



# Master of the Rolls

THE RIGHT HON. SIR GEOFFREY VOS

Irish Law Society Industry Event  
Wednesday 09 October 2024

## AI and the GDPR

### Introduction

1. The EU's Artificial Intelligence Act<sup>1</sup> began entering into force on 1 August 2024. It will take some 2-3 years to come fully into operation.
2. The Council of Europe's Treaty<sup>2</sup> on AI and human rights, democracy and the rule of law was opened for signature on 5 September 2024 by European Justice Ministers in Vilnius. The Council of Europe's Treaty was billed as the first-ever international legally binding treaty aimed at ensuring that the use of AI systems is consistent with human rights, democracy and the rule of law. The Treaty has already been signed by the EU, the UK and the USA, amongst others.
3. I want to consider this evening how these developments, alongside the General Data Protection Regulation<sup>3</sup> (the GDPR), are likely to affect the adoption and development of AI processes in general and automated decision making in particular.
4. It is interesting, I think, to notice how, in the field of AI, the distinction between private law and regulation has already become blurred. That is something that has happened also in relation to digital assets and digital trading, where many countries have regulated and, in the course of so doing, have introduced changes to their private law by a sidewind. This is something that I regard with a little scepticism, but, in relation to digital assets and digital trading, will have to be left for another lecture.
5. As technology advances, it is important, I think, not to impede its beneficial adoption by premature regulation, before the dangers posed by those technologies are clearly understood. I am not saying that that has happened, but it is something to which the EU and its member states and other countries should be alive. This point was made specifically in relation to the EU's AI Act in Mario Draghi's report last month on the future of European competitiveness.<sup>4</sup> He said expressly that:

---

<sup>1</sup> Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence.

<sup>2</sup> The Council of Europe's Framework Convention on artificial intelligence and human rights, democracy, and the rule of law ([CETS No. 225](#)).

<sup>3</sup> Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.

<sup>4</sup> [https://commission.europa.eu/document/download/97e481fd-2dc3-412d-be4c-f152a8232961\\_en](https://commission.europa.eu/document/download/97e481fd-2dc3-412d-be4c-f152a8232961_en).

“the EU’s regulatory stance towards tech companies **hampers innovation**: the EU now has around 100 tech-focused laws and over 270 regulators active in digital networks across all Member States. Many EU laws take a precautionary approach ... For example, the AI Act imposes additional regulatory requirements on general purpose AI models that exceed a pre-defined threshold of computational power – a threshold which some state-of-the-art models already exceed” (emphasis added).<sup>5</sup>

6. It is also important to ensure that impediments are not placed in the way of technologies that facilitate international commerce. Inadvertent changes to municipal laws that are widely used in international trade can create problems for frictionless technology-assisted trade, particularly where such changes do not align with each other and with internationally applicable regulatory regimes.
7. And before turning to the two specific, and perhaps somewhat technical, problems that I want to address tonight, I should say that automated decision making is very far from a minority sport. I read only last week in the Financial Times that OpenAI is betting that AI-powered assistants will hit the mainstream by next year. Reasoning AI agents that can complete complex tasks on behalf of consumers and people generally, and also, of course, businesses are reported to be the “newest front in the battle between tech companies”. So, the legal points I shall be discussing really do matter.

#### Two serious problems for businesses and individuals using AI

8. With that introduction, let me explain the two serious problems that I want to talk about. They very much concern industry and business, the tech community and the lawyers to the tech community.
9. The first problem relates to the use of automated decision-making affecting people’s individual rights by Governments, global corporations, like Apple and Google, and even SMEs. It is, in a nutshell, whether the use of such AI is prohibited by article 22 of the GDPR.
10. The second problem concerns the data used to train AI tools. The question here is whether the owners of that data retain any residual rights once the machine has been trained on data that has been placed in the public domain. There is also the subsidiary, but important, issue of whether the work product, whether text or computer code, produced by these AI tools are also holding and utilising copyright or copyleft material.
11. I will end with a brief look into the future.

#### The article 22 problem

12. Article 22 of the GDPR (and of the UK GDPR) provides, as you all know, that: “The data subject shall have the right not to be subject to a decision based solely on automated

---

<sup>5</sup> See page 26 of the report.

processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her”.

