



Brief for oral evidence to the House of Lords Select Committee on Artificial Intelligence

ORG's Policy Director, Javier Ruiz, gave evidence to the House of Lords Select Committee on Artificial Intelligence. These are his briefing notes, making the case for truly open IP in AI and against the concept of data as property, and asking for better public procurement of AI in the public sector.

1. Who should own data and why?

Firstly we need to establish that pure data as such cannot be owned in the UK as property, as established by the Court of Appeal in 2014 (Datateam).

We have a complex system that covers intellectual property, rights, contracts, confidentiality, and personal data.

There is a critical problem here: It is not just AI providers who create the new insights or value. You need the data as well as software, indeed the distinction becomes blurred.

Data is as important or more as the specific software. AI has only advanced because nowadays there is a lot of available data to train systems.

Public sector bodies and the NHS are in a more powerful position than they realise and need to become more assertive vis a vis technology companies.

Public procurement rules for AI need to be adapted to include the role of data in services and the capture of benefits.

We want to keep value for the public and access to knowledge . It is now common to have policies mandating:

- open access publishing in public funded research.
- open source in public funded software .
- open access digital archives

But what is the value in AI processes and how is it captured? In most similar cases when we talk about value we mean intellectual property.

Establishing fairness – or a public benefit - over the distribution of the goods coming from AI implies a clear framework for IP.

This is particularly important when we deal with public sector data and non-commercial societal benefits.

If we take the example of DeepMind and the NHS, not to pick on them but as a well known case where information is available.

DeepMind claims ownership of any “Developed IP” developed in their dealing with the NHS, according to contracts available online.

But what specific IP is created if any?

Google claims over: Designs, works, inventions, software, data, techniques, algorithms, know-how or other materials.

In many cases the actual outcome is hard to define. AI can produce **insights**. How are these captured and embodied needs a lot more clarity. It is far from clear that any output of a computerised process can be claimed as intellectual property.

An engineer in a leading AI company told us they cannot fully share what the machine learning system learns without fully sharing the full model, and are worried that complete transparency would mean having to open up their systems completely.

This convergence of machine learning models and the insights they generate are, data and code, is a hard problem, but the status quo of having AI companies getting all the insights is not acceptable.

If you cannot even separate what the model learns from then model itself we need some proper work to understand what these claims are really about.

So far DeepMind publishes its **insights** in the form of research papers, which are openly available, but we are not sure what will happen in the future.

DeepMind is helping improve NHS algorithms. In the best known case the standardised Acute Kidney Injury algorithm helps hospitals provide consistent quality decision making. The AKI algorithm was build through a process of consensus of various expert committees. It could be very difficult to establish intellectual property ownership and restrict its use.

Now with AI this may change. New versions of such algorithms can be created by computers on the basis of patient data and NHS resources. DeepMind would claim all IP over such algorithms but this is problematic.

Computers are responsible for generating the “Developed IP” in ways that challenge existing intellectual property arrangements and the concept of the author or creator

Under the UK copyright act:

“work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken” (s. 9(3) CDPA)

This approach is based on the idea that the user or in some cases the programmer are the only actors that are involved in some form of creative process, while the commissioner of the work or the owner of the computer are not.

This is true, but this simple approach may not reflect the generation of new copyright works through artificial intelligence. Both the code and the training data are required to create new

“works”. The more sophisticated the AI, the more it learns from data and not from its original programmer.

Copyright in AI works created autonomously need to incorporate the role of the data as part of the “arrangements necessary”.

There are similar issues with patents and the role of the inventor.

In some cases it may even be that the output of the process is data in a form that would not be recognised as a “work” under copyright. Other intellectual property rights could apply

The issue of data is also relevant for individuals in other settings, not just the NHS or public sector bodies.

Once we establish that an AI provider should not retain automatically exclusive ownership of all IP generated with someone else’s data, there are still practical issues.

We may still want to provide some protection to the investments in AI, while promoting disclosure and wider access (objectives of most IP law, but also critical in public sector procurement).

For accountability purposes this is an issue as well.

Most ethical frameworks or principles on AI demand a degree of transparency and explainability. This is hard in AI. It may well be that we can only fully trust open source software that can be fully inspected, as it is already established practice in the field of security. This is notwithstanding the inherent limitations of certain types of autonomous machine learning for transparency.

