Recent Cases Highlight Growing Conflict Between AI and Data Privacy

April 20, 2020 Lee Johnston

PRACTICES Media Entertainment and Sports, Anti-SLAPP and First Amendment Rights, AI and Deep Learning, Privacy and Cybersecurity, Media and Entertainment Litigation

No one can question the explosive growth in the use of artificial intelligence ("AI"). Seizing on its powerful predictive capabilities, private sector companies and government entities alike now employ machine-learning ("ML") algorithms to assist in diverse applications ranging from detecting and preventing fraudulent online credit card transactions to optimizing traffic flow to blocking dangerous phishing attempts.

Against this backdrop of AI's speed and efficiency, however, lies increasing concerns about the protection of personal data. Machine-learning algorithms are not born with the advanced predictive capabilities seen in products like Alexa or Google Maps. Rather, like their human counterparts, they require food, care and training that can only be provided by being fed vast amounts of real-life and experimental data about the experiences, perceptions and interactions of humans; i.e., personal data, to achieve deep learning.

Recent cases highlight the growing tension between AI's thirst for and use of huge amounts of personal data to train ML algorithms and personal data's status as a protected commodity under recent U.S. privacy laws like Illinois's Biometric Information Privacy Act (BIPA) and the California Consumer Privacy Act (CCPA), as well as under older privacy statutes like the Health Insurance Portability and Accountability Act of 1996. Many of these cases have been brought even where the allegedly illegal use of the personal data is intended to correct the discriminatory "bias" in ML algorithms, or otherwise achieve laudable goals, such as enhancing healthcare providers' ability to predict patients' future clinical events.

Recent Cases Involving Collection and Use of Machine Learning "Training" Data

Janecyk v. International Business Machines, Case No. 1:20-cv-00783 (N.D. Ill.) (filed January 22, 2020). This putative class action, arises out of IBM's use of publicly available images to create the "DiF" (Diversity in Faces) dataset.[1] The plaintiff, Tim Janecyk, is a photographer who uploaded photos of himself and others at political rallies to the photo sharing site Flickr, which in turn used these and other images to create a database of 99 million images for use as a reference library to train AI models. According to Janecyk, IBM coded a subset of the photos to describe the appearance of the people in the photos, and then offered its collection to researchers as a tool to help reduce bias in facial recognition models.

Notwithstanding its good intentions, IBM now faces potential liability under BIPA of $1,000 to $5,000 per violation for each Illinois resident "who had their biometric identifiers, including scans of face geometry, collected, captured, received, or otherwise obtained by IBM from photographs in its Diversity in Faces Dataset."

Mutnick v. Clearview AI, et al., Case No. 1:20-cv-00512 (N.D. Ill.) (filed January 22, 2020). This putative class action arises out of Clearview AI's creation of a facial recognition database of millions

of Americans trained from more than 3 billion photos Clearview scraped from online social media and other internet-based platforms such as Venmo.[2] The plaintiff, David Mutnick, alleges that Clearview's AI facial recognition database has been sold to over 600 law enforcement agencies, as well as other private entities, to biometrically identify individuals who had no knowledge of, and did not consent to, Clearview's capture and use of their biometric data. In addition to monetary damages under BIPA, the plaintiff recently filed a motion for preliminary injunction, seeking to stop any further dissemination or use of the biometric data and affirmatively requiring Clearview to implement more robust security measures to protect database from further data breaches.[3]

Burke v. Clearview AI, Inc., Case No.: 3:20-cv-00370-BAS-MSB (S.D. Cal.) (filed February 27, 2020). The *Burke* putative class action alleges the same facts and claims complained of in *Mutnick*, but also seeks relief under CCPA based on Clearview's alleged failure to inform consumers "at or before the point of collection" about the biometric information it was collecting and the purposes for which this data was going to be used.[4] Seeking to side-step the absence of a private action for this claim under CCPA, the Burke complaint frames the CCPA violations as violations of California's Unfair Competition Law (UCL), which prohibits business practices that violate other laws.[5]

Dinerstein v. Google, Case No. 1:19-cv-04311 (N.D. Ill) (filed June 29, 2019). The *Dinerstein* putative class action alleges that through a series of corporate transactions allowing it to acquire and absorb an AI data-mining company called DeepMind, and its partnerships with healthcare systems, including the University of Chicago, Google illegally obtained access to hundreds of thousands of patients' medical files in violation of HIPPA. According to the *Dinerstein* plaintiffs, Google utilized this ill-gotten personal healthcare data to "train" machine-learning diagnostic and search algorithms, which in turn it seeks to patent and commercialize in a fee-for-service, subscription or standalone service. The *Dinerstein* complaint asserts a panoply of claims against the University of Chicago, including violations of the Illinois Consumer Fraud and Deceptive Business Practices Act, 815 ILCS 505 ("ICFA") and breach of express and implied contract, and against Google for tortious interference with contract and unjust enrichment.[6]

These recent cases underscore the increased risk associated with the collection and use of data to train machine-learning algorithms and AI models. To mitigate these risks, companies should first inventory the source of all data used to train and develop AI, and then document the intellectual property rights and privacy consents associated with each data set. By doing this audit-based work proactively, companies can better evaluate the risks associated with each data set and develop strategies to mitigate those risks.

---

1. Shortly after the Janecyk state court case was filed, a second complaint against IBM based on the same alleged conduct was filed in Illinois federal court, styled Vance v. IBM, Case No. 1: 20-cv-577 (N.D. Ill.).

2. On February 5, 2020, a second putative class action complaint was filed against Clearview in Illinois federal court alleging similar claims for relief under BIPA. See Hall v. Clearview AI, et al., Case: 1:20-cv-00846 (N.D. Ill).

3. In late February 2020, Clearview disclosed that its client list had been hacked. See Plaintiff's Mem. of Law in Support of Motion for Preliminary Injunction, Dkt. No. 32 at p. 10 (citing Betsy Swan, Facial-Recognition Company that Works with Law Enforcement Says Entire Client List Was Stolen, The Daily Beast (Feb. 26, 2020) ("Clearview Client List Stolen").

4. The Burke plaintiffs recently consented to the transfer of venue in the case to the Southern District of New York, where two other Clearview AI cases are pending. See Joint Motion to Transfer Venue, Dkt. No. 10 (filed April 14, 2020). The two New York federal cases are styled Calderon et al v. Clearview AI, Inc. et al, Case No. 1:20-cv-01296-CM (S.D.N.Y.) and Broccolino v. Clearview AI, Inc., Case No. 1:20-cv-02222-CM (S.D.N.Y.)

5. Cal. Bus. & Prof. Code §§ 17200. The language of the CCPA attempted to avoid the "backdoor" assertions of CCPA violations through the UCL. See Cal. Civ. Code section 1798.150(c) ("Nothing in this title shall be interpreted as the basis for a private right of action under any other law."); cf. Cal-Tech Communications, Inc. v. Los Angeles Cellular Telephone Co., 20 Cal. 4th 163, 182 (1999) (statutes containing "absolute bar" to relief may not be recast as UCL violations). However, this provision under the CCPA is untested, and the California attorney general has advocated for a CCPA amendment permitting a more expansive private right of action under the CCPA.

6. The Dinerstein plaintiffs' contract and tort claims are based on the University's failure to abide by the HIPPA restrictions set forth in the hospital admission and treatment forms signed by the patients and the HIPPA-based privacy policy disclosures provided by the University.