13. This article, as it seems to me, has very wide potential effect, and yet it has only recently been the subject of its first major CJEU decision in the *SCHUFA Holding* case<sup>6</sup> (*SCHUFA Holding*).
14. The question, in a nutshell, is whether article 22 **prohibits** decision-making within its ambit or provides just another right for the data subject to enforce in the event of its violation.
15. It also obviously has repercussions if AI or automated decision-making were to be used in our judicial processes in the future. If AI were ever to be used in judicial decision-making, an automated decision could arguably not be effective.
16. It can probably be accepted without much debate that the wording and application of article 22 is quite complex.
17. The *SCHUFA Holding* case held that a credit reference agency was making an automated decision affecting an individual’s legal rights, when it created an AI driven credit repayment probability score about an individual. Lenders, in fact, rely heavily on these scores in deciding whether to lend or not to lend, amongst other things. The CJEU thought that the obligation to comply with article 22 fell on the credit reference agency, not just on the lender.
18. This decision does not, of course, answer the question I have posed, but it does perhaps give us some pointers.
19. Paragraphs 42-44 of the Court’s decision make clear that article 22 gives data subjects the right not to have solely automated decisions made about them or significantly affecting them, if they result in legal consequences for them. The scope of the decisions in question are to be construed broadly as the outcome of the case itself demonstrates.
20. Paragraphs 52-3 of the judgment do not finally resolve the question I have posed when it says that article 22(1): “lays down a prohibition in principle, the infringement of which does not need to be invoked individually by such a person”. Such decisions are, therefore, only authorised when they fall within one of the exceptions mentioned in article 22(2), namely where: (a) the decision is **necessary** for entering into, or performance of, a contract between the data subject and a data controller, (b) it is authorised by EU or national law, and (c) where it is based on the data subject’s explicit consent.
21. On one analysis, both article 22 and the *SCHUFA Holding* decision are unsatisfactory since they inevitably make the legality of an automated decision-making process dependent on **the purposes for which the outcome of that process is used**. The words of article 22 are “which produces legal effects”, so the legality of the decision depends

---

<sup>6</sup> Joined Cases C-26/22 and C-64/22 of 7 December 2023.

on the use made of its product or outcome. Arguably, one cannot know in advance whether a specific automated decision is lawful or not.<sup>7</sup>

22. The decision that the automated creation of a credit repayment probability score about an individual constitutes **a decision** makes article 22 of wide-ranging application. Moreover, the EU's AI Act has denominated a wide range of AI systems as "High Risk AI systems".<sup>8</sup> These include "AI systems intended to be used to evaluate the creditworthiness of natural persons or to establish their credit score", law enforcement AI systems and those intended to be used in the administration of justice.
23. All in all, it is looking more and more as if the provision in article 22 of the GDPR will be given a prohibitive interpretation beyond the immediate context of the GDPR. It looks as if the CJEU at least may favour an interpretation that effectively prevents Governments, global corporations, and SMEs utilising automated decision-making wherever it would affect people's individual rights, unless the process is specifically authorised by statute or consented to. There are likely to be many instances where these entities are already using AI in this way or in a way that comes very close to the article 22 situation.
24. We may, I suppose, end up with a situation in which local authorities, Amazon and government pension authorities ask customers to consent to automated decision-making every time they contact you, just as we are asked 20 times a day to consent to cookies or additional cookies. We shall see.

#### AI training data

25. It is axiomatic that large language models operate by sifting through vast tracts of data to find the next most likely word to answer the searcher's prompt or inquiry. It is, therefore, equally axiomatic that restricting the LLM's access to data will restrict its ability to respond ever more accurately and reliably to such inquiries.
26. Yet despite the fact that LLMs have now been widely and publicly available since the launch of GPT4 in March 2023, there have been few legal challenges to the use of the data on which all LLMs depend. I shall, though, mention in a moment one that is likely soon to come to trial in London. Despite what is often said, it is possible, at least in principle, to identify the data on which LLMs have been trained. The producers of LLMs describe the problems of copyright and perhaps also copyleft infringement as complex problems about which, presumably, we should not need to worry.
27. I suspect that many of us ordinary mortals, in 2024, expect that, once we have committed our data to social media or to the worldwide web or into any other public space, it is fair game for the producers of LLMs to access it and to use it in training their machines. You may expect to see that data regurgitated in some form somewhere sometime, as soon as you knowingly allow it to be released into what we used to call the public domain. Whether one's act of release into the public domain constitutes

---

<sup>7</sup> I am indebted to Matthew Lavy KC for making this point to me.

