

Facebook and Artificial Intelligence: A Review of Good Practices

Beatriz Pérez -Acuña, Celia Fernández-Aller

in: Matthias C. Kettemann (ed.), How Platforms Respond to Human Rights Conflicts Online. Best Practices in Weighing Rights and Obligations in Hybrid Online Orders (Hamburg: Verlag Hans-Bredow-Institut, 2022)

Facebook and Artificial Intelligence: A Review of Good Practices

Beatriz Pérez -Acuña, Celia Fernández-Aller¹

UNIVERSITY OF HUELVA//UNIVERSIDAD POLITÉCNICA DE MADRID

Introduction. Facebook and Artificial Intelligence

This article compiles the ethical controversies reflected by journalists appearing in the media about Facebook in relation to the use of artificial intelligence (AI) in the different products of this company, with a special focus on content moderation. Different applications of AI are shown, with both positive and negative impacts.

A human rights approach (HRA) is being used to avoid the risks of “ethics washing” (Floridi, 2019) and the co-optation of AI ethics by big tech (Jobin, 2022). Many ethical recommendations are currently in place for developing algorithms (Fjel, Achten, 2020). However, recommendations that focus on human rights are still scarce. As Fjel suggests: “On its own, a set of principles is unlikely to be more than gently persuasive. Its impact is likely to depend on how it is embedded in a larger governance ecosystem, including for instance relevant policies (e.g. AI national plans), laws, regulations, but also professional practices and everyday routines”.

The HRA is based on strong principles (such as universality of human rights, participation, transparency, accountability, sustainability...), and it has the element of enforcement, which allows rights holders to claim rights. Additionally, human rights are universally accepted, and this makes a great difference compared to the wide variety of AI codes of ethics. Principles alone cannot guarantee ethical AI, due to the problematic implementation (lack of proven methods) and the lack of accountability.

International human rights law is a governance regime with significant potential relevance to the impacts of AI. There is a strong connection between AI governance and human rights laws and norms. “64% of our documents contained a reference to human rights, and five documents took international human rights as a framework for their overall effort. Existing mechanisms for the interpretation and protection of human rights may well provide useful input as principles documents are brought to bear on individuals cases and decisions, which will require precise adjudication of standards like “privacy” and “fairness,” as well as solutions for complex situations in which separate principles within a single document are in tension with one another” (Fjel, Achten, 2020).

Facebook and artificial Intelligence at the centre of media controversy

In September 2021, the 37-year-old Haugen, who worked as an engineer at Facebook until May 2021, in the division for civic integrity, a unit within the organisation that aims to recommend policies that protect the general public. The former employee of the company, who exposed the organisation's darkest secrets, left the company, but hours before leaving, she pulled copies of dozens of documents that she said “prove that Facebook has always been aware that its algorithms fuel division, promote hate, spread fake news, and can

¹ Both authors have contributed equally.

have a profound impact on the emotional and physical health of teenagers. And instead of correcting this, the company has looked the other way, privileging its growth and profits rather than choosing to protect its customers" (Gomez, 2021 para.3). She disclosed tens of thousands of internal Facebook documents to the US Securities and Exchange Commission and the Wall Street Journal: "last week in a room on Capitol Hill, lawmakers from both parties gathered to hear testimony from Frances Haugen" (Gómez, 2021, para.1).

Gómez (2021) reports that according to Haugen, "the main problem is that Facebook's virality depends on how the division in charge of the company's growth amplifies the algorithm to ensure that the content "sticks more and is reproduced". "The more this happens, the more time users spend on the networks and the more profit Facebook makes, because it can expose them to more ads. And numerous studies have shown that the more controversial and divisive the content, even if it is false, the more traffic it generates. Haugen argues that Facebook is aware of this situation, but prioritises its own coffers" (Gómez, 2021, para. 6).

Haugen confirms what Facebook denies, among others, that "its social networks are ideal platforms for spreading disinformation, with the documents leaked by Haugen confirming that Mark Zuckerberg and his staff also know this, although they deny it" (Bécares, 2021 para.4). Artificial Intelligence, used to filter and select information, is still many years away from being perfected. Haugen says: "that the company encourages content that angers, polarises and divides". On the other hand, he says that "internal studies show that the company is lying to the public about significant progress against hate, violence and misinformation" (Zornoza, 2021 para.3).

According to the Facebook engineer, "it doesn't matter whether the information is true or false" (Genbeta, 2021 parr.12). The algorithm does not prioritise this aspect and Haugen says that, "it is since 2018 that this aspect is so marked. In fact, he recalled how European leaders have openly questioned this fact" (Genbeta, 2021 parr.12). With this, the former employee says: "Facebook makes more money when more content is consumed. People engage more with things that provoke an emotional reaction. And the more anger they are exposed to, the more they interact and the more they consume" (Perez, 2021 para.12). These internal studies point, Haugen argues, to the fact that "Facebook has been lying about significant progress against hate, violence and misinformation. According to an internal report, after all the changes, hate on the platform would have been reduced by 3-5%" (Perez, 2021 para.11). Haugen accepted the job at Facebook in 2019, making it a condition to work against disinformation on the platform.

One of the serious problems Haugen sees is that, as users: "You have your phone. You could see 100 pieces of content if you sit down and scroll for just five minutes. Facebook has thousands of options it could show you. The algorithm chooses among these options based on the type of content you have interacted with most in the past" (Genbeta, 2021 parr.11). Facebook is showing us in our 'Feed'. This is where the famous algorithms come into play, deciding what the consumer sees and does not see. The algorithm chooses some content over others "Social networks make us sick" accuses the platform of being intentionally as addictive as tobacco (Araújo, 25 September, 2020). All these statements by Haugen have put the social network in the spotlight of an ethical debate.

An internal audit warns Facebook that its inaction in the face of hate speech is "a step backwards for civil rights" (Araújo, 10 July, 2020). According to the former directive, when we are confronted with an information space that is filled with exalted, hateful or polarising content, social peace is somehow breached: "it erodes our civic trust, it erodes our faith in others, it erodes our ability to want to care for others" (Genbeta, 2021 parr.8). As a consequence, according to Genbeta (2021) "today's version of Facebook can fragment our societies and cause ethnic violence around the world" (para.8), he said in reference to hate

crimes that take place from some people to others because of their skin colour or religion. And he recalled "the ethnic cleansing that took place in Myanmar in 2018, when the military used Facebook to present its genocide and encourage hate speech towards the Rohingya ethnic minority" (Genbeta, 2021, para.9).

The differences between nudity in pornography and a nude body in art are not at all clear; the AI's selection is bizarre. The most emblematic statues of the renaissance have already been victims of this problem. On the other hand, we know about sex trafficking through the net. It has been discussed in the past how Facebook and its social media empire cares more about people's bodies than about racism or hate speech. The Vienna Tourist Board said that "museums in Austria's capital have faced many online challenges in displaying their works. After Vienna's Natural History Museum posted images of the Venus of Willendorf, a 25,000-year-old Palaeolithic limestone statue", Facebook deleted the images and labelled them pornographic. "In 2016, Facebook removed the Pulitzer Prize-winning photograph "Napalm Girl", which shows a naked girl fleeing a napalm attack during the Vietnam War. The photo had to be restored after AI censorship, reminding us of the long list of photographs removed by Facebook in its history. Facebook, and other social networks, may even have more excuses to remove nude images in part because of the threat of SESTA-FOSTA, a US law against sex trafficking" (Bécares, 20 October 2021).

Another source of debate and dispute is Facebook, which found itself in trouble after a new bias in its algorithms was revealed that had discriminatory results. In this case, the company was the cause of controversy: "after it became known that Facebook's AI labelled a video of black people as "primates", the social network apologised and said that the error was "unacceptable"". In this case the company was controversial because its AI labelled a video with black people as primate-related content" (Erard, 6 September 2021, para.1). The gross labelling error caused by the artificial intelligence was detected. According to the report, the video in question belongs to the British Newspaper *Daily Mail*. It was uploaded to Facebook on 27 June 2020 and, as described by the NYT, shows "black men in altercations with white civilians and the police" (Erard, 6 September 2021 para.3).

Below the player, the social network's recommendation algorithm "asked users if they wanted to continue watching videos about primates" (Erard, 6 September 2021, para.3). The problem that is putting Facebook in trouble is not new. Several companies have come under scrutiny in recent years for algorithmic biases that favour discrimination. This has been seen mainly in facial recognition technologies that "are especially erratic when processing dark skin tones" (Erard, 6 September 2021).

