

OpenAI Class Action Likely to Increase Scrutiny of Web scraping and Data Collection Practices

by Philip N. Yannella and Timothy W. Dickens

July 5, 2023

Summary

A class action lawsuit filed against OpenAI and its primary investor, Microsoft, seeks damages and injunctive relief for the alleged theft and commercial misappropriation of consumer personal data processed by and used to train large language model AIs, including ChatGPT.

The Upshot

- The Complaint relies in part on novel legal theories regarding consumers' rights and economic interests in their publicly available personal information.
- The case is likely to increase scrutiny on AI data collection and disclosures surrounding third-party implementation of AI tools, in particular web scraping activities.

The Bottom Line

While unlikely to halt AI's rapid advance, this case is almost certain to increase internal and external scrutiny of new AI tools and the manner in which third parties integrate AI technologies into service offerings.

On June 28, a group of plaintiffs filed a class action lawsuit against OpenAI—creator and publisher of the generative AI tool ChatGPT—as well as OpenAI's primary investor, Microsoft. The 151-page complaint is the first significant U.S. class action to assert that generative AI tools violate consumer privacy rights.

The Complaint, filed in the Northern District of California, challenges the core of the generative AI models. It alleges that OpenAI stole and “commercially misappropriated” the personal data used to train ChatGPT and related generative AI tools by scraping, collecting, and processing the data of millions of individuals without their consent or authorization. The Complaint further alleges that once OpenAI trained its tools on stolen data, it rushed them to market “without implementing proper safeguards or controls to ensure that they would not produce or support harmful or malicious content and conduct that could further violate the law.” Plaintiffs allege violations of the Electronic Communication's Privacy Act (ECPA) and the Computer Fraud and Abuse Act (CFAA); invasion of privacy laws in California and Illinois; unfair competition laws in California, Illinois, and New York; as well as violations of the Illinois's Biometric Information Privacy Act (BIPA). In addition to these statutory claims, the Complaint contains claims of negligence, invasion of privacy, intrusion upon seclusion, larceny/receipt of stolen property, conversion, unjust enrichment, and failure to warn.

The data at issue in the Complaint can be broken into two primary data sets: the training data scraped from the internet and the data collected and processed by AI enabled tools, such as ChatGPT's implementation for Snapchat and Stripe. However, given the volume of data and the number of data sources at issue, establishing and responding to claims regarding how specific groups of users' data was collected or processed is likely to be extremely difficult.

A primary hurdle for plaintiffs' theory of the case is that the much of the training data at issue is publicly available on the internet. Courts have been generally unwilling to block the scraping of public internet data when challenged by competing online businesses (see *HiQ Labs v. LinkedIn Corp.*, 31 F.4th 1180 (9th Cir. 2022)). Additionally, the processing of personal information

by generative AI models is arguably different in nature from the scraping of information from public websites for profiling purposes. Plaintiffs may be able to find some traction for an argument by looking to Justice Gorsuch's dissenting opinion in the Fourth Amendment case *Carpenter v. U.S.*, in which he laid a foundation for an individual claiming a property interest in personal information. The case is 138 S.Ct. 2206, 2268 (2018).

Similarly, plaintiffs are likely to face difficulty in establishing disclosure-based claims regarding the collection of information through AI-powered tools. Plaintiffs will have to identify what information is collected through which tools, what disclosures were provided and whether the given disclosures were adequate, and to what extent OpenAI can be responsible for disclosures and processing conducted by third parties that integrate OpenAI tools.

On the other hand, the Complaint places a significant emphasis on the potential risks to children associated with generative AI. While the Complaint does not allege violations of the Children's Online Privacy Protection Act (COPPA), the Complaint does assert that OpenAI trained its tools using children's data and "designed ChatGPT to be inappropriate for children." It also alleges that OpenAI deprived children of the economic value of their personal data. Increased legal scrutiny of children's online privacy has led to a surge in state children's privacy laws, such as the California Age Appropriate Design Code Act, as well as FTC regulatory activity (see Epic Games consent decree). Claims that generative AI models violate