<https://medium.com/@RocketMeUpCybersecurity/privacy-preserving-machine-learning-how-to-train-models-without-compromising-data-866a825097d2>

<https://www.smarsh.com/blog/thought-leadership/generative-AI-and-data-privacy-the-challenge-of-PII-use-in-training-data-sets>

<https://www.zscaler.com/blogs/product-insights/ai-cybersecurity-navigating-gdpr-privacy-laws-and-risk-management>

The three main factors are: AI data volume needs, individual privacy rights, and data privacy laws.

* AI needs a lot of data to work with, and organizations often choose to make PII a part of the training set for the AI. This choice is at the very least not very far off from a data breach, thus making this a major issue. There are also other current standards that this may violate, including the principles of purpose limitation and data minimization.
* Under many current data privacy laws, individuals have the right to have their PII changed, accessed, and deleted from the storage of any organization in possession of such data. It appears however that once PII is used in training an AI model, it is remarkably difficult, if not impossible, to remove that and only that data point. Even if this is done, it is likely that the AI model’s performance will degrade as a result of that removal. Some of the laws that govern these standards are: EU’s GDPR, California’s CCPA/CPRA, and even HIPAA.

These problems are solved at least in part by privacy preserving machine learning (PPML), which aims to essentially facilitate the training of these AI models without compromising data security and privacy. Some of the key principles are:

* Data confidentiality
  + Encrypt or anonymize (or even redact) all data used to train an AI model.
  + Minimize the amount of PII used in training to only what is strictly necessary.
* Model security
  + Models must be made to resist reverse engineering so as to protect all data, especially PII, from a breach.
  + Models must remain effective in the presence of an attack to resist breaches.
* Compliance and Ethics
  + Comply with data privacy laws, such as GDPR, HIPAA, CCPA/CPRA, or others.
  + Data must always be handled through proper channels to ensure security. How a company handles its users’ data will directly affect if and how much its users trust the company.

Techniques:

* Differential privacy
  + Add controlled noise to model output or data so that an AI model can be helpful to the user without exposing PII.
  + Alternatively, a privacy budget can be adhered to, which limits how much information about an individual can be revealed. This entails more risk, and proper value must be placed on each data point.
* Federated Learning: train models locally, and only share model updates rather than transferring all sensitive data to a central server.

Overall, it appears that the main challenges with managing PII ethically in AI models relate to how the data is used in AI models. Since they use massive data sets, companies often train the models using PII they have, however this presents a huge security risk. It’s not easy to fix this though, as specific data points can’t always be targeted and removed from the AI’s data pool without causing performance degradation. This makes compliance with data privacy law tricky.

The principles of PPML are going to be very necessary to lean on for our policy decisions. These ideas outline how to build an AI model from the ground up so that it is both effective and compliant. We will need to focus on how to change what already exists, but it would be very good to focus on how models should be trained for the future.