

Monitoring and analysing social media in relation to IP infringement



Executive Summary

The growth of e-commerce has been well documented, but how the rise of different technologies and consumer habits has affected intellectual property rights (IPR) infringement on the internet and, in particular, on social media platforms, is not clear. It is within this context, the European Union Intellectual Property Office (EUIPO) decided to conduct a study to better understand the volume and frequency of IPR infringement on social media⁽¹⁾. This study was organised around three strands of interest to better understand the current activities and trends related to counterfeit physical products and the piracy of digital content.

The first part of the study aims to provide a comprehensive picture of the social media uses related to possible IPR infringement activities or promotion; the second part measures the relative presence of IPR infringement on physical products and digital content on social media compared to genuine products or licit copyright-protected digital content; and the third part intends to identify key indicators in order to better recognise IPR infringement business models on social media.

The scope of the analysis was conducted by data mining four social media platforms (Facebook, Twitter, Instagram, and Reddit) in six European countries: Germany, Spain, France, Italy, Poland and Sweden, although conversation traceability on the social media channels selected for the study was not always possible. In order to mitigate this limitation, language criteria were used to allocate conversations to the targeted countries. Consequently, English was added to the six European Union (EU) languages considered for the study (French, German, Italian, Polish, Spanish, Swedish) as it is widely used on social media, by non-native speakers.

The study applied a methodology based on social intelligence analytics (SIA) and a quantitative and a qualitative analysis. In targeting conversations related to IPR infringement of physical products and digital content, one of the limitations of the study was the right to access certain conversations on social networks. Therefore, the study addressed only public conversations. Another issue highlighted by the study was the

⁽¹⁾ As highlighted in the [EUIPO 2020 Status Report on IP infringement](#) published in June 2020.

difficulty in distinguishing between licit and illicit content. Finally, the results presented in this report naturally represent a restricted sample and the analyses can only speak for the six countries, seven languages and the four social media channels included. Nonetheless, this variety of sources and the breadth of the data means the conclusions drawn can be considered as holding a broader relevance. Likewise, the significance of this report regarding the behaviours identified on social media can be considered indicative of wider habits that are currently prominent on social media in relation to IP infringement.

Social media uses for IPR infringement activities or promotion

A total of 3.9 million conversations related to the categories and brands chosen to represent physical products were extracted for the study using SIA. Among these categories of physical products, toys, perfume and cosmetics had the highest volumes of conversation. Regarding digital content, films, music and video games had the highest volume of conversations.

An IPR infringement-related conversation was identified as any conversation concerning a breach of an IPR. The study identified 11 % of conversations regarding physical products could be possibly related to counterfeits, and 35 % of conversations on digital content could be possibly related to piracy. Although the quantitative analysis was conducted carefully, the study highlighted the difficulty to identify IPR infringement with certainty.

When examining where there were the highest number of IPR infringement incidences, clothing, footwear and jewellery all featured in the top 3 for physical products. E-books, TV shows and music recorded the highest number of infringements for digital content. The clear conclusion to be drawn from this evidence was that these products collectively had the highest rate of conversations that were identified as being possibly related to IP infringement.

The relative presence of conversations related to counterfeit products and pirated content on social media compared to genuine products or licit digital content

The second part of the study, based on the analysis of three different aspects: platforms, languages and timelines, identified certain trends in IPR infringement activities and efforts to promote them.

In the case of IPR infringement for physical products, the study revealed that Instagram was the social network with the highest total volume of conversations. At product category level, Instagram was also the main platform for conversations about watches, toys, perfume and cosmetics, jewellery and footwear. Twitter was found to have the highest volume of conversations about clothing and toys, and Reddit was used most often for conversations regarding pharma and headgear. In contrast, Facebook⁽²⁾ showed a lower volume of conversations, which could be explained by an efficient approach from the platform to identify and delete infringing content⁽³⁾. This, however, cannot be corroborated by the evidence in this report and, in fact, it may well be that Facebook's low number of recorded conversations was due to IPR infringement-related conversations taking place in private rather than in public. Although all platforms have a direct messaging function, the prominence of private groups appears to be a phenomenon unique to Facebook. As this study only addressed public conversations, this hypothesis could not be confirmed but it should be borne in mind when noting the comparably low numbers of conversations suspected of IPR infringements identified on Facebook by SIA.

In the case of IPR infringement for digital content, a different pattern appeared. Reddit became particularly important for conversations related to films and TV shows, while Twitter was preferred for conversations related to music and e-books.

Regarding the language parameter, the first part of the study highlighted that English was predominant and that some languages, such as Polish and Swedish, were used infrequently. Despite this, it was observed that a high volume of French, German, Italian

⁽²⁾ Private Facebook groups are excluded from this study.

⁽³⁾ Facebook's measures to face IPR infringement include a 'global notice-and-takedown program, a robust repeat infringer policy, and additional specialised measures going beyond notice-and-takedown.' – [Facebook transparency report](#).

and Spanish conversations were found on Instagram regarding physical products, except for pharma-related mentions, which took place more frequently on Twitter.

As the study's duration was 6 months (April to September 2020) and took place during the COVID-19 pandemic, it was observed that the results obtained were affected in part by the lockdown. In order to contextualise the results, the overall volume of IPR-related conversations was compared in parallel with the overall pattern of IPR conversations recorded over the last 3 years (June 2017 to July 2020). This highlighted that, for physical products, the pharma category showed a noticeable increase of conversations, and the clothing category peaked around the end of the lockdown. In contrast, the digital content trend showed a progressive increase in conversations that could be interpreted as a learning-curve effect as users became more familiar with both licit and illicit digital content consumption.

IPR infringing business models used on social media

The topic modelling and the qualitative analysis conducted to identify possible IPR infringing business models highlighted that it was very difficult to find a pattern regarding IPR infringement on social media. This could be explained by the following observations.

- In the case of a physical product, the analysis revealed that conversations related to IPR infringement concerned mainly promotions and commercial activities. The providers of counterfeit products copied proven and successful business models already used by the legitimate brands.
- In the collected conversations, piracy of digital content was driven by the users' intention to access pirated content. Indeed, all conversations were about finding ways to access content illegally. Consequently, providers of pirated digital contents assumed a passive role, relying on users to disseminate the information and attract new users to their content.

Although tailor-made IPR infringement business models could not be established in this study, the analysis led to the definition of what can be termed a 'bundle-of-clues' approach to help to identify the conversations related to IPR infringement for both physical products and digital content. A more reliable approach to detection would also

be made easier by using trained machine learning models that could provide predictions and recommendations.

The study demonstrates that social media platforms are tools for recurrent IPR infringement for both physical products and digital content. Furthermore, identifying this type of content is complex due to the varied and constantly changing approaches used by providers of counterfeit products or pirated content, which are widespread across multiple platforms, languages and content types.