
Using Artificial Intelligence to evaluate privacy policies

“Claudette meets GDPR” Project

Q&A

1. What is the aim of the “Claudette meets GDPR” project?

“Claudette¹ meets GDPR” is a project conducted by researchers from the European University Institute (EUI) in Florence², with the support of BEUC.

The main aim of the project is to test whether the legal evaluation of privacy policies of online services under the General Data Protection Regulation (GDPR) can be automated using artificial intelligence (machine learning). For this purpose, the researchers analysed the privacy policies of 14 companies and used this analysis to train an artificial intelligence tool (“Claudette”) to automatically assess whether a privacy policy complies with the GDPR.

2. Which privacy policies were analysed? Why these?

The team of researchers analysed the privacy policies of Google, Facebook (and Instagram), Amazon, Apple, Microsoft, WhatsApp, Twitter, Uber, AirBnB, Booking, Skyscanner, Netflix, Steam and Epic Games.

These companies were chosen as they are among the most used services in various sectors (e.g. social networks and messaging, travel, transport and accommodation services, entertainment, etc.) but also because they are among the biggest online players and should therefore be setting a good example for the market to follow.

3. How to assess privacy policies under the GDPR? What would a perfect policy look like (what’s the “golden standard”)?

According to the GDPR (art. 12-14), consumers should be given information about data processing – among others, the identity of the controller, legal basis and purpose of processing, recipients of personal data, right to rectify and complain – using plain and intelligible language. This means that a GDPR-compliant privacy policy should be **comprehensive** (including all the required information) and **comprehensible** (using clear and understandable language). Furthermore, the types of envisioned processing should be **lawful**, i.e. complying with GDPR’s rules and principles.

The [report](#) contains a detailed explanation of the “golden standard” (sections 2.1, 3 and 4). Everyone – journalists, civil society, consumers themselves – can use it to assess the (likelihood of) compliance of privacy policies of other platforms and services.

4. What are the results of the evaluation?

The results of the analysis – which happened in June 2018 – indicate that, about one month after the entry into application of the GDPR, there is a significant need for improvement. None of the analysed privacy policies fully meets the requirements established by the GDPR.

¹ “Automated CLAUse DETeCTER”, <http://claudette.eui.eu>.

² A team of researchers from the European University Institute in Florence, University of Bologna and University of Modena and Reggio Emilia, led by Professors Hans-W. Micklitz and Giovanni Sartor

The 14 privacy policies comprised 3,658 sentences (80,398 words) in total. 401 sentences (11.0%) were marked as containing unclear language. 1,240 (33.9%) were marked as “potentially problematic” or as providing “insufficient” information.

The other key takeaway is that the task of analysing the content of privacy policies can be, to a significant degree, realised by computers using artificial intelligence, if a sufficiently large data set is created. In practice, this means that many more privacy policies need to be manually analysed before the analysis can be fully conducted by machines alone.

5. What are the problems that have been identified?

- Companies do not provide all the information which is required under the GDPR transparency obligations. For example, companies do not always inform users properly regarding the third parties with whom they share or get data from.
- Processing of personal data often does not happen according to GDPR requirements. For instance, clauses stating that that by simply using the website of the company, the user agrees to its privacy policy.
- Policies are formulated using vague and unclear language³, which makes it very hard for consumers to understand the actual content of the policy and how their data is used in practice.

6. Why has BEUC decided to support this project?

The GDPR aims to give consumers more transparency and control over how companies collect and use their personal data. During the weeks before and after the 25th of May 2018⁴, most companies amended their privacy policies to bring them in line with the new Regulation. Users were faced with a *tsunami* of updated policies and new consent requests. It is very hard for them to assess whether their rights are being respected.

Consumer organisations whose job is to defend users’ rights and keep the market in check must support consumers in protecting their privacy and autonomy, yet their means are also limited. The same can be said for enforcement authorities in charge of monitoring compliance with the GDPR requirements.

In a connected world where everything we do entails the processing of personal data in one way or another, privacy policies will become even more complex and omnipresent. New tools are necessary to monitor and analyse what’s in those policies.

BEUC decided to support this project for two reasons. First, to have a snapshot of the level of compliance with the GDPR transparency requirements. Second, to support the development of an AI-powered legal evaluation tool which could greatly facilitate the lives of consumer organisations, consumers and the work of enforcement authorities when testing whether companies’ privacy policies comply with the GDPR.

³ For example, language qualifiers such as “may”, “might”, “some”, “often”, and “possible”

⁴ Date of entry into application of the GDPR

7. How does the “Claudette” platform function and what is it analysing?

The researchers from the EUI developed a web crawler which has been used to monitor, on a daily basis, the privacy policies of a predefined list of online services. The data retrieved by the crawler was then processed using supervised machine learning technology.

The machine learning tool (“Claudette”) is trained to scan the content of the privacy policies and assess it on the basis of a “gold standard” outlining what a fully GDPR-compliant privacy policy would look like. This “gold standard” was developed by the researchers using the requirements established in the GDPR and the relevant guidelines from the Data Protection Authorities⁵.

The result of this process is a colour coded “annotated” privacy policy where all clauses are categorised, singling out those clauses that could be non-compliant or, at least, problematic and therefore deserving special attention. You can see all the results at <http://www.claudette.eu/gdpr/>. What is currently displayed on the website is the outcome of human-preformed annotation.

8. Is “Claudette” ready to analyse privacy policies all by itself?

These are preliminary results and so far “Claudette” has only been trained with a small number of privacy policies, therefore the results of the automated scanning are not 100% accurate. More data is needed to obtain higher quality results. The more privacy policies “Claudette” can learn from, the more accurate its assessment of other privacy policies will be. For the purposes of this research, the results of the automated scanning were manually checked. This is the reason why the experiment was limited to 14 services. In the future, human intervention might be necessary to a much lower degree.

9. Is this a one-off experiment? If not, what are the next steps?

BEUC and EUI intend this to be a long-term project to develop automated tools that can help private and public enforcement actions. BEUC will bring this research to the attention of the data protection authorities and will continue monitoring market developments closely. We do not rule out taking further legal actions as appropriate.

10. What does this project mean for the future of market control and enforcement?

Artificial intelligence tools can enable consumer organisations and enforcement authorities to systematically undertake automated analysis of millions of privacy policies or – for example also - contract terms and conditions.

In a digital economy and society, technology can also help to ensure better compliance of business with fundamental consumer rights and guarantee that infringements cannot go unnoticed.

⁵ Guidelines of the former Article 29 Working Party, now European Data Protection Board