

**EVIDENCE GIVEN TO THE ALL PARTY PARLIAMENTARY GROUP  
FOR ARTIFICIAL INTELLIGENCE:  
EVIDENCE MEETING 1 – DATA – 22 JANUARY 2018**

My name is Tim Pullan, founder & CEO of ThoughtRiver a leading legal technology that uses artificial intelligence to automatically risk profile contracts. Prior to that I ran a data analytics business for Experian in Asia Pacific. By profession I am a lawyer with over 20 years' experience in data privacy advising predominantly large multinationals on data issues in various parts of the world.

I have 3 comments I wish to make:

- The generation of data-driven knowledge should be encouraged as a driver of both our economy and our liberal enlightened society. It is both a public and private good.
- Citizens need to be protected against the purposes for which data-driven knowledge can be used. Data protection law does not provide effective protection.
- Encouraging the growth of data driven innovation whilst protecting citizens is not incompatible. But it does require radical thinking.

### **Data-driven knowledge**

AI is the next wave of industrial revolution. Done well, this will humanise our lives by reducing repetitive and low value work. It will accelerate the speed with which we find cures to illnesses.

AI is not new. Back in the eighties and nineties Experian and others were using huge datasets and AI techniques to identify credit profiles in the consumer population. It may have been the original 'computer says no' solution. But by semi-automating credit evaluation, it also played a key role in expanding the availability of credit to the broader population.

With my own business ThoughtRiver, we use AI technology to identify meaning in complex contract texts. In doing so, we massively reduce the amount that lawyers need to read.

In both cases the role of AI is to convert data into information at scale. This typically requires large amounts of high quality data.

Since the eighties AI tools have become vastly more powerful. But when it comes to data, the fundamentals have not changed. To produce high quality Intelligence you need high quality, accurately categorised data. If I tell the machine these are examples of blue cars, it must be so otherwise the machine will learn badly. So much of the data out there is not usable by these criteria.

Experian revolutionised credit evaluation by recognising that data categorisation was key but takes a lot of

work. The answer was collaborative data pools for credit providers. At ThoughtRiver, we too recognise that the future of AI in the legal industry rests on collaborative categorisation. The same is certainly true for other industries.

If we fail to do this we will find ourselves in a state of data poverty - unable to compete with key datasets controlled by foreign corporate and national interests.

If the UK is to seize the opportunities of this new industrial wave, I believe Government has a role to play in sponsoring and creating the enabling regulatory environment for industry-level data trust schemes that make datasets widely available to innovators and encourage wide scale collaborative categorisation.

### **Protection of citizens**

Just as we should tear any barriers to our quest for knowledge, we should not as a society be squeamish about prohibiting specific uses of that knowledge in a paternalistic way. It is not incompatible to say that we want to know whether young male drivers are more likely than young female drivers to have accidents, whilst prohibiting insurers from using that information to discriminate when pricing policies.

As corporate datasets have grown so we have seen the use of AI in marketing to exploit consumers in new, far more intrusive ways. What marketers refer to as 'customer engagement' strategies through channels like social media have now evolved into sophisticated and potentially insidious forms of mind control which policymakers should consider regulating with regard to specific groups e.g. prohibiting use of targeting algorithms for children, gamblers etc.

It is odd that this intrusiveness phenomenon has arisen during the same 17 years in which privacy legislation has massively expanded. Something is definitely not working.

For the consumer the reason is its reliance on consent as the cornerstone protection. In practice this does not work - research has shown that it would take 70 days to read all the consent notices we see during a typical year. We click through. It is only when we get hounded that we care, and then withdrawing consent across possibly hundreds of companies is not easy. People rightly feel they don't control their data because it is sprayed across thousands of corporate databases.

From a business perspective, absolute compliance is sometimes impossible. The legal presumption that each item of personal data can be duly considered and individually tagged for specific action simply does not reflect corporate realities. Personal data is sprayed across databases, messaging systems, devices, back-ups, back-ups of back-up. A corporate might pass an audit but it will nevertheless have instances of non-compliance all of which are potential liabilities. With its harsher standards GDPR makes the disparity between requirement and reality even more stark. We may be leaving the EU but the truth is GDPR will remain the effective legal standard. The size of regulatory penalties for non compliance which will in future be levied by regulatory bodies over which we have little control should be considered a major future economic risk. Furthermore, now that the UK has lost its influence over EU data protection law we don't have the realistic option of being able to amend the law to fix any issues in the foreseeable future.

## **A radical data governance solution**

There needs to be a focus on practical as well as regulatory steps to ensure effective protection for consumers.

Businesses often ask how they are going to cope with GDPR. I say easy, don't process any personal data and it doesn't apply to you. In the area of consumer intelligence, UK is leading the way in collaborative categorisation that does not use personal data - for example, homegrown companies like InfoSum enable anonymous collaboration vital to sharing in sensitive areas like healthcare.

But this being a big subject, I would like to be bold propose a radical and practical data governance programme for government.

It lies in work first done by Doc Searle at Harvard in the late nineties. Doc Searle proposed that instead of having to throw our data into a corporate data swamp every time we buy a service each of us should genuinely control our own data via personal data stores. If a company wants information about us, they request temporary access to it from our personal data store. If we want to turn off all marketing from anyone for a couple of weeks, a flick of a switch is all it takes. If we want our internet search history only used for profiling for as long as we agree, it is kept in our store. The personal data store would be able to verify identity to corporates where required and agreed.

To make this work, it would be illegal for any online or offline service to refuse to accept personal data stores when signing up new customers. It would also be illegal for them to take away copies of any data on temporary licence. But whilst they have access to it via the personal data store, effective consent is assured.

The insight is this: generally speaking consumer marketers don't care who you are. They care what you are. They only need to know who you are because of regulation or because they need a way to join up all the disparate pieces of information they collect about you. Data protection law focusses on data that identifies who we are, not what we are. So by enabling commercial organisations to have a relationship with you, but minimising the burden of personal data processing we find mutual ground.

I have seen many businesses try and fail to launch personal data stores. In my view to be successful this needs Government sponsorship but actual ownership by an independent trust and I recommend it as a way of making data protection tangible and meaningful to the average consumer.

## **Conclusion**

We need to remove any regulatory barriers to the pursuit of knowledge via AI, including the use of data for these purposes. But at the same time we should be ready and willing to regulate specific harmful uses of that knowledge in the UK. Finally, to make data protection properly meaningful to citizens we need to make it real and tangible in the form of functionality that can be used in their day to day lives.

**Tim Pullan**  
**22 January 2018**

