Volume 2, No. 1 | January–February 2019

RAIL

The Journal of Robotics, Artificial Intelligence & Law

fastcase FULL COURT PRESS

RAIL
The Journal of Robotics, Artificial Intelligence & Law
Volume 2, No. 1 | January–February 2019

THE JOURNAL OF ROBOTICS, ARTIFICIAL INTELLIGENCE & LAW (ISSN 2575-5633 (print) /ISSN 2575-5617 (online) at $495.00 annually is published six times per year by Full Court Press, a Fastcase, Inc., imprint. For customer support, please contact Fastcase, Inc., 711 D St. NW, Suite 200, Washington, D.C. 20004, 202.999.4777 (phone), 202.521.3462 (fax), or email customer service at support@fastcase.com.

Publishing Staff
Publisher: Morgan Morrissette Wright
Journal Designer: Sharon D. Ray
Cover Art Design: Juan Bustamante

Cite this publication as:

The Journal of Robotics, Artificial Intelligence & Law (Fastcase)

A Full Court Press, Fastcase, Inc., Publication

Editorial Office

711 D St. NW, Suite 200, Washington, D.C. 20004
https://www.fastcase.com/

POSTMASTER: Send address changes to THE JOURNAL OF ROBOTICS, ARTIFICIAL INTELLIGENCE & LAW, 711 D St. NW, Suite 200, Washington, D.C. 20004.

Articles and Submissions

Direct editorial inquires and send material for publication to:

Steven A. Meyerowitz, Editor-in-Chief, Meyerowitz Communications Inc., 26910 Grand Central Parkway, #18R, Floral Park, NY 11005, smeyerowitz@meyerowitzcommunications.com, 646.539.8300.

Material for publication is welcomed—articles, decisions, or other items of interest to attorneys and law firms, in-house counsel, corporate compliance officers, government agencies and their counsel, senior business executives, scientists, engineers, and anyone interested in the law governing artificial intelligence and robotics. This publication is designed to be accurate and authoritative, but neither the publisher nor the authors are rendering legal, accounting, or other professional services in this publication. If legal or other expert advice is desired, retain the services of an appropriate professional. The articles and columns reflect only the present considerations and views of the authors and do not necessarily reflect those of the firms or organizations with which they are affiliated, any of the former or present clients of the authors or their firms or organizations, or the editors or publisher.

QUESTIONS ABOUT THIS PUBLICATION?

For questions about the Editorial Content appearing in these volumes or reprint permission, please call:

Morgan Morrissette Wright, Publisher, Full Court Press at mwright@fastcase.com or at 202.999.4878

For questions or Sales and Customer Service:

Customer Service
Available 8am–8pm Eastern Time
866.773.2782 (phone)
support@fastcase.com (email)

Sales
202.999.4777 (phone)
sales@fastcase.com (email)
ISSN 2575-5633 (print)
ISSN 2575-5617 (online)

Will "Leaky" Machine Learning Usher in a New Wave of Lawsuits?

Brian Wm. Higgins*

This article examines the causes of action that might be asserted against a developer who publishes, either directly or via a machine learning as a service cloud platform. A leaky machine learning model is also explored along with possible defenses, using the lessons of cybersecurity litigation to frame the discussion.

A computer science professor at Cornell University has a new twist on Marc Andreessen's pronouncement that software is "eating the world."[1] According to Vitaly Shmatikov,[2] it is "machine learning [that] is eating the world" today. His personification is clear: machine learning and other applications of artificial intelligence ("AI") are disrupting society at a rate that shows little sign of leveling off. And with increasing numbers of companies and individual developers producing customer-facing AI systems, it seems all but inevitable that some of those systems will create unintended and unforeseen consequences, including harm to individuals and society at large.

Researchers like Shmatikov and his colleagues are starting to reveal those consequences, including one—"leaky" machine learning models—that could have serious legal implications. Below, the causes of action that might be asserted against a developer who publishes, either directly or via a machine learning as a service ("MLaaS") cloud platform, a leaky machine learning model are explored along with possible defenses, using the lessons of cybersecurity litigation to frame the discussion.

Background

Over the last nearly two decades, both the plaintiffs and defendants bars have contributed to a body of case law now commonly referred to as cybersecurity law. The development of this practice

Robotics, Artificial Intelligence & Law / January–February 2019, Vol. 2, No. 1, pp. 9–16.

