# GDPR, AI and Machine Learning in the Age of Data Privacy

INTOUCH INTERNATIONAL™

## Executive Summary

Over the past few years, there's been a massive cultural and legal shift in the way consumers view and secure their personal data online. As of May 2018, the implementation of the General Data Protection Regulation (GDPR) allows European consumers more "ownership" of their data and the ability to remove that data from marketing systems.

Marketers leveraging artificial intelligence (AI) tools might be especially impacted by users hesitant to share their data, since AI models work much better when they incorporate larger, more representative data sets with personal information.

With 88% of U.S. internet users concerned about the privacy and security of their personal information on the web, and 67% of U.S. internet users advocating for GDPR-type privacy laws, it's time to take a look at the impact of this shift on U.S. advertisers as well.

Key questions about data, privacy and mandating consumer "opt-ins" have been raised after the Cambridge Analytica scandal and other recent Facebook and Google breaches.

- What rights do we have as consumers?
- What data is protected under GDPR?
- Will AI models be limited if users don't want to provide data?

This POV will answer these questions and provide recommendations for marketers leveraging AI in the new age of data privacy and security.


## Background: See Something, Say Something

The GDPR is already impacting the way large companies do business in Europe.

Google has been under fire in Europe for allegedly tracking users' locations, even while Location History was turned off.

A September data breach of Facebook, in which up to 90 million users' data was compromised, showcased the company's new way of operating during attacks on its platform. In the past, Facebook could have quietly implemented and tested a fix before going public. Instead, Facebook announced the security issue, patch and its ongoing investigation within 72 hours.

GDPR outlines a timeframe in which companies like Facebook have to provide detail about hacks and breaches. Failure to do so can result in violation and penalty up to 4% of their global revenue.

In October, Google announced it would shut down Google+ because of a similar data breach. It's reasonable to assess that Google's easiest and simplest path was to shut down the platform rather than refine it.

These recent situations act as a great case study for future compromised stewards of data — being quiet in the face of exposed user data is no longer an option.

Even though GDPR doesn't affect U.S.-only companies directly right now, many of the GDPR mandates may be a requirement in the United States as well in the future. If a U.S. company is marketing anywhere in Europe, these new rules still apply.

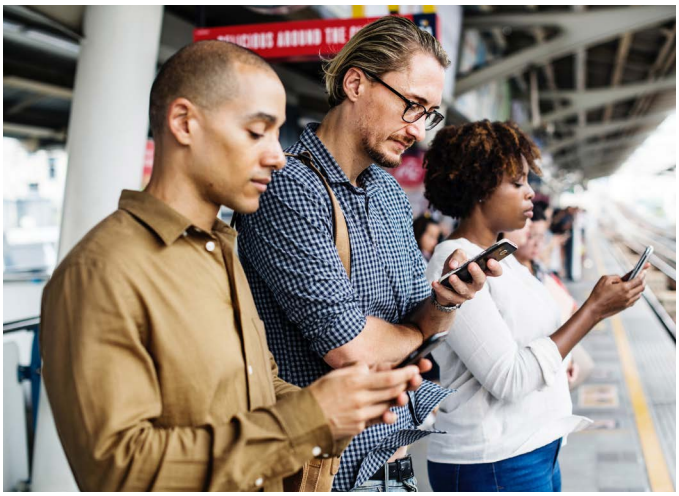
## GDPR Considerations

### OPT-INS

First and foremost, GDPR requires companies to gather consent from users to store their personal data for any use (analytics, profiling, AI, etc.). Personal data can be anything from a name, address, photo, email address, medical information, to an IP address.

Marketers need to make it very clear on their websites what data they are collecting, how they'll use it — in plain language — and how users can opt out in the future. This opt-in process should shape best practices for UX moving forward.

The primary goal for GDPR is for consumers to be more informed about how their information is being used, and become more intentional about approving its use.



### RIGHT TO BE FORGOTTEN

A main premise of GDPR is that users have the right to erase the data companies have stored about them and request proof of the erasure. This is important for AI models, because if companies are using that data to train AI, they'll have to remove any requested records from the model.

AI processors may have to become more granular about storing data and avoid pooling it unless it is anonymized (more on that below). This could require significant infrastructure development to support the changes.

An important clarification: if a user previously consented to share their data, then an AI model is trained but consent is subsequently removed, the AI model may not need to be retrained with the new dataset. However, any future training would need to avoid using removed data.

### RIGHT TO EXPLANATION

The idea that GDPR requires companies to be able to explain to users how their data is collected and used, if it is used to automate "significant" decisions, is somewhat controversial. Some legal scholars doubt the legality and feasibility of such a right. Ultimately, it's better to err on the side of transparency and caution when interpreting GDPR Articles 13, 14 and 15, and supply this information where possible.

