signal Processing Magazine, vol. 29, no. 6,

[13] N. Postman, C. Weubgarter, *Teaching as a subversive activity*, 1971, ISBN 13: 9780440085621

Authors

Philip Boreham is the Director of Engineering for Doble Engineering's Innovation Centre for Online Systems (ICOS) in the UK. Work in the power industry designing test environments for fault simulation in generation equipment led to the next 20 years developing innovative solutions for distributed online condition monitoring applications with a focus on condition-based maintenance. Philip is motivated by the discovery of new techniques in failure detection and condition assessment of power systems, supported by a degree in Mechanical Engineering and membership of the IET and IEEE.

Dr. Tony McGrail is the Doble Engineering Solutions Director for Asset Management and Monitoring Technology. He has several years experience as a utility substation technical specialist in the UK, focusing on power transformer test and assessment, and as a T&D substation asset manager in the USA, focusing on system reliability and performance. Tony is a Fellow of the IET and a Member of IEEE, CIGRE, IAM and ASTM. He has degree in Physics with a subsequent Ph.D. in Applications of AI to Insulation Assessment.

Dr. Imene Mitiche is a research fellow at Glasgow Caledonian University, working in collaboration with Doble Engineering on data analytics using signal processing and machine learning for the next generation asset monitoring instruments. She obtained a BSc. in Computer and Electronic Systems Engineering (Software development) with First Class Honours and a MSc. in Telecommunications Engineering with Distinction from Glasgow Caledonian University. She then pursued and obtained her PhD in Machine Learning Application for High Voltage Condition monitoring. Imene is a member of IEEE and IET. Her other interests are data science, machine learning operations and edge-cloud implementation.