Another clear example, at Facebook "hundreds of engineers are rebuilding how its ads work, as announced by Graham Mudd" (Bécares, 11 August 2021, para.2), one of the company's top advertising executives. "Facebook claims it needs to track us to make Instagram free, and the question is who would pay €2 a month if they made it paid" (Bécares, 3 May 2021). Facebook's new rhetoric about making advertising more privacy-conscious could even be considered a defeat considering its previous statements about its use of people's information. Take its big campaign to object to Apple's ad tracking on the grounds that it was acting anti-competitively and hurting small businesses that relied on ads to reach customers: "Now Facebook says it is working on new approaches that respect privacy. It is worth remembering that it also has to do this because of legislation" (Bécares, 11 August 2021 para. 6).

The European Union is considering a ban on micro-targeted ads as part of a broad legislative proposal called the Digital Services Act, and the US government has recently signalled its "interest in monitoring "user surveillance" by "dominant internet platforms" (Bécares, 11 August 2021 para. 6).

Two years ago, "Apple threatened to remove Facebook and Instagram apps from its app shop" (Clarín, 2021). They were concerned that these platforms were being used as "a tool for trading and selling handmaids in the Middle East. In internal documents they stated that they were underestimating "confirmed abusive activity". There were domestic workers from the Philippines who shared on their social media accounts of being abused. Within Facebook, "you can see accounts with posed photographs of African and South Asian women, with ages and prices listed next to their images" (La Nación, 25 October 2021). The newspaper 20 Minutos (2021) notes that: "The Philippine government has a team of workers assigned to track" Facebook posts in an attempt to protect its citizens seeking work in Middle Eastern countries (AP, 25 October 2021).

Apple CEO Tim Cook criticises Facebook and others for not respecting privacy. According to Cook, these networks prioritise "conspiracy theories and incitement to violence" (Bécares, 29 January 2021). These are topics that engage more people in these conversations. In this way, the companies that run these social platforms have the opportunity to collect more personal information from citizens and thus have more data to sell.

Another schandal was the recent communications outage, which not only created chaos that ended up costing billions of dollars in losses, but also revealed the need for the use of this network by users. As a consequence, the CEO apologised for the fall of Facebook, WhatsApp and Instagram, somehow avoiding further criticism. In his own words: "Facebook, Instagram, WhatsApp and Messenger are back online now. Sorry for the disruption today. I know how much you rely on our services to keep you connected to the people you care about," Zuckerberg said in a terse message posted on his popular social network to alleviate the crisis cabinet (La voz de Galicia, 2021 parr.3).

The Metaverse project

Another demand that will not be easy to achieve is that of a report on the problems that may arise from the development of the metaverse, such as possible psychological or human rights harms caused by the use and abuse of the platform, and whether these can be avoided or are inherent to the evolution of the technology. Once this is done, and with all the information available, they are calling for a non-binding advisory vote by the board on whether shareholders consider it appropriate to proceed with the implementation of Zuckerberg's grand project. This CEO is gambling everything on the metaverse, and he is taking a huge risk. No tech company has ever invested so much in a single project to date. For example, this group is requesting that, "Meta produce a report analysing why the company's moderation policies have not been effective in controlling the dissemination of hateful, violent or misinformation content, and that this research be presented to the shareholder meeting. They also call for a further investigation to measure the actual and potential impact of targeted advertising on the human rights of users of the various platforms that make up Meta, and for that information to be published on the company's website by 1 June 2023" (Rodriguez, 2021 final).

Zuckerberg appears not only as CEO of the company, but as an individual defendant, exposing the social media mogul for the first time to possible financial and even criminal penalties. Racine believes that, in this case, "adding Mr Zuckerberg to the lawsuit is fully justified, and sends the message that any corporate leader, starting with the CEO himself, should be held accountable for his own actions" (Merino, 20 October, 2021 para.1) The metaverse is a hypothetical, but it is one that is likely to become a reality in a few years. "Facebook already has a gigantic list of plans to be the leading provider of virtual space for people to escape the real world" (Gonzalez, 28 October 2021 final).

Just as the company has a few years to try to do better this time, we as users also have a few years to consider its less bright and optimistic implications. "The internal business restructuring will make "Metaverse" the new Product group that will encompass all the different Augmented Reality and Virtual Reality related developments that the company holds" (Miguel, 2021 para.7). "Metaverse", the term used by Facebook to define its next steps in building a virtual universe that functions as "the next version of the internet", is so popular that it functions as a noun with a life of its own. Don't call it Facebook, call it Meta. The company changes its name and bets everything on the metaverse. The metaverse in Facebook's idea: a three-dimensional virtual zone in which humans, represented by avatars, interact in various ways. It is an evolution of the internet. Metaverse users access the environment through terminals that project them into the virtual environment.

"Mark Zuckerberg kicked off Facebook Connect 2021 by talking about the future, and for him it has only one name: the metaverse" (González, 28 October 2021). The Facebook CEO and company spent more than an hour describing a virtual world in which we will all want to live and, of course, spend money. The idea is by far a new one and virtual worlds are a thing of the past thanks to video games. However, Facebook's grand vision is to turn this into a kind of alternate reality in which to live, in the manner of the film, Ready Player One: "In fact, Zuckerberg believes the metaverse will replace the mobile web in the future and transcend the plane of screens" (Gonzalez, 28 October 2021). The problem is that no such world currently exists, and if it did, there is no way to enter it. Again, however, none of this is really accessible. Such spaces would require virtual and mixed reality equipment that simply does not yet exist, let alone be accessible to the masses, but which "Facebook clearly wants to build. Oculus is a sample of this and will be the starting point" (Gonzalez, 28 October 2021).

Facial recognition

Facebook's facial recognition had been active and at the centre of privacy debates for several years, and was intended to recognise members of the social network in any photo uploaded to it. Two years ago we even saw intentions to use mobile phone cameras to recognise faces wherever we go, something that made more than a few people panic. That doesn't mean that Meta is no longer interested in facial recognition: "the company advocates the technology to help people know when someone uploads photos of them, to unlock devices like Apple's Face ID or to prevent fraud" (Lopez, 2 November 2021 para.3). However, "at the same time, he admits that there are "growing concerns" in society and that nothing is yet clear at the regulatory level" (López, 2 November 2021 para.3).

Facial recognition systems are becoming increasingly common in everyday technologies. Mobile phone unlocking is the most obvious example, but not the only one. "We've seen for example facial recognition in security cameras, as we are sure to wonder, not so much according to new research, which has shown that this system can be bypassed with a virtual face" (Rus, 13 August 2021, para.2). Artificial intelligence is used to recreate a face that meets most of the face characteristics that facial recognition systems look for. Creating a face similar to the majority of the population is all it took for these researchers to overcome a number of existing facial recognition systems. "The researchers' moral of the story is that facial recognition systems may not be as secure as they seem. They propose using alternatives or extra checks so as not to rely entirely on a facial recognition system that, as we can see, is relatively easy to circumvent" (Rus, 13 August 2021, final para.).

TextStyleBrush

Then it's the turn of the project about handwriting. A new Facebook project manages to imitate most handwriting by reading a single word. Based on this word, the artificial intelligence learns what the typeface looks like and uses it to write any text: "TextStyleBrush is the name of Facebook AI's new project. It is an artificial intelligence that can copy the style of a text found in a photograph. What undoubtedly makes it impressive compared to other similar artificial intelligences is the small amount of data it requires: just a few letters from a word. In other words, you could theoretically write your name and the AI would know how to imitate your handwriting" (Rus, 14 June 2021 para.2). Facebook explains that while most AI systems can do this with specific tasks, in this case TextStyleBrush is flexible enough to understand all types of fonts from real environments. That is, it understands texts with all types of calligraphy, applied styles, thicknesses, colours, shading or rotations without apparent problems. They indicate that to achieve this, its artificial intelligence tries to understand the typography it analyses from a more holistic point of view. Rus comments: "They say there are still limitations and specific points to overcome, for example when the AI has to deal with metallic text and reflections" (Rus, 14 June 2021, final para.).

The actual usefulness of having an AI copy your handwriting. As is often the case with such developments in artificial intelligence, it can have both positive and potentially negative ethical consequences: "In theory such an AI could be used to forge someone's handwriting and speak on their behalf, for example" (Rus, 14 June 2021, para. 5). But also for positive applications. For example, it can be used to instantly translate real-world texts by simply photographing or pointing the camera. Be that as it may, for the time being it is only a public research by Facebook. We'll see if in the next few trials the improvement can become popular when applied to a real app or tool.