As it applies to AI, this means users have a right to know when they are the subject of an AI model and how the data is being used. This can sometimes be difficult because of the "Black Box" problem in AI — when we know the inputs and outputs of a system, but not always how the data is used in training.

To satisfy this right, marketers should give consumers a simple explanation about how the data is used when possible, explaining the logic and significance of the AI model.

- What points of personal data are used?
- What are the technical components (neural net, logistic regression, etc.)?
- What is the significance or consequence of using the data?

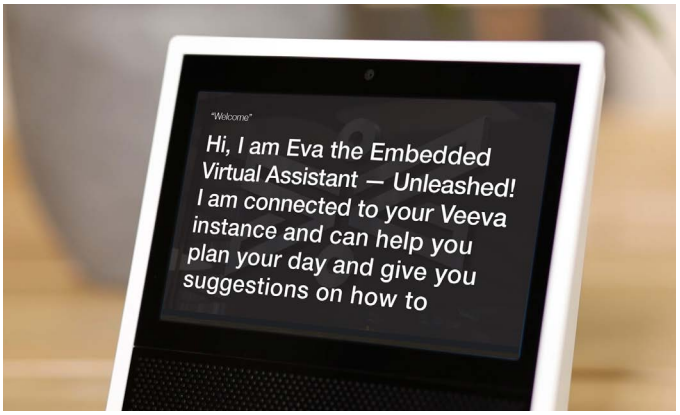
This approach allows users to be more informed about what their consent accomplishes without giving away any proprietary information or confusing consumers with extremely complicated jargon.

### THE RIGHT TO OPT OUT OF AUTOMATED DECISION MAKING

Article 22 of the GDPR states that if the AI is responsible for a significant outcome (approving an auto loan, determining mortgage rates, calculating healthcare/insurance outcomes), consumers have a right to have humans become involved. AI cannot be the only determining factor in these decisions.

If AI is used to assist HCPs in any decision making, they'll need to ensure the AI is not the sole factor in any actionable determinations.

This is especially important in healthcare. Intouch's Cognitive Core engine is intended to be *augmented intelligence* rather than traditional AI, because it augments what humans can do through automation and decision-intelligence support.




Effective AI does not replace human judgment, especially when livelihoods and health are at stake. Even in more benign tasks

like social media monitoring and moderation, AI tools can be extremely helpful and efficient, but ultimately humans still need to be involved.

#### ETHICAL USE OF DATA

Data providers are now responsible for unethical use of data by third parties: in other words, GDPR holds marketers accountable.

For example, if a healthcare organization sends data to an AI provider, the healthcare organization is responsible for ensuring that the data is used in a responsible way.

This may have helped prevent Cambridge Analytica from using Facebook user data unethically. Companies paying closer attention to how data is used moving forward is a good thing.

## Insights

In the short term, GDPR has resulted in some companies making it easier for consumers to view and download their data, even in the United States. For example, on Facebook, new tools allow users to download all their data and see what Facebook "knows" about them. In the longer term, companies must determine how to store and return data back to consumers and develop infrastructure to accommodate.

#### ANONYMIZATION AND PSEUDONYMIZATION

One way that marketers can continue to use "personal data" to an extent is to anonymize or pseudonymize the data. When data is irreversibly anonymized, it is no longer considered "personal data" and may continue to be used for AI models.

Pseudonymized data (replacing identifying characteristics of data with a pseudonym or non-identifying value) remains categorized as personal data but still provides limited protection.

There are several ways to anonymize or pseudonymize data, including:

- Randomization
- Generalization
- Masking

Considerations for anonymizing data include assessing the risk that a third-party "bad actor" could identify subjects using the anonymized data. Data minimization and collection techniques can help limit this risk.

## Conclusion

This cultural shift may result in new ways of innovating for AI. In the years to come, consumers will get to enjoy being more intentional about how their data is used. Marketers will continue to face the need to adapt to this shift both culturally and legally.

Just like a camera with more pixels will showcase higher photo detail, AI models with more data points provide a clearer picture than those with fewer data points.

Developers must find methods of bridging the trust gap with consumers, helping them feel more comfortable about contributing their data for better models that help both marketer and consumer.

For pharma, which has been known to have some trust issues with consumers, GDPR could have another benefit by allowing brands to foster trust.

The healthcare industry is uniquely positioned to adapt to the new rules and regulations. Regulations like GDPR can make it harder for companies to advertise due to the extra scrutiny placed on their data collection and retention, but pharma is no stranger to a tough regulatory environment. This means compliant U.S. pharma marketers may already have an advantage as it comes time to apply these same standards domestically.



*Author: Andrew Grojean, Innovation Manager, Intouch Solutions*

**Want to learn more about GDPR, AI and Machine Learning?**
Contact Ariel Salmang at +44 20 32392129
ariel.salmang@intouchint.com
**intouchint.com**