We need to find ways to open models for inspection while protecting the IP of the developers. Maybe you could have proprietary software made available under certain conditions without giving the right to reuse it.

In addition, Testing the replication of results could mean making data available as well as algorithms.

There could be some temporary data exclusivity arrangements, similar to the field of biological medicines.

Pharmaceutical companies have special arrangements over trial data for certain medicines, which is made available to check but remains restricted for commercial exploitation for a period of time.

Finally a note on contracts. Companies are not making arrangements that assign most commercial benefits to the developer of AI. This can bypass IP law.

Public procurement contracts involving AI need to incorporate considerations of IP that are not as one sided as the current contracts of DeepMind.

2 Do you think the Government's AI review addresses the key issues relating to data and AI? Does it go far enough?

The review is focused on growing the AI sector and is not sufficiently dedicated to ensuring that growth provides a broad societal benefit, despite some general rhetoric.

The report recognises that access to data is critical and that there are problems, but these are framed as obstacles to overcome rather than genuine concerns that may want us to STOP specific programmes of AI.

We broadly agree with some recommendations on improving Access to data, but we believe that the concept of Data Trusts needs much more discussion.

Data Trusts would need to address the distribution of benefits, public interest, etc. Not just as an afterthought around commercial exploitation of results, as it is presented in the review.

Creating frameworks can simplify data access but individuals need to understand each instance of processing, who is involved and for what purpose, not just the generic framework.

The interests of individuals whose data is being processed are not reflected in the proposal. The assumption is that a vague notion of progress and the advancement of science is enough.

We have no direct comments on the recommendations on skills, but in any case this should include broad multidisciplinary skills, such as ethics, and not just computer science.

The recommendation to form a framework jointly by the ICO and the Alan Turing institute is problematic because the remit of the ICO is very narrow and focused on data protection. The issues raised by AI may well go beyond strict data protection and touch on broader privacy as a human right, discrimination, competition or financial abuse. We need a much broader approach.

The other recommendations in the review to align bodies and stakeholders completely miss giving any voice to consumers and citizens via civil society organisations.

3. Is the Data Protection Bill, as currently being considered in Parliament, fit for future challenges? Is the GDPR?

GDPR provides very limited rights to explanations and objection to automated profiling.

The DPB requires numerous improvements, but in this context there is one area that deserves attentions:

1. Article 173 of the Bill gives non-profits the rights in article 80 of GDPR to support data subjects, but GDPR also has optional powers for NGOs to take on cases without the need to be instructed. This is particularly important in complex data environments such as AI, where people may not know they are affected.

4. Do you believe we need a specific AI watchdog or regulator to deal with the issues surrounding data and privacy?

AI issues are specific to domains and sectors. It would be very hard to avoid overlaps, and hence leaving gaps as well.

For data privacy the ICO is already the regulator and there is no need to create a new one.

Other regulators may need to incorporate AI in their work plans.

However, It would be helpful to have a trusted independent body to provide expert support to established regulators, with the power to issue recommendations and guidance and also investigate on requests from the public. If they cannot get enough information investigations should be triggered under specific areas: privacy, competition, discrimination, finance, etc.

The IPO should start some work on the impact of AI in copyright and patents.

The Equality and Human Rights Commission needs to understand how discrimination operates in machine learning.

Public procurement and public-private partnerships based on AI should be exemplary and set the tone, and procurement oversight bodies

5. How can data be managed so that it is used for the public good?

There are many ethical frameworks out there. Everyone claims to be working for the public good. The public good needs to be build into processes by looking in detail at the wider impacts and the long terms consequences of specific activities, not just the immediate results or vague promises.

6. What technical approaches might help to preserve privacy while also ensuring the benefits of AI are realised?

PI to cover.
Discussion on

7. How can we mitigate against unintended prejudices in the data that is used for training AI models?

PI to cover
One additional point. We cannot use AI to fix social problems and avoid making human choices. Even without bias we may disagree with a computer decision.

8. If there was one recommendation you would like to see the committee make at the end of this inquiry, what would it be?

Public procurement guidelines of AI systems to ensure public benefit.