Glasses research and Immersive realities

A recent foray by Facebook into the realms of augmented reality (AR) could have a significant impact on the limits of artificial intelligence (AI) and its presence in our daily lives. "Just days after unveiling its AR glasses produced in collaboration with RayBan, Mark Zuckerberg's company announced the development of a project that could teach AI to understand the world as we see it." A quick glimpse of this process among others, Facebook keeps a record of all the websites you visit: "However, to reach this end, Facebook's experiment would have to see, hear and remember the personal information its users perceive on a daily basis" (Olaskoaga, 2021). In this vein, Facebook also announces work on virtual reality: "The company has already commercialised a popular virtual reality headset, the Oculus Quest 2, and plans to move from VR to augmented reality in the next few years" (Gilbert, 15 October 2021).

Facebook is researching AI systems to see, hear and remember user actions. Now the company's intentions with its AI developments are revealed. "If you were suspicious of their glasses in collaboration with Ray-Ban, you were right" (Márquez, 2021, para.2). Recording your day through a pair of glasses and sharing your life with the company has consequences. What is done with all that information? Although they assure users that they will be protected. Now, through research, it aims to discover the scope of its technology: "We know that Facebook is investing a lot of resources in its smart glasses and augmented reality, but at the moment, beyond recording video and taking photos, they don't have great capabilities. That could change in the future thanks to artificial intelligence. How? Well, one of the possibilities is to give the AI the ability to understand what is seen from our eyes, listen and manage to "remember scenes" (being able to answer questions like "where did I leave my keys?") or "remember words" (being able to answer questions

like "what did John say the other day?"). A most striking project that at the same time generates multiple doubts" (Pérez, 2021, parr.1 and 2).

Facebook wants the artificial intelligence of the future to be able to "forget" irrelevant data. "Researchers at the social network are working on a method to train an AI to forget information that loses relevance. They would achieve this by setting expiry dates. Facebook's AI project is very interesting, especially in times when the volume of available information is increasing. For this reason, the researchers propose a method called Expire-Span. It is designed to enable artificial neural networks to more efficiently classify and store information related to the tasks they have to perform" (Erard, 5 September 2021).

Facebook uses AI to remember everything the user does: "Facebook has developed a system for its Artificial Intelligence (AI) mechanisms that teaches them to forget certain information when it is not important to perform their functions, saving on memory and processing costs. The new technology, called 'Expire-Span', is "the first operation of its kind" according to the company, and aims to make AI neural networks resemble the workings of the human brain, which constantly forgets data, as Facebook said in a statement" (Altmann, 14 May 2021, paras 1 and 2). The use of Expire-Span offers benefits to some common AI mechanism tasks such as character-level language modelling, and improves efficiency in long-term context tasks in language, reinforcement learning, object collision and algorithm tasks.

Facebook has been shown to profit from every piece of data given to the platform: "One point that is very much worth commenting on is Facebook's ability to make money from users who do not pay for the service. The average monthly revenue per user (globally) was \$1.99 during the last quarter". "If we narrow it down to users in the US and Canada, the figure is much more impressive: \$8.63 per month per user. More than what is achieved by services like Netflix or Spotify, directly paid services that, between family accounts, trial months or specific discounts have lower average revenues". "This is possible thanks to Facebook's use of our data. Theoretically, and security scandals aside, Facebook does not sell our data to third parties, but sells third parties access to us thanks to the use of our data" (Lacort, 2018 parr.7,8 and 9).

Ego4D

This may change in the future thanks to artificial intelligence. One of the possibilities is to endow AI with the ability to understand what is seen from our eyes, to listen and to "remember scenes", being able to answer questions such as: "where did I leave my keys" or "remember words", "what did John say the other day? (Erard, 2021). A most striking project that at the same time raises many doubts about the ethical ground in terms of the use applied: "Ego4D is a long-term project in which Facebook plans to investigate the ability of AI to understand and interact with the world as we do, from a first-person perspective" (Márquez, 2021 parr.4). The company's ambitions with AI projects are becoming increasingly troubling from an ethical standpoint. As certain devices will give it the power to constantly analyse people's lives using material obtained from first-person recording. In the media, privacy is still at stake. For the moment, "Zuckerberg's company says it is only a research project and not a commercial development" (Márquez, 2021 para.6). Hard to believe, when they have just released a pair of glasses that can provide results from this research. His ideas about product projection are not far-fetched, as many technology companies have a similar vision regarding the uses of Artificial Intelligence and Augmented Reality: "With the entry of Metaverso, it would seem impossible not to think about such developments" (Márquez, 2021 parr.7).

Ego4D is Facebook's project for AI to see as humans see. Its development belongs to the artificial intelligence department where, in collaboration with 13 universities around the world, have created a

database to teach AI to understand the typical images and photos recorded from first-person devices "The Ego4D project is a powerful exercise in the impact that artificial intelligence can have. Algorithms where the more information we give it and the more we let it into our daily lives, the more accurate the answers it can give us" (Pérez, 14 October 2021 final). Where to draw the line is also an important debate that will need to be addressed.

While it is common for algorithms to work with data sets of videos and photos seen from afar, Facebook wants to anticipate a situation where first-person videos become more common. The problem is that while the AI is able to identify an image of a fair, it does not have such an easy time when the image of the fair is from the perspective of the person viewing it. The same is true for all kinds of situations where the angle is not from afar: "Next-generation AI systems will need to learn from a completely different kind of data: videos that show the world from the centre of the action, rather than on the sidelines" explains Kristen Grauman, a researcher at Facebook (Perez, 14 October 2021 para.4). According to Gilbert (2021): "Ego4D, which will use data "from 13 universities and labs in 9 countries, which have collected more than 2,200 hours of first-person video in a variety of settings, and drawn from the routines of up to 700 participants" (Gilbert, 2021, para.4). As Facebook explains, this is a twenty-fold increase in the amount of material that was available to help train the algorithms. (Perez, 14 October 2021, para.5).

If we consider how useful it might be for AI to be able to see and hear everything we do in the first person, we discover several utilities: "The Facebook team suggests five possibilities: Episodic memory: asking when something happened. By having a record of our life, the AI can answer questions that in many cases only we know. Prediction: anticipating certain routine steps. For example, in a recipe, the AI can warn us if we have skipped a step. Manipulation of objects: the AI can guide us to perform certain steps, for example, playing an instrument or giving instructions on how to position our body. Audiovisual diaries: by having a record of what we see, it would be possible to ask who said what and when. And so, for example, remember what number the teacher said or what time he/she stayed. Social interaction: by improving first-person understanding, algorithms could help improve vision or sound" (Perez, 14 October 2021, para.6 onwards). Huge potential... at the cost of teaching it our point of view. The possibilities of applying artificial intelligence from the point of view of the user's eyes opens up many possibilities, but also raises many privacy concerns. "Ten days ago, the leak of a gigantic database of Facebook users was revealed; data on 530 million people from all over the world, of which 11 million belonged to people living in Spain. Within this data are of course telephone numbers, dates of birth, places of work and, in some cases, the type of civil relationship" (Castillo, 2021, para.1).

Ego4D, Facebook's project to make AI see as humans see, its development belongs to the artificial intelligence department where, in collaboration with 13 universities around the world, they have created a database to teach AI to understand the typical images and photos recorded from first-person devices "The Ego4D project is a powerful exercise in the impact that artificial intelligence can have. Algorithms in which the more information we give it and the more we let it into our daily lives, the more accurate the answers it can give us" (Pérez, 14 October 2021 final). Where to draw the line is also an important debate to be addressed.

While it is common for algorithms to work with datasets of videos and photos viewed from afar, Facebook wants to anticipate a situation where first-person videos become more common. The problem is that while the AI is able to identify an image of a fair, it does not have such an easy time when the image of the fair is from the perspective of the person viewing it. The same goes for all kinds of situations where the angle is not from afar: "Next-generation AI systems will have to learn from a completely different kind of data:

videos that show the world from the centre of the action, rather than from the margins" explains Kristen Grauman, a researcher at Facebook (Perez, 14 October 2021 para.4). According to Gilbert (2021): "Ego4D, which will use data "from 13 universities and labs in 9 countries, which have collected more than 2,200 hours of first-person video in a variety of settings, and drawn from the routines of up to 700 participants" (Gilbert, 2021, para.4). As Facebook explains, this is a twenty-fold increase in the amount of material that was available to help train the algorithms. (Perez, 14 October 2021, para. 5).