ISSN 2575-5633 (print) / ISSN 2575-5617 (online)

area was inevitable, given the estimated 8,000 data breaches involving 11 billion data records made public since 2005.[3]

After some well-publicized breaches, lawsuits against companies reporting data thefts began appearing more frequently on court dockets across the country. Law firms responded by marketing "cybersecurity" practice groups whose attorneys advised clients about managing risks associated with data security and the aftermath of data exfiltrations by cybercriminals.

Today, with an estimated 70 percent of all data being generated by individuals (often related to those individuals' activities), and with organizations globally expected to lose over 146 billion more data records between 2018 and 2023 if current cybersecurity tools are not improved,[4] the number of cybersecurity lawsuits is not expected to level off anytime soon.

Ransomware Attacks

While data exfiltration lawsuits have remained the bulk of cybersecurity litigation, few were surprised when the plaintiffs' bar began targeting other cyber issues, most recently ransomware attacks, especially those affecting healthcare facilities.

In ransomware, an organization's computer systems are frozen by malicious software attacks until a ransom is paid. Those alleging injury in such cases say that, while frozen, a defendant was unable to effectively deliver critical services, which caused harm and injury. Where the business is in the healthcare field, plaintiffs claim the organization's delivery of patient-related services was adversely affected.

What data exfiltration and ransomware litigation have in common is the access to and security of confidential and private customer data. When commercial and government organizations fail to take adequate steps to control access to that data or maintain the privacy of their customers' data, litigation ensues.

Leaky machine learning models are likewise built on data, often using tens of thousands of personal data records, and businesses that do not take steps to maintain the privacy of that data may also become embroiled in litigation.

Machine Learning and Membership Inference

In their research, sponsored in part by the National Science Foundation and Google, Shmatikov and his colleagues in early 2017 "uncovered multiple privacy and integrity problems in today's [machine learning] pipelines" that could be exploited by adversaries to infer if a particular person's data record was used to train machine learning models.[5]

They describe a healthcare machine learning model that could reveal to an adversary whether or not a certain patient's data record was part of the model's training data.

In another example, a different model trained on location and other data, used to categorize mobile users based on their movement patterns, could reveal by way of query whether a particular user's location data was used.

These scenarios certainly raise alarms from a privacy perspective, and one can imagine other possible instances of machine learning models revealing the kind of personal information to an attacker that might cause harm to individuals (an attack does not necessarily involve a nefarious purpose, either; an attack could come from law enforcement's investigations of individuals).

While actual user data may not be revealed in these attacks, the mere inference that a person's data record was included in a data set used to train a model, what Shmatikov and previous researchers refer to as "membership inference," could cause that person (and possibly the thousands of others whose data records were used) embarrassment and other consequences.

The membership inference problem is not limited to the discriminative classification machine learning models investigated by Shokri et al. More recently, researchers at the University College London reported membership inference in generative models, specifically generative adversarial networks ("GANs").[6] Generative models are used to generate new data from the same underlying distribution associated with a particular training data set.

The new "synthetic" data can be used to enhance an existing data set, often images or video data. Hayes and his colleagues describe synthetic health-related images generated by a generative model that could leak information about an individual's health if the individual's record was used to train the generative model.

In another example, if images from a database of criminals are used to train a face-generation algorithm, membership inference could leak an individual's criminal past.

Possible Membership Inference Causes of Action

Assuming for the sake of argument that a membership inference disclosure of the kind described above becomes legally actionable, it is instructive to consider what businesses facing membership inference lawsuits might expect in terms of statutory and common law causes of action so they can take steps to mitigate problems and avoid contributing more cyber lawsuits to already busy court dockets (and of course avoid leaking confidential and private information).

These causes of actions could include:

- invasion of privacy;
- consumer protection laws;
- unfair trade practices;
- negligence;
- negligent misrepresentation;
- innocent misrepresentation;
- negligent omission;
- breach of warranty; and
- emotional distress, among others.[7]

Negligence

Negligence might be alleged, as it often is in cybersecurity cases, if plaintiff (or class action members) can establish evidence of the following four elements:

- the existence of a legal duty;
- breach of that duty;
- causation; and
- cognizable injury.

Liability might arise where defendant failed to properly safeguard and protect private personal information from unauthorized

access, use, and disclosure, where such use and disclosure caused actual money or property loss or the loss of a legally protected interest in the confidentiality and privacy of plaintiff's/members' personal information.

Misrepresentation

Misrepresentation might be alleged if plaintiff/members can establish evidence of a misrepresentation upon which they relied and a pecuniary loss resulting from the reliance of the actionable misrepresentation. Liability under such a claim could arise if, for example, plaintiff's data record has monetary value and a company makes representations about its use of security and data security measures in user agreements, terms of service, and/or privacy policies that turn out to be in error (for example, the company's measures lack robustness and do not prevent an attack on a model that is found to be leaky).

