IP Protections for Data and Machine Learning Models

September 2021

Key Elements and Terminology

- **Two key ingredients to create a machine learning model: the learning algorithm and the training data.**
  - The learning algorithm is applied to training data to train a model, which is the output of the process.
  - Many learning algorithms are open source or commoditized, so good training data can be the more scarce ingredient.

IP Rights (or Lack Thereof)

- Patents. Both learning algorithms and trained models are likely to be mathematical formulas that are generally not subject to patent protection as they are "abstract ideas".
  - *See* Alice Corp. v. CLS Bank International, 573 U.S. 208 (2014).
- Copyrights. Learning algorithms, models, and factual data used to train models generally will not be copyrightable.
  - *See* Feist Publications, Inc., v. Rural Telephone Service Co., 499 U.S. 340 (1991).
  - The software implementation of a model is copyrightable, but the software implementation may be trivial in many cases.
- Database rights. Data may be protected by database rights in Europe, but they are not applicable in the U.S. and most other jurisdictions.
  - See "*Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.*"

Contractual Protections

- Data License Restrictions
  - Prohibit recipients from disclosing data
  - Prohibit recipients from using data for training models
- Allocation of "ownership" over trained models.
  - Ownership concepts are vague and cumbersome as applied to trained models because there is usually no applicable IP framework.
  - Neither the provider of training data nor a learning algorithm automatically "owns" trained models.
  - Trained models are unlikely to be "derivative works" or “Derived Data".
  - Rights and restrictions can be established by **a contract drafted with the unique dynamics of machine learning in mind**.

Practical Security

- A trained model is often a classic trade secret, like the recipe for Coke or Thomas's English Muffins.
  - A competitive solution is to keep the model from competitors.
- Unique data that yields a trained model with a standard learning algorithm can also be a trade secret
- Practical measures to maintain confidentiality:
  - Run a trained model with a Saas architecture behind an API.
  - Provide collaboration partners with access to data only in a data platform where you can prevent export.
  - Classic contractual protections: monitor and audit.