Data management

Another intriguing controversy, about data privacy and the ethical consequences of AI, is "why Louis Barclay, a developer at the company who designed a tool that allows users to remove news stories that appear on Facebook, has been permanently banned from the tech giant's platform" (Asher, 2021, para.1). Barclay had been the creator of a browser extension called Unfollow Everything: "Louis Barclay, a UK-based developer, is the creator of a browser extension called Unfollow Everything. It allowed users to automatically unfollow all their friends and pages on Facebook, leaving their newsfeed blank. Barclay uploaded the Unfollow Everything app to Google Chrome in July 2020. The scientists wanted to study the impact of not having a news feed on how satisfied Facebook users were, as well as the amount of time they spent on the platform. In July of this year, Barclay received a letter from Facebook's lawyers, ousting him from the company" (Asher, 2021, para.2). Yet another example of how AI interferes with the way we think, as it selectively chooses the people, news and propaganda it wants, impacting the minds of users, without being able to delete and discard such content.

There is also the OnlyFans platform with exclusive sexual content: "OnlyFans is a subscription platform for exclusive and explicit sexual content. In it, there are profiles similar to those of other social networks, such as Facebook, Twitter or Instagram, but with the difference that to 'follow' or subscribe to someone's content you might have to pay". According to Avila (2021) there are free subscriptions: "There are also free subscriptions, from content creators who show their material without having to pay for it, and who generate income through other alternatives as part of their actions, or simply do not profit through the platform" (Avila, 2021 para.2). "Some media have labelled the service as a mechanism that facilitates virtual prostitution" (Avila, 2021 para.5).

Facebook's algorithm that conditions what is seen on users' home pages has been a common controversy for many years: "People's use (or participation) in the social network is decreasing. They saw that people were hiding 50% more posts" (Bécares, 26 October 2021, para.5). "What did happen is that they browsed more content in Groups, which is where they could find more posts of interest to them. Meaningful social interactions - the comments between friends that Facebook optimises - also dropped by 20%" (Bécares, 26 October 2021, para.6). Other projects were also carried out that were able to conclude how Facebook's algorithms can direct users towards divisive content. "The project also concluded that Facebook's algorithms can lead users down the path of conspiracy theories" (Bécares, 26 October 2021, para.9).

The discrepancy between Facebook's public claims about the effectiveness of its AI and the reality of the user experience has long baffled researchers and other regular users of the platform: "AI has minimal success in removing hate speech, violent images and other problematic content, according to internal company reports" (Wells et al, 18 October 2021 para.1).

On hate speech, the documents show, Facebook employees have estimated that "the company removes only a fraction of posts that violate its rules, a low single-digit percentage, they say. When Facebook's algorithms

are not confident enough that content violates the rules to remove it, the platform shows that material to users less frequently, but the accounts that posted the material go unpunished. Employees were analysing Facebook's success in enforcing its own rules on content that it details internally and in public documents such as its community standards" (Wells et al, 18 October 2021 para.3).

The Wall Street Journal series, based on documents and interviews with current and former employees, describes how "the company's rules favour elites; how its algorithms foster discord; that it has long known that drug cartels and human traffickers openly use its services; and how anti-vaccine activists use Facebook, among other issues. An article on the effects of Instagram on the mental health of teenage girls prompted a Senate hearing in late September. Examples of content that Facebook's AI should have detected but could not be seen include close-up videos of a person shooting someone and videos of car crashes with "visible dismemberment and entrails", according to the documents. "Other violations of Facebook's policies that were leaked through AI included violent threats directed at transgender children" (Wells et al, 18 October 2021 para 22).

In March, another team of Facebook employees came to a similar conclusion "estimating that these systems were removing posts that generated between 3% and 5% of hate speech views on the platform and 0.6% of all content that violated Facebook's anti-violence and violence policies. Incitement" (Wells et al, 18 October 2021 para.9).

In 2016, pop star Selena Gomez flew to Facebook's Menlo Park headquarters to pose for photos with Zuckerberg and Facebook COO Sheryl Sandberg to celebrate her status as the most-followed account on Instagram. Not long after, she was shocked to read a comment from a user on one of her Instagram posts: "Go kill yourself," according to the star's spokesperson, (Wells et al, 18 October 2021 para.44).

Another episode is hate speech in regional elections: In March, staffers preparing for regional elections in India said hate speech was a major risk in Assam, where there is growing violence against Muslims and other ethnic groups. "Assam is of particular concern because we do not have an Assam hate speech classifier," according to a planning document (Notimundo, 2021).

AI must also be trained in foreign languages. The social network's algorithms present problems when it comes to combating violence against minorities or human exploitation. According to a December 2020 memo, "that report also makes public Facebook's inability to address hate speech and harmful content outside the United States. In fact, hate speech and misinformation is substantially worse among non-English speaking users, according to the Facebook Papers. Much of Facebook's moderation infrastructure is not sufficiently resourced to operate in languages other than English, and its software has difficulty understanding, for example, certain dialects of Arabic" (The Debate, 2021 para.7). In that sense, "Facebook's moderation algorithm was only able to identify 0.2% of the harmful material in Afghanistan. The rest of the harmful content was detected by staff, even though the company lacked moderators who spoke Pashto or Dari, the country's main languages" (The Debate, 2021 para. 8).

Globally there are concerns about the use of the platform, "Facebook's artificial intelligence content moderation is unable to read some of the languages used on the platform. Facebook employees have expressed concern about how the system has allowed the platform to be used for nefarious purposes, according to documents seen by The Journal" (Channels, 2021, para.1). A former vice president warns: "A former vice president of the company has explained to the newspaper that Facebook perceives the potential harm in foreign countries as "the cost of doing business" in those markets" (Channels, 2021, paras 7 and 8). Thus, "drug cartels and human traffickers have used Facebook to recruit victims" (Canales, 2021, para.10).

Or the use of Facebook to incite violence: "In Ethiopia, some groups have used Facebook to incite violence against the people of Tigrayan, content that went unnoticed due to the lack of moderators who speak the native language. The company had also failed to translate its community rules into the languages used in that region, according to the Wall Street Journal" (Channels, 2021 para. 13).

A team of more than 350 specialists at Facebook is focused on stopping these organisations and trying to detect risks. Facebook has a huge blacklist full of names of users with criminal and terrorist backgrounds: "Individuals and organisations from all five continents are registered on this list, which Facebook has never wanted to make public for, among other reasons, ensuring the "safety" of its employees. Facebook uses this list to veto content and conversations that can be generated on its platform and that explicitly mention names such as those on the list. Drug cartels, criminal gangs, terrorist organisations and even political parties complete a list with thousands of entries. The list dates back to 2012, when voices began to be heard warning of the risk that these social networks were being used by terrorist organisations to propagandise and recruit new recruits. The list was also key after many experts singled out Facebook for facilitating the assault on Capitol Hill earlier in the year" (Aguilar, 2021, para. 1).

This algorithm affects agencies and advertisers, they cannot rely on the reach of organic traffic because, as we have seen, it practically no longer works, they have a limited budget with which they have to maximise results, unless the Digital Marketing agency is highly specialised in a specific sector, it usually has clients from different sectors and, not knowing in depth all their audiences, they may have added difficulties when it comes to achieving results with some Facebook campaigns.

Facebook's artificial intelligence systems sift through billions of posts looking for items that might match the company's definitions of content that violate its rules. The screening algorithms, called classifiers, are the foundation of the company's content moderation system. Building these classifiers is laborious and complex, requiring an army of humans to flag a large number of posts according to a set of rules. Engineers then take these examples and train their systems to determine the likelihood of other posts violating the rules.

Facebook executives have long said that AI would address the company's chronic problems by keeping what it considers hate speech and excessive violence, as well as underage users off its platforms. However, the problem is not over: "Facebook's algorithm today has little to do with what it was a few years ago. Now the algorithm is far more complex and uses AI - specifically machine learning - to help users connect with other users rather than to help them consume content in isolation" (Nielfa, 2022, -Artificial Intelligence and the Facebook Ads algorithm-).

There are other applications with positive feedback: "Facebook's algorithm can predict up to 96 hours in advance whether a patient is going to show deterioration. Since April last year, Facebook AI has been creating and sharing disease prognostic models to help health experts determine how best to plan and allocate resources in their area. Now, they are open-sourcing all of them for governments and researchers to use" (Hernandez, 2021). Facebook points out that these models are research solutions intended to help hospitals in the coming days and months with resource planning: "A model for monitoring and forecasting covid-19 or improving learning environments for robots are some of the announcements made by the company during the Facebook AI Innovation Summit" (Hernandez, 2021).

