Communications Processing

People have more choices than ever in the market for electronic communications services. Increasingly, they are choosing services that depend on processing electronic communications data. For example, many messaging services process communications data to provide "smart" features and to keep users safe. These services often rely on artificial intelligence (AI) and machine learning, which involve teaching computers to take actions without being explicitly programmed. Doing so requires the processing of content and metadata of messages, which enables the services to "learn" from how people use the service.

Facebook is concerned that the proposed ePrivacy Regulation will impede the development of technologies based on AI and machine learning. We describe some of these technologies below and make recommendations for how the ePrivacy Regulation can be amended in order to promote the continued development of such technologies, while also protecting the confidentiality of people's communications.

Smart messaging features

Unlike basic SMS messaging services — which merely transmit messages — Facebook's Messenger is defined by its "smart" features that depend on AI and machine learning. For example, M Suggestions (currently available in e.g. FR, ES, UK, IRL) can recognise when people are messaging about getting together and help coordinate a "plan" by suggesting the creation of an event or plan.

Facebook Messenger also provides features such as translations, suggested replies, text-to-speech (a crucial feature for the visually impaired), favourite contacts (which are based on an understanding of whom a person communicates with most frequently), link previews (which are rendered when a URL is recognised in a message thread), and more.

These features rely on the processing of message content and metadata. For example, translations are developed through Natural Language Processing (NLP) and Natural Language Understanding (NLU), which rely on the ability of machines to read and understand human language. We're currently processing 2 billion text translations per day, and 800 million people see translations each month.

Features that keep people safe

Facebook scans communications data to keep people safe when they're using our services, as part of our efforts to detect child pornographic material; child grooming activities; terrorist material; non-consensual intimate images (sometimes called "revenge porn"); financial scams; and spam. Facebook is expected by many policymakers and citizens alike to invest more in machine learning based on the

scanning of communications data to address violating content.

Protecting confidentiality

Facebook agrees with the main goal of the proposed ePrivacy Regulation: protecting the confidentiality of people's online communications. As drafted, however, the Regulation is unlikely to improve existing confidentiality protections against unlawful surveillance. Moreover, it would treat the features described above — the features people sign up for when they choose to use Messenger — as *threats* to confidentiality.

Facebook strongly believes in transparency and control; they're fundamental to Facebook's services. But an over-reliance on consent - which we see in Article 6 - will undermine its value in delivering effective transparency and control. Moreover, as explicitly recognised in the GDPR, other legal bases for data processing, such as legitimate interest or contractual necessity, might be more appropriate in specific cases. In such cases, companies need to provide clear information to users, offer the right to object and provide appropriate safeguards in line with the GDPR.

Recognising additional legal bases is appropriate because:

- When used too frequently, consent is a weak protection for privacy - people click away prompts and notices without paying proper attention. Consent should be used sparingly and only when it's actually meaningful.
- "[...] there are simply too many consent requests for an individual user to consider, watering down the psychological effect of being confronted with a consent transaction. Jolls and Sunnstein (2006, p.212), for instance, have found that consumers learn to tune out messages that they see often. [...] The Dutch implementation of article 5(3) of Directive 2009/136/EC (the 'cookie law') is a good example of how excessive consent requests have led to 'consent fatigue'." Source: Schermer et al., "The Crisis of Consent: How Strong Legal Protection May Lead to Weaker Consent in Data Protection," *Ethics & Technology* (March 2014).
- Facebook needs to process communications data to protect its users from terrorists, spammers, child abusers and fraudsters who attempt to abuse our services. None of these will give consent.
- Two-party consent will not allow for translation or smart messenger services upon request of the receiver when the sender does not give his consent. Translation would then have to take place under GDPR after reception of the message.
- An over-reliance on consent will impede the development of AI and the features relying on it.

Example: identifying terrorist content

Facebook processes images and text to help detect content that violates Facebook's policy prohibiting the promotion of terrorist activity. If rules on confidentiality of

communications would require consent for such scanning from both the recipient and the sender, such operations would, in many cases, be impossible.

Facebook's policy on terrorist activity

There's no place on Facebook for terrorism. We remove terrorists and posts that support terrorism whenever we become aware of them. When we receive reports of potential terrorism posts, we review those reports urgently and with scrutiny. And in the rare cases when we uncover evidence of imminent harm, we promptly inform authorities.

Reporting & Enforcement

Today, 99% of the ISIS and Al Qaeda-related terror content we remove from Facebook is content we detect before anyone in our community has flagged it to us, and in some cases, before it goes live on the site. We do this primarily through the use of automated systems like photo and video matching and text-based machine learning, which is paired with human review. Once we are aware of a piece of terror content, we remove 83% of subsequently uploaded copies within one hour of upload. Processing images and text is a fundamental part of these operations.

- *Image matching:* When someone tries to upload a terrorist photo or video, our systems look for whether the image matches a known terrorism photo or video. This means that if we previously removed a propaganda video from ISIS, we can work to prevent other accounts from uploading the same video to our site. In many cases, this means that terrorist content intended for upload to Facebook simply never reaches the platform.
- Facebook, Microsoft, Twitter and YouTube have begun sharing hashes – i.e. unique digital fingerprints – of the most extreme and egregious terrorist images and videos we have removed from our services — content most likely to violate all of our respective companies' content policies. Participating companies can add hashes of terrorist images or videos that are identified on one of our platforms to the database. Other participating companies can then use those hashes to identify such content on their services, review against their respective policies and definitions, and remove matching content as appropriate.
- *Language understanding:* Facebook has recently started to experiment with using AI to understand text that might be advocating for terrorism. We're currently experimenting with analyzing text that we've already removed for praising or supporting terrorist organizations such as ISIS and Al Qaeda so we can develop text-based signals that such content may be terrorist propaganda. That analysis goes into an algorithm that is in the early stages of learning how to detect similar posts. The machine learning algorithms work on a feedback loop and get better over time.
- *Removing terrorist clusters:* We know from studies of terrorists that they tend to radicalize and operate in clusters. This offline trend is reflected online as

well. So when we identify Pages, groups, posts or profiles as supporting terrorism, we also use algorithms to "fan out" to try to identify related material that may also support terrorism. We use signals like whether an account is friends with a high number of accounts that have been disabled for terrorism, or whether an account shares the same attributes as a disabled account.

Legal bases

Additional legal bases are needed (other than consent) to allow for scanning and machine learning for the purpose of terrorist content removal. A deliberate violator of Facebook's policy on violent extremism would not consent to their content being scanned for this purpose.

Legitimate interest or contractual necessity could serve as appropriate legal bases for this purpose. Facebook would then assess and balance both the interest in processing the communications data as well as the appropriate safeguards to be provided to protect the interests and the fundamental rights of the user.