<sup>8</sup> Article 6(2) and paragraph 9(a) of annex III to the EU's AI Act makes many AI systems and specifically those intended to be used in the administration of justice into "High Risk AI systems".

consent within the meaning of articles 6 and 7 of the GDPR is a far more tricky point. On its face, however, it would seem that it does not. Apart from anything else, it is hard to see how consent can be withdrawn, as is required by article 7.3, once LLMs have actually been trained on that data.

28. Despite what I have said, the problem is very acute. It is not so because of the likelihood that an individual would sue OpenAI or Microsoft for using text we have asked to be summarised by an LLM – though a class action might, I suppose, be brought along those lines.
29. It is acute because large language models are being increasingly used by developers of software across the globe to write their code. The LLMs are, of course, only able to write code so brilliantly because they are trained on huge tracts of coding source material. Make no mistake that it is said that LLMs can write code far more accurately, and of course far more quickly, than any human programmers.
30. We are all familiar with the problem of musicians worrying that AI will write better songs than they can. There are already numerous extraordinary publicly available song-writing apps, like *Suno* and many others, that provide the most startling results. The musicians' worry is, of course, that the apps are trained on copyrighted material that is publicly available from platforms like Spotify and Apple Music.
31. The case I mentioned a moment ago is being brought by Getty Images against Stability AI, which has produced an AI image creation programme. Getty says that Stability AI has infringed its intellectual property rights by using its image libraries to train its Stable Diffusion programme, and by reproducing the outputs from that programme that allegedly reproduce part of Getty's copyright works.
32. But to return to the code problem, even if the source material is the subject of "copyleft" rather than copyright, there are serious repercussions when training materials are protected. Copyleft, as you will all know, is the practice of allowing protected works to be used on the internet on condition that the same rights are preserved in the work product. Where the original source code carries a license fee, the work product would be subject to the same fee.
33. Once LLMs have been used to generate code, the users are likely to embed that code in their products without regard to the licensing rights of the original source material on which the AI model was trained.
34. This problem is likely, I guess, to be the subject of significant litigation in the future, even if some might think that open source copyleft material owners have only themselves to blame if their code data is used to create ever cleverer programmes without their being paid a fee. But once again, we shall see.

### Conclusions

35. So, let me return to the thesis that I mentioned in my introduction. The two AI "problems" that I have highlighted this evening, are problems created in part at least by regulation getting ahead of private law. Domestic legislation can create exceptions

to the article 22 problem, and allow automated decision-making about individuals in defined circumstances. I am not aware that it has yet done so in any of the jurisdictions likely to be represented here tonight.

36. We all need to be careful not to impede the development and adoption of new technologies, whilst also being astute to ensure that people's basic human rights are not infringed by the new processes. The European Law Institute's annual conference is being held here in Dublin for the rest of this week. The question I have identified will be the focus of our discussions. We are fortunate to have many distinguished and expert speakers with us to discuss these problems.
37. Let me add two final footnotes. Article 6(2) and paragraph 9(a) of annex III to the EU's AI Act makes, as I have said, AI systems concerned with the administration of justice into "High Risk AI systems". The definition is as follows: "AI systems intended to be used by a judicial authority or on their behalf to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts". It may be thought that that definition would encompass almost any such system when applied to justice. We must, as I say, be careful not to stop innovation dead in its tracks, even if the world is not yet ready for automated judicial decisions, unless they are supervised by a human judge.
38. You may think that the two problems I have spoken about are nerdy and over-technical. The problem is that future generations will, I think, speaking for myself anyway, want to make use of AI, LLMs and automated decision-making to improve their everyday lives. The lawyers and legal systems perhaps owe it to consumers, present and future, that we protect them from real cyber-abuses without preventing or hampering innovation.
39. I would be happy to answer any questions you may have.