In some cases, actual reliance on statements or omissions may need to be alleged.

State Consumer Protection Laws

State consumer protection laws might also be alleged if plaintiff/members can establish (depending on which state law applies) deceptive misrepresentations or omissions regarding the standard, quality, or grade of a particular good or service that causes harm, such as those that mislead plaintiff/members into believing that their personal private information would be safe upon transmission to defendant when defendant knew of vulnerabilities in its data security systems.

Liability could arise where defendant was deceptive in omitting notice that its machine learning model could reveal to an attacker the fact that plaintiff's/members' data record was used to train the model. In certain situations, plaintiff/members might have to allege with particularity the specific time, place, and content of the misrepresentation or omission if the allegations are based in fraud.

Defenses

For their part, defendants in membership inference cases might challenge plaintiff's/members' lawsuit on a number of fronts. As an initial tactic, defendants might challenge plaintiff's/members' standing on the basis of failing to establish an actual injury caused by the disclosure (inference) of data record used to train a machine learning model.[8]

Defendants might also challenge plaintiff's/members' assertions that an injury is imminent or certainly impending. In data breach cases, defendants might rely on state court decisions that denied standing where injury from a mere potential risk of future identity theft resulting from the loss of personal information was not recognized, which might also apply in a membership inference case.

Defendants might also question whether permission and/or consent was given by a plaintiffs/members for the collection, storage, and use of personal data records. This query would likely involve plaintiff's/members' awareness and acceptance of membership risks when they allowed their data to be used to train a machine learning model.

Defendants would likely examine whether the permission/consent given extended to and was commensurate in scope with the uses of the data records by defendant or others.

Defendants might also consider applicable agreements related to a user's data records that limited plaintiff's/members' choice of forum and which state laws apply, which could affect pleading and proof burdens. Defendants might rely on language in terms of service and other agreements that provide notice of the possibility of external attacks and the risks of leaks and membership inference.

Many other challenges to a plaintiff's/members' allegations could also be explored.

Preventive Measures

Apart from challenging causes of action on their merits, companies should also consider taking other measures like those used by companies in traditional data exfiltration cases. These might include proactively testing their systems (in the case of machine

learning models, testing for leakage) and implementing procedures to provide notice of a leaky model.

As Shmatikov and his colleagues suggest, machine learning model developers and MLaaS providers should take into account the risk that their models will leak information about their training data, warn customers about this risk, and "provide more visibility into the model and the methods that can be used to reduce this leakage."

Machine learning companies should account for these foreseeable risks and associated consequences and assess whether the risks are acceptable compared to the benefits received from their models.

Conclusion

If data exfiltration, ransomware, and related cybersecurity litigation are any indication, the plaintiff's bar may one day turn its attention to the leaky machine learning problem. If machine learning model developers and MLaaS providers want to avoid being the target of this attention and the possibility of litigation, they should not delay taking reasonable steps to mitigate the leaky machine learning model problem.

Notes

* Brian Wm. Higgins is an Intellectual Property & Technology partner at Blank Rome LLP, working with clients to strategically protect, enforce, and defend their intellectual property rights. He may be reached at higgins@blankrome.com.

1. M. Andreessen, *Why Software is Eating the World,* Wall Street Journal (Aug. 20, 2011).

2. https://www.cs.cornell.edu/~shmat/.

3. Privacy Rights Clearinghouse—Data Breaches (available from https://www.privacyrights.org/data-breaches) (accessed Aug. 27, 2018).

4. Juniper Research, https://www.juniperresearch.com/home.

5. *See* R. Shokri et al., *Membership Inference Attacks Against Machine Learning Models,* Proceedings of the 38th IEEE Symposium on Security and Privacy (2017).

6. J. Hayes et al., *LOGAN: Membership Inference Attacks Against Generative Models,* Proceedings on Privacy Enhancing Technologies, Vol. 1 (2019) (available on arXiv; Aug. 21, 2018).

7. *See, e.g., Sony Gaming Networks & Cust. Data Sec. Breach Lit.,* 996 F.Supp. 2d 942 (S.D. Cal 2014) (evaluating data exfiltration causes of action).

8. *See In re Science App. Intern. Corp. Backup Tape Data,* 45 F. Supp. 3d 14 (D.D.C. 2014) (considering "when, exactly, the loss or theft of something as abstract as data becomes a concrete injury").