A deepfake is a video in which a person's voice and face are changed by artificial intelligence software, making the altered video look authentic: "Facebook has developed an artificial intelligence that wants to identify deepfake images and then track down their creators" (Erard, 2021). This technique is mostly used

with public figures: "Deepfakes abound on the web and some are very difficult to identify. The technology behind these fakes allows for increasingly realistic end results, and that is a real concern. Facebook is one of the many companies working on tools to detect them, and in the last few hours it has unveiled some notable advances. The social network has teamed up with Michigan State University (MSU) to develop a method that not only aims to identify deepfakes, but also to trace their origin. What does this mean? It means that, from the analysis of an image, the technology can determine whether it was artificially generated and detect which generative model was used to produce it" (Erard, 2021 par.1 and 2).

Good practices...or bad practices?

Positive and negative impacts of Facebook's algorithms

As discussed in previous sections, Facebook's use of algorithms has positive and negative impacts.

Journalists have highlighted more the negative ones: the company is an example of bad practice in relation with privacy of users, the fuel of division, the promotion of hate, the spread of fake news, and the profound impact on the emotional and physical health of teenagers.

Literature has also emphasised more the negative impacts: Facebook may seem like a social network on its surface, but its real business is trading in influence through personal data. It is more a personalized advertisement platform than a social medium (Veliz, 2021). There is a report of the British parliament that compares Facebook with a digital ganster (House of Commons, 2019).

Researchers coming from the technological field have also raised concerns. Although machine-learning used by Facebook could have a positive impact in disinformation governance, "there are several challenges that still need to be addressed. Firstly, artificial intelligence tools are not only used to counter disinformation, but they are accelerating the threat, for example by providing systems to generate fake news, images and videos. Secondly, once a detection system has been deployed, malicious actors can modify their behaviour or exploit adversarial attacks to avoid detection. Therefore it is necessary to devise detection systems that are robust to changes in the underlying distribution of the data. Finally, there is a general lack of solutions that effectively combine together multiple modalities (text, video, speech and network analysis), therefore providing more robust and accurate solutions" (Camacho, David². Special Session at the International Joint Conference on Neural Networks (IJCNN) 2022. In conjunction with IEEE World Congress on Computational Intelligence (WCCI) 2022).

A good practice which can be highlighted is in relation with minors: Facebook is not going to use Instagram Kids application, which has been developed recently. It is due to concerns related to minor's rights (<https://about.instagram.com/blog/announcements/pausing-instagram-kids>).

Facebook had an opportunity to show the world their engagement with people's rights in the study called Experimental evidence of massive-scale emotional contagion through social networks (Kramer et al., 2014). It was a collaborative Research Ethics endeavor between Facebook and Cornell University's Departments

² Camacho's research group is leading a project called "CIVIC: Intelligent characterisation of the veracity of the information related to COVID-19", which goal is to combine the knowledge of experts in communication and journalism with experts in Artificial Intelligence techniques in order to implement a tool for the general public aimed at characterising automatically information related to COVID-19.

of Communication and Information Science. In it, Facebook researchers directly manipulated Facebook users' news feeds to display differing amounts of positive and negative posts from the people they followed in order to determine whether their subsequent posts were affected by the positivity or negativity of the set of posts they were viewing. This effect, that more positive or negative posts read by a user could change their own emotional state positively or negatively, is the 'emotional contagion' referenced in the article. Facebook allowed the scientists (both internal and those from Cornell) access to the huge amount of data that was produced by manipulating what Facebook users saw according to computerized determination of positivity and negativity levels.

A good practice could have been also the commission Facebook did of a Human Rights Impact Assessment (HRIA) to evaluate its role in the genocide of the Rohingya in Myanmar. The HRIA was found to be largely ineffective at uncovering the human rights harms of Facebook's AI-enabled tools and identifying appropriate mechanisms to mitigate those harms moving forward (Latonero, 2021).

As it is highlighted by the literature (Benesh, 2020) "Facebook, Inc. is quietly running the largest system of censorship the world has ever known, governing more human communication than any government does, or ever has". There is a problem with the human rights impacts of the companies: they can not be held accountable for impacts produced outside their main establishment country. The United Nations treaty on Business and Human Rights is not legally binding yet. For this reason, rights of citizens affected by bad practices can not be enforceable yet.

There are some examples of good practices in some other social networks, as Care2.com (<http://www.care2services.com/online-social-action-network>) is an online social action network with over 50 million members around the world. Care2 was founded in 1998 with a simple mission: to help make the world a better place. Today, Care2 is a highly-engaged social network of over 40 million citizen activists standing together for good and making extraordinary impact - locally, nationally and internationally - by starting petitions and supporting each other's campaigns. Care2 has been a pioneer of online advocacy since its inception. They provided the first central platform for online petitions, and were the first to help nonprofit organizations tap into this passion to grow their organizations. Care2 has now helped 2,000 nonprofit organizations recruit more than 50 million donor leads.

The Care2 community realizes the power of online petitions in a world where the government can be frustratingly ineffective. Together, they have protected wolves around Denali National Park, won justice for victims of discrimination, saved dogs from euthanasia, helped get GMOs out of Hershey chocolates, stopped an old growth oak grove from destruction, shut down an abusive farm, stopped the killing of rare owls, and much more.

Another good example is Networks For Good (<https://networksforgood.com/>) whose key function is the user's ability to create their own groups, events and share their experiences and beliefs with others in the business. By doing so, this creates a new level of meaningful engagement and generates positive momentum in delivering a meaningful engagement strategy across an organisation that will secure lasting relationships between employers, employees, members and organisations.

The role of digital rights

Facebook's algorithms have positive and negative impact on digital rights (positive, in the right to digital health, for example; negative, in the right to privacy and data protection, equality, freedom of expression, free information).

The European Commission unveiled recently two of the most important components of its digital agenda: The Digital Services Act (DSA) and the Digital Markets Act (DMA). The DSA proposal introduces new rules on how online marketplaces and content hosting platforms deal with illegal content, including special transparency and auditing obligations for very large platforms with more than 45 million monthly active users in the EU, a threshold surpassed by several services including Facebook, YouTube, Twitter and TikTok. The Commission did not force platforms to monitor and censor what users say or upload online.

The proposal for a Regulation (DSA) takes over the same two principles as the e-commerce Directive, namely that they are not liable for illegal content that they host or transmit, as long as they do not have actual knowledge of it, and that there is no general obligation to monitor to prevent the publication or transmission of such content. It therefore maintains the general legal framework of liability of the Directive, but nevertheless introduces new obligations for hosting service providers and online platforms: the creation of specific processes for requesting the removal of illegal content, the introduction of mechanisms for users whose content has been removed to defend themselves on the grounds of their right to freedom of expression and information, the obligation to cooperate with the competent authorities of the Member States in the process of removing illegal content and identifying certain users, the obligation to cooperate with the competent authorities of the Member States in the process of removing illegal content and identifying certain users, and the obligation to ensure that users are able to exercise their right to freedom of expression and information (Pérez de las Heras, 2022).

The DSA includes a number of due diligence obligations: some affect all intermediary service providers and include, among others: the establishment of a single point of contact for direct communication, the designation of a legal representative in the Union in the case of providers not established in the Union but providing services on European territory, information in their general terms and conditions on the restrictions imposed by the intermediary on the use of their services, including the content moderation policy and the use of algorithmic decision-making systems (Art. 10-12).

The DSA recognizes a second set of additional obligations specifically addressed to online platforms, unless they are small and medium-sized enterprises. These specific obligations include the establishment of internal complaint systems to handle the possible removal of illegal content, cooperation with out-of-court dispute resolution services, preferential treatment of takedown notices of illegal content handled by so-called "trusted reporters" (introduced in Art. 19 of the proposal), suspension of the use of the content of the platform (Art. 19 of the proposal) and the suspension of the use of the content of the platform (Art. 19 of the proposal), the temporary suspension of services to recipients with a continuous history of infringements, the reporting of activities suspected of constituting serious offences, the collection of information necessary to enable the traceability of business customers offering products or services on the platform and the proper identification of advertising displayed on its interfaces, including information on the criteria used to select advertising recipients (arts. 14-24). For large platforms, defined as those with at least 45 million active monthly users in the EU, additional obligations are added to identify potential risks and, where appropriate, mitigate them. These obligations include the preparation of a risk analysis to determine the risks related to the distribution of illegal content, the introduction of mitigating measures for the risks identified, the carrying out of independent audits once a year to monitor compliance, the description of the parameters used in their recommendation systems and the obligation to designate a compliance officer responsible for supervising compliance with the established tasks (arts. 25-33), (Pérez de las Heras, 2022).

The proposed Regulation provides also for a regime of public guardianship of these obligations. The monitoring and protection scheme is inspired by the GDPR system itself. Thus, a Digital Services Coordinator responsible for compliance with the future Regulation in each Member State and for processing user complaints is envisaged, as well as the creation of a European Digital Services Committee in which the competent authorities of each Member State will participate. Finally, a system of financial penalties is also envisaged in the event of non-compliance with the Regulation, although the Member States are given discretion to impose them within the limits established by the proposal, which can reach up to 6% of the service provider's annual revenue (Articles 40-49).

European lawmakers have the chance, with the Digital Services Act, to ensure that public interest researchers, including academia, journalists, and civil society organizations, have access to the data needed from large platforms.

Concerning impacts of Facebook algorithms on human rights, the Proposal of a regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) is also an important step. It is a risk-based framework and AI applications are categorised by unacceptable risk, high risk, limited risk and minimal risk. The facial recognition system would be high risk. Many other AI applications used by Facebook would be limited or minimal risk. These means less duties for the platform.

Another interesting regulation is the recent proposal of a European Directive on due diligence. The European Parliament has a Report with recommendations to the Commission on corporate due diligence and corporate accountability (2020/2129(INL)). The draft directive proposes that EU companies with more than 500 employees and a turnover of 150 million euros should be obliged to put in place specific measures to prevent human rights and environmental abuses along their supply chains, through so-called "due diligence". The draft proposes that in sectors with a higher risk of exploitation, such as agriculture and the fashion industry, it would apply to companies with more than 250 employees and a turnover of 40 million, while SMEs would be exempted. Companies from non-EU countries operating in the single market and above these thresholds would also be covered (Facebook could be an example). Under the Commission's proposal, companies could be held liable for damages committed by their subsidiaries, subcontractors and/or suppliers inside or outside their borders; and victims of these rights violations could file complaints in EU courts. This is an important step in establishing the right to remedy for victims of corporate malpractice.

Conclusions

Machine learning algorithms must put people first. Algorithms dominate our lives, and while some change our lives for the better, we are beginning to realise that letting a machine make decisions may not be such a good idea. Algorithms are beginning to dominate our reality: they do so when they recommend a song on Spotify, a series on Netflix or a purchase on Amazon, but also when they are in charge of choosing the best candidates for a job or granting a loan.

That makes for a bittersweet feeling about these algorithms that make our lives easier but also condition them: "we should be able to audit these algorithms to take back control of both those algorithms and our lives" (Pastor, 15 February 2017 paras. 1 and 2). Those who program these algorithms are humans with their own experiences and opinions that to a greater or lesser extent can end up affecting these systems: "This

means that we end up with algorithms that are biased, sexist, conditioned, subjective and as unfair as we ourselves are" (Pastor, 13 April 2018 para.25).

Humans are good at distinguishing which content can hurt sensibilities, but machines still have a lot of trouble differentiating between hate speech, race, sex, politics, etc. This is one of the great challenges of artificial intelligence. The problem of detecting such content and comments has not been solved perfectly by Facebook's artificial intelligence systems, but Zuckerberg is optimistic about the future: "Humans are good at distinguishing what content can be hurtful, but machines still have a lot of trouble. Hate speech is one of the hardest problems to tackle, but I think artificial intelligence can get us there in five to ten years" (Mollejo, 2018, para.4). Zuckerberg insists on the use of artificial intelligence to "police" Facebook content. This would make it possible to differentiate directly and in real time when such content or comments are part of a debate and are acceptable, or when there is aggression involved and should be moderated and censored.

It is not entirely clear whether this is the system they use on Facebook to try to moderate and censor inappropriate content, but what is certain is that the system is used on Instagram: "Another problem in the air, that of DeepText and other similar systems is that of false positives: blocking comments and content that were not toxic" (Pastor, 13 April 2018 parr.19). It would classify the comment in special categories such as 'prohibited behaviour', 'racism', 'sexual harassment' or 'bullying'. "Among the main risks is that of false positives: the problem is that a word that has certain meanings can be understood differently depending on the context or the passage of time" (Pastor, 13 April 2018 parr.14). Google has also been working on this problem for some time. Therefore, we find light and shadows of AI as a moderator and content cleaner: "The problem is even more worrying when we see how those who most use artificial intelligence to filter, moderate or censor do so in a dangerous way: without apparent control" (Pastor, 13 April 2018 parr.14).

The role of the human being in ensuring that such content does not become toxic is more important than ever. Mark Zuckerberg's social network has millions of users and thousands of posts per day. But there are those who break ethics and decide to spread false news, drug advertisements, harassment, racism. One of the people who fights for this "justice" on the social network is Joaquín Quiñonero, director of applied machine learning at Facebook. His 'weapon': artificial intelligence, the option for the future. Quiñonero came to Facebook after working at Microsoft" (Fiter, and Encabo, 18 November 2019, para.1). The latest data published by Facebook highlights how this technology is revolutionising its fight against this serious problem: "in recent months they have removed nearly 3.2 billion fake accounts and more than 11 million posts that incited hatred". The expert wanted to represent what kind of frauds the algorithms are up against (Fiter, and Encabo, 18 November 2019, parr.3). Artificial intelligence is also behind the identification of people appearing in photographs or videos.

This contribution concludes that platforms' efforts to combat disinformation through AI are not enough. The international regulation framework (specially the European Union framework) is an opportunity to help platforms moderating contents.

The impacts of machine learning algorithms on ethical values and on human rights are many, both positive and negative. The second ones need more effort from different stakeholders, such as states, private sector, civil organizations and intergovernmental organizations.

References

- 20Bits. 5 October 2021 "WhatsApp, Facebook and Instagram suffer a global downtime of more than six hours". <https://www.20minutos.es/tecnologia/actualidad/caida-mundial-de-whatsapp-facebook-e-instagram-no-permite-usar-las-apps-4844020>
- 20Bits. 7 November 2021 "A Facebook study reveals that 360 million of its users are compulsive addicts to this social network". <https://www.20minutos.es/tecnologia/estudio-facebook-revela-millones-usuarios-adictos-compulsivos-esta-red-social-4882477/>
- Aguilar, Albert R. 14 October 2021 "Filtran la -lista negra- de nombres de Facebook: estos son los -terroristas o criminales- que están vetados en la red social". <https://www.businessinsider.es/lista-negra-facebook-llena-terroristas-criminales-948005>
- Altmann, Gerd. 14 May 2021 "Facebook brings artificial intelligence closer to the human brain and teaches it to forget" PORTAL TIC EUROPA PRESS. <https://www.europapress.es/portaltic/sector/noticia-facebook-acerca-inteligencia-artificial-cerebro-humano-le-ensena-olvidar-20210514150033.html>
- Scope. Technology. 19 October 2021 "What is Ego4D, Facebook's controversial new project" <https://www.ambito.com/tecnologia/facebook/que-es-ego4d-el-nuevo-y-controversial-proyecto-n5301603>
- Antoñanzas, Diego. 2021 "Facebook uses artificial intelligence to guess love relationships" <https://diegoantonanzas.com/facebook-usa-inteligencia-artificial-para-adivinar-relaciones-amorosas/>
- Araújo, Santi. 10 July 2020 "An internal audit warns Facebook that its inaction on hate speech is "a step backwards for civil rights"" <https://www.genbeta.com/redes-sociales-y-comunidades/auditoria-interna-advierde-a-facebook-que-su-inaccion-discurso-odio-paso-atras-para-derechos-civiles>
- Araújo, Santi. 25 September 2020 "Social media makes us sick : former Facebook exec accuses the platform of being intentionally as addictive as tobacco" <https://www.genbeta.com/redes-sociales-y-comunidades/redes-sociales-nos-enferman-ex-ejecutivo-facebook-acusa-a-plataforma-ser-intencionalmente-adictiva-como-tabaco>
- AP, 25 October 2021 "Apple threatened to pull Facebook app", 20Minutos Actualidad. <https://www.20minutos.com/noticia/327976/0/apple-amenazo-con-retirar-la-aplicacion-de-facebook>
- Asher Hamilton, Isobel. 30 October 2021 " Facebook bans developer for good after he created an app that allowed users to delete their news feed" <https://www.msn.com/es-es/noticias/tecnologia/facebook-expulsa-para-siempre-a-un-desarrollador-despu%C3%A9s-de-que-creara-una-aplicaci%C3%B3n-que-permit%C3%ADa-a-los-usuarios-eliminar-su-feed-de-noticias/ar-AAPqvj2?getstaticpage=true&automatedTracking=staticview&parent-title=los-jeans-con-flores-bordadas-que-te-enamorar%C3%A1n-y-puedes-hacer-t%C3%BA-misma&parent-ns=ar&parent-content-id=BBsRwzk>
- Avila, Lucia. 24 August 2021. "OnlyFans, a dangerous social network: What is it and how does this adult content platform work?" https://www.ondacero.es/noticias/sociedad/onlyfans-peligrosa-red-social-que-como-funciona-esta-plataforma-contenido-adulto_202108246124f3fb1827770001a3cfb0.html
- BBC News World. 3 February 2022. Updated 4 February 2022 "Facebook: Meta's unprecedented plunge on the stock market after the social network's first drop in active users" <https://www.bbc.com/mundo/noticias-60244251>
- Bécares, Barbara. 11 August 2021. "Facebook says it's looking at how to continue its big ad business, but knowing less data about you." <https://www.genbeta.com/actualidad/facebook-dice-que-esta-estudiando-como-seguir-su-gran-engocio-publicidad-sabiendo-datos-ti>
- Bécares, Barbara. 14 April 2021. "European Union, its first regulation to regulate the use of artificial intelligence" DevelopWordNews. <https://www.genbeta.com/actualidad/union-europea-sacara-su-primer-reglamento-para-regular-uso-inteligencia-artificial-adelanto-que-plantea>
- Bécares, Bárbara. 20 October 2021 "Vienna Museums show their nude art on OnlyFans to avoid being censored by Facebook." <https://www.genbeta.com/actualidad/museos-viena-muestran-su-arte-desnudos-onlyfans-para-evitar-ser-censurados-facebook>
- Bécares, Barbara. 26 October 2021 "The Facebook Papers lay the company bare and reveal everything from human trafficking to why Instagram is losing young people" <https://www.genbeta.com/actualidad/facebook-papers-dejan-compania-al-desnudo-desvelan-trafico-personas-que-instagram-pierde-a-jovenes>
- Bécares, Bárbara. 29 January 2021 "Tim Cook criticises Facebook and others for not respecting privacy, and Apple will launch a feature to prevent tracking" <https://www.genbeta.com/actualidad/tim-cook-critica-que-facebook-otras-no-respeten-privacidad-apple-lanzara-funcion-para-impedir-su-rastreo>

