5.4 Legality

Authors: Neel Guha, Peter Henderson, Lucia Zheng, Mark Krass, Daniel E. Ho

In this section, we describe how US law may influence, constrain, or foster the creation and use of foundation models.[115] We note that the legal landscape surrounding algorithmic tools remains uncertain. We highlight issues pertaining to (1) model training, (2) liability for model predictions, and (3) protections for model outputs.

Though understanding how the law affects foundation models is crucial, it is important to recognize that the law cannot be the only lens through which we evaluate the construction, maintenance, and use of foundation models. Ethical frameworks are necessary to understand where legally permissible applications of foundation models may still be ill-advised for the harms they inflict and are discussed in more depth in §5.6: ETHICS and §5.1: FAIRNESS. Studying the potential for misuse and possible security concerns (see §5.2: MISUSE and §4.7: SECURITY) is critical for preventing harmful outcomes *ex ante*, as opposed to the *ex post* treatment that legal mechanisms often provide.

5.4.1 Training.

Training foundation models will require accumulating vast amounts of multi-modal data, raising questions around data collection and data use.

First, the ability for model creators to grow datasets via web scraping will be governed by the manner in which courts will interpret terms of service provisions and, notably, the U.S. Computer Fraud and Abuse Act (CFAA), which criminalizes accessing a server "without authorization" [Wajert and Rottman 2019]. Courts are in conflict on these questions, and recent cases have sought to clarify the circumstances under which web scraping may be barred.[116] The restrictiveness of data access would fundamentally affect the diversity of data practitioners can use to train foundation models [Levendowski 2018].

Second, much of the data contained in training sets will be copyrighted and potentially protected by intellectual property law. However, copyright law recognizes exceptions when individuals may be permitted to use copyrighted material.[117] Some scholars believe that the legal permissibility of training datasets will largely rest on whether courts interpret the process of model training as "transformative" under fair use doctrine [Lemley and Casey 2020]. Though the question of what qualifies as transformative is highly context dependent, the general rule is that transformative uses are those "that add something new, with a further purpose or different character, and do not substitute for the original use of the work" [Office 2021]. Already, the recently released Github Copilot tool is bringing these arguments to the fore [Gershgorn 2021].

Finally, some training datasets may run afoul of privacy laws. Illinois, for instance, enables individuals to sue for improper collection or use of biometric data (e.g., retina or iris scans, fingerprints, voiceprints, or scans of hand or face geometry).[118] Foreign privacy laws like the E.U.'s General Data Protection Regulation (GDPR) — which will affect American model creators if datasets contain information from E.U. citizens — would require data subjects to be informed about the purpose of data collection. Further issues could arise for laws like the California Consumer Protection Privacy Act (CCPA), which provide individuals with a "right to be forgotten," raising questions as

[115] Our perspective here centers on US law and legal frameworks. Discussions of the implications of foundation models with respect to other countries may consequently take different perspectives.

[116] *Van Buren v. United States*, 141 S.Ct. 1648 (2021).

[117] *See, e.g.*, , 17 U.S.C §107 to 112.

[118] IBM is the defendant in a current class action alleging that IBM's collection and use of this data (including for machine vision purposes) violates this statute. See Class Action Complaint at 2, Vance v. Int'l Bus. Machines Corp., No. 20 C 577 (N.D. Ill. filed Jan. 24, 2020).

to whether model creators will need to "remove" training data from models [Villaronga et al. 2018; Ginart et al. 2019].

5.4.2 *Output liability.*

Though foundation models themselves are task agnostic, fine-tuned models — or the representations learned by foundation models themselves — may be used for traditional prediction tasks. Where these tasks form components of larger decision-making systems, foundation models will thus influence actions, decisions, or policies. When these result in harm, model creators — and the individuals operating them — may be legally responsible.

Embedding foundation models in physical systems (e.g., self-driving cars, electric grid management, medical diagnostics, etc.) may result in physical harm to individuals. Here, courts will likely resolve questions of liability under tort doctrine [Lemley and Casey 2019; Selbst 2020]. Key open questions include the interplay between the liability of users, foundation model providers, and application developers, as well as the standards courts will use to assess the risk profile of foundation models. Deployments in particularly sensitive domains (e.g., medicine) will require regulatory approval, and the development of standardized processes to assess safety [Wu et al. 2021g].

Fine-tuned foundation models that classify individuals in ways that correlate with protected attributes (e.g., race, gender) may face challenges under civil rights laws. Scholars have noted that claims for disparate treatment resulting from foundation models may be brought in the context of hiring, housing, or credit lending [Gillis and Spiess 2019; Scherer et al. 2019]. Exactly how courts will adjudicate these issues is far from clear. Scholars have noted for instance, that the courts' traditional views on "discrimination" would actually prevent machine learning practitioners from implementing many algorithmic fairness techniques [Xiang 2021; Ho and Xiang 2020].[119]

U.S. law recognizes special privileges and limits on governmental entities. Thus, the use of foundation models by governmental entities — at a local, state or federal level — will implicate special considerations, in addition to equal protection claims. The use of models for risk assessment — or in other settings which result in a deprivation of life, liberty, or property — will invite procedural due process claims.[120] When models are used by administrative agencies (e.g., the Environmental Protection Agency) for instance, plaintiffs may allege that such use violates basic standards of due process, reasonableness / non-arbitrariness, and transparency.

5.4.3 *Legal protections for outputs.*

Model outputs — and by extension the model creators responsible for the models — may also be afforded certain legal protections. First, content produced by generative models may implicate free speech issues. The extent to which courts will find First Amendment protections for machine generated content is unclear. Scholars have discussed a number of open questions, including whether "AI speech" is protected [Massaro et al. 2016] or if model outputs are in effect the human programmer's speech [Kajbaf 2019]. Others have noted the possibility of disclosure requirements (akin to safety disclosures for pharmaceutical drugs or other substances), also implicating speech doctrine, under which models would be forced to share with listeners that their content is machine generated [Lamo and Calo 2019]. These issues could have wide ranging consequences, affecting whether individuals can use foundation models to mass produce speech, or whether model creators could be held liable for content generated by foundation models.

[119] For more information on how models may embed certain biases, see §5.1: FAIRNESS.

[120] Procedural due process recognizes that plaintiffs usually have certain basic rights during any deliberation that will deprive them of life, liberty, or property (e.g., the right to cross-examine adverse witnesses).