- Bécares, Bárbara. 3 May 2021 "Facebook claims it needs to track us to make Instagram free, and the question is who would pay €2 a month if they made it paid" <https://www.genbeta.com/actualidad/facebook-reconoce-que-necesita-rastrear-nos-instagram-sea-gratis-pregunta-quien-pagaria-2eur-al-mes-hacen-pago#:~:text=This%20will%20translate%20to%20the%20same%20as%20it%20has%20been%20until%20now>
- Bécares, Barbara. 4 October 2021, "The former Facebook employee who leaked its internal documents: they allow hate and disinformation, they just want to make money". <https://www.genbeta.com/actualidad/exempleada-facebook-que-ha-filtrado-sus-documentos-internos-permiten-odio-desinformacion-solo-quieren-ganar-dinero>
- Benesch, S. 2020. "But Facebook's Not a Country: How to Interpret Human Rights Law for Social Media Companies. Yale Law School. <https://digitalcommons.law.yale.edu/cgi/viewcontent.cgi?article=1004&context=jregonline>
- Camacho, David. Special Session at the International Joint Conference on Neural Networks (IJCNN) 2022. In conjunction with IEEE World Congress on Computational Intelligence (WCCI) 2022. <https://wcci2022.org/>
- Channels, Katie. Business Insider 17 September 2021 "Facebook's AI moderation is unable to interpret all languages, leaving users in some countries more exposed to harmful content." <https://www.businessinsider.es/facebook-tiene-problemas-moderar-contenidos-no-sean-ingles-933247>
- Castillo, Alba. 15 April 2021 "Facebook, investigated by the European Union for the leak of millions of its users' phone numbers". <https://www.20minutos.es/noticia/4658218/0/facebook-investigada-por-la-union-europea-por-la-filtracion-de-millones-de-numeros-de-telefono-de-sus-usuarios/>
- Clarín, 19 September 2021. "Apple almost removed Facebook from the App Store: human traffickers used the social network to sell victims". The Vanguard <https://www.lavanguardia.com/tecnologia/20210919/7731847/apple-estuvo-punto-eliminar-facebook-app-store-trafficantes-personas-usaban-red-social-vender-victimas-pmv.html>
- Delgado, Monica. 10 November 21 "Graham Mudd removes from Facebook, advertising linked to sensitive issues" Consumer tick. <https://www.consumotic.mx/tecnologia/elimina-facebook-publicidad-ligada-a-temas-sensibles/>
- The Debate. 25 October 2021 "Facebook ignored reports warning of hate speech and human trafficking" <https://www.eldebate.com/tecnologia/20211025/facebook-ignora-informes-alertaban-sobre-difusion-mensajes-odio-trata-personas.html>
- Erard, Gabriel. 6 September 2021 "Facebook labels black people as primates" [Iper textual](https://hipertextual.com/2021/09/facebook-discriminacion-inteligencia-artificial) and the AI. <https://hipertextual.com/2021/09/facebook-discriminacion-inteligencia-artificial>
- Erard, Gabriel. 5 September 2021 "Facebook wants the artificial intelligence of the future to be able to "forget" irrelevant data." <https://hipertextual.com/2021/05/facebook-inteligencia-artificial-datos-irrelevantes>
- Erard, Gabriel. 16 June 2021 "Facebook intends to rely on reverse engineering to detect 'deepfakes' and trace their origin" <https://hipertextual.com/2021/06/facebook-detecta-origen-deepfake>
- Erard, Gabriel. 6 September 2021 "Facebook AI labels black people as "primates"" <https://hipertextual.com/2021/09/facebook-discriminacion-inteligencia-artificial>
- Fernández de Lara, Carlos. 4 June 2021 "Facebook to uphold Trump veto for 2 years; then revisit its case." <https://www.forbes.com.mx/facebook-mantendra-veto-a-trump-por-2-anos-luego-volvera-a-revisar-su-caso/>
- Fiter, Miguel and Encabo, Ignacio. 18 November 2019. International Artificial Intelligence Congress: "this is how Facebook is working with artificial intelligence to identify hate" The Independent, in science and technology. <https://www.elindependiente.com/futuro/2019/11/18/asi-trabaja-facebook-con-la-inteligencia-artificial-para-identificar-al-odio/>
- Fjeld, Jessica and Achten, Nele and Hillgoss, Hannah and Nagy, Adam and Sri Kumar, Madhulika, Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI (January 15, 2020). Berkman Klein Center Research Publication No. 2020-1, Available at SSRN: <https://ssrn.com/abstract=3518482> or <http://dx.doi.org/10.2139/ssrn.3518482>
- Floridi, L (2019), «Translating principles into Practices of Digital Ethics: Five Risks of Being Unethical». *Philosophy and Technology*, 32.
- Gascón, Marta. 28 October 2021 "What are the 'Facebook Papers': keys to understanding the new scandal that Zuckerberg's social network is facing" <https://www.20minutos.es/tecnologia/actualidad/que-son-los-facebook-papers-claves-para-entender-el-nuevo-escandalo-al-que-se-enfrenta-la-red-social-de-zuckerberg-4868034/>

- Gascón, Marta. 28 October 2021. "Instagram Kids: what is this social network project for children and why Facebook has now paralysed it". <https://www.20minutos.es/tecnologia/ciberseguridad/instagram-kids-que-es-este-proyecto-de-red-social-para-ninos-y-por-que-ahora-facebook-lo-ha-paralizado-4835999/>
- Genbeta. 5 October 2021 "The ex-Facebook employee who leaked its internal documents: they enable hate and misinformation, they just want to make money."
- Gilbert, Ben. 15 October 2021. "Facebook prepares an artificial intelligence that will track your every move." Business insider. <https://www.businessinsider.es/facebook-prepara-ia-rastrear-todos-movimientos-948387>
- Gómez Maseri, Sergio. 10 October 2021, "Facebook: the story of a growing scandal". <https://www.eltiempo.com/mundo/eeuu-y-canada/facebook-la-historia-de-un-escandalo-cada-vez-mayor-624207>
- González, Gabriela. 28 October 2021, "Facebook's Metaverse is a virtual world ideal for spending money and which there is no way to enter yet" <https://www.genbeta.com/actualidad/metaverso-facebook-mundo-virtual-ideal-para-gastar-dinero-al-que-no-hay-forma-entrar-todavia>
- Gonzalez, Gabriela. 8 April 2016 Last updated, 11 March 2021 "85 out of every 100 adults in the world use a Facebook-related account" <https://hipertextual.com/2016/04/uso-de-facebook-en-el-mundo>
- Hernandez, Noelia. 1 July 2021 "Facebook puts its open-source artificial intelligence at the service of science" https://www.lespanol.com/invertia/disruptores-innovadores/innovadores/tecnologicas/20210701/facebook-inteligencia-artificial-codigo-abierto-servicio-ciencia/593191230_0.html
- House of Commons 2019 Digital, Culture, Media and Sport Committee, Disinformation and Fake news: final report.
- The Nation. 25 October 2021 "Apple threatened to pull Facebook app" <https://www.lanacion.com.ar/agencias/apple-amenazo-con-retirar-la-aplicacion-de-facebook-nid25102021/>
- Jobin, Anna. Principles in Ethics Guidelines & Policy Perspectives Summer School: ARTIFICIAL INTELLIGENCE, ETHICS & HUMAN RIGHTS Universidad Politécnica Madrid, 22 June 2022 Dr. Anna JOBIN
- La voz de Galicia 5 October 2021 "Mark Zuckerberg apologises for the fall of Facebook, WhatsApp and Instagram" <https://www.lavozdegalicia.es/noticia/sociedad/2021/10/05/mark-zuckerberg-pide-perdon-caida-facebook-whatsapp-instagram/00031633435695869723899.htm#:~:text=%C2%ABFacebook%C2%20Instagram%C2%20WhatsApp%20and,e n%20his%20popular%20social%20network%20>
- Lacort, Javier. 27 September 2018. Updated 15 November 2018 "This is how Facebook makes money: the other advertising giant has two aces up its sleeve." <https://www.xataka.com/empresas-y-economia/asi-como-gana-dinero-facebook-otro-gigante-publicitario-tiene-dos-ases-manga>
- Latonero, Mark and Agarwal, Aaina. Human Rights Impact Assessments for AI: Learning from Facebook's Failure in Myanmar, Carr Center for Human Rights Policy, 2021, <https://carrcenter.hks.harvard.edu/files/cchr/files/210318-facebook-failure-in-myanmar.pdf>
- Lopez, Miguel. 2 November 2021 "Meta will delete the facial recognition data of more than a billion people used on Facebook: more knockout blows from Zuckerberg" https://www.xataka.com/privacidad/meta-borrara-datos-reconocimiento-facial-mil-millones-personas-usados-facebook-golpes-efectos-zuckerberg?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+xataka2+%28Xataka%29
- McKay, Tom. 8 January 2017 "What really happened to that artificial intelligence that Facebook shut down because it had created its own language". <https://es.gizmodo.com/lo-que-realmente-sucedio-con-aquella-ia-que-facebook-ap-1797424875>
- Merino, Marco. 20 Oct. 2021 "Mark Zuckerberg, indicted in his personal capacity in Cambridge Analytica case: a conviction would expose him to heavy penalties."
- Miguel Trula, Esther. 29 July 2021 "Facebook thinks the networks of the future will look like Ready Player One. And that's why it's creating the Metaverse" <https://magnet.xataka.com/en-diez-minutos/facebook-cree-que-redes-futuro-se-pareceran-a-ready-player-one-eso-esta-gestando-metaverso>
- Mollejo, Veronica. 15 May 2018 "Facebook turns to artificial intelligence to fight hate" <https://www.redbull.com/es-es/tecnologia-facebook-inteligencia-artificial-odio>
- Nielfa, Jaime S. 1 January 2022 "7 Artificial Intelligence Tools to Optimise Facebook Ads Campaigns and The Fall of Organic Traffic on Facebook Ads" <https://scoreapps.com/blog/es/facebook-ads-herramientas-ia/>
- Notimundo. November 2021. Facebook was concerned about hate speech violations in Assam ahead of the 2021 elections.

- <https://noticiasmundomundo.news/facebook-estaba-preocupado-por-las-violaciones-del-discurso-de-odio-en-assam-antes-de-las-elecciones-de-2021/>
- Olaskoaga, Andrés. 5 January 2021 "Facebook, experiments with artificial intelligence that listens, sees and remembers your personal information" Muy interesante, Ciencia y tecnología. <https://www.muyinteresante.com.mx/ciencia-tecnologia/facebook-inteligencia-artificial-recuerda-informacion-personal/amp/>
- Pastor, Javier 13 April 2018. "Mark Zuckerberg: artificial intelligence will censor, hate messages, before they are published". <https://www.xataka.com/robotica-e-ia/zuckerberg-y-el-papel-de-la-inteligencia-artificial-como-moderadora-y-censuradora-de-contenidos>
- Pastor, Javier. 15 February 2017 "Los algoritmos que control nuestras vidas son tan injustos como nosotros mismos" <https://www.xataka.com/privacidad/los-algoritmos-que-controlan-nuestras-vidas-son-tan-injustos-como-nosotros-mismos>
- Perez, Enrique. 14 October 2021 "The next (and somewhat disturbing) challenge for Facebook's AI is to understand and remember the world as seen through our eyes". <https://www.xataka.com/privacidad/siguiente-algo-perturbador-desafio-ia-facebook-entender-recordar-mundo-visto-a-traves-nuestros-ojos>
- Perez, Enrique. 27 September. 2021 "Facebook halts creation of 'Instagram Kids' for under-13s after controversy with toxicity report." <https://www.xataka.com/aplicaciones/facebook-detiene-creacion-instagram-kids-para-menores-13-anos-polemica-informe-toxicidad>
- Perez, Enrique. 4 October 2021, Updated 8 October 2021 "Facebook is going through one of its most delicate moments after weeks of leaks: everything that has come to light so far". <https://www.xataka.com/empresas-y-economia/facebook-atravesa-uno-sus-momentos-delicados-semanas-filtraciones-todo-que-ha-salido-a-luz-ahora>
- Pérez de las Heras, Beatriz (2022). Los derechos digitales en la Unión Europea: del liberalismo económico a la protección jurídica. Revista General de Derecho Europeo, ISSN-e 1696-9634, N°. 57.
- Rodriguez, Pablo. 14 December 2021 "Why no scandal will topple Zuckerberg at Meta: more than 200 shareholders have been trying to disempower him since 2018 without success" <https://www.xataka.com/empresas-y-economia/que-no-hay-escandalo-que-tumbe-a-zuckerberg-meta-200-accionistas-tratan-restarle-poder-2018-exito#:~:text=No%20hay%20esc%C3%A1ndalo%20de%20Facebook,acumulado%20en%20todos%20estos%20a%20a%C3%B1os.>
- Rus, Chistian 9 December 2020 "US sues Facebook for anti-competitive practices: asks it to divest itself of Instagram and WhatsApp" <https://www.xataka.com/empresas-y-economia/estados-unidos-demanda-a-facebook-practicas-anticompetitivas-pide-que-se-desvincule-instagram-whatsapp>
- Rus, Chistian. 13 August 2021 "researchers have created a "master face" capable of fooling facial recognition systems"
- Rus, Chistian. 14 June 2021 "Facebook says its artificial intelligence can mimic your handwriting just by looking at a word" <https://www.xataka.com/robotica-e-ia/facebook-dice-que-su-nueva-inteligencia-artificial-puede-imitar-tu-escritura-solo-ver-palabra>
- Véliz, Carissa 2021 Privacy is Power: Why and How you should take back control of your data. Ed. Melville House.
- Week. Technology. 14 October 2021 "Facebook adjusts its policies to act against mass attacks and sexual harassment" <https://www.semana.com/tecnologia/articulo/facebook-ajusta-sus-politicas-para-actuar-contra-los-ataques-masivos-y-el-acoso-sexual/202110/>
- Well, Georgia; Seetharaman, Deepa and Horwitz, Jeff. 14 September 2021 "Revealing documents: Facebook knew Instagram is toxic for teenage girls" The Wall Street Journal. <https://www.lanacion.com.ar/el-mundo/documentos-reveladores-facebook-sabia-que-instagram-es-toxico-para-las-adolescentes-nid14092021/>
- Wells, Georgia; Seetharaman, Deepa and Horwitz, Jeff. 18 October 2021 "Facebook says AI will clean up the platform. Its own engineers have doubts" <https://www.latercera.com/que-pasa/noticia/facebook-dice-que-la-ia-limpiara-la-plataforma-sus-propios-ingenieros-tienen-dudas/J3EAKYOGPJDPJNHILN3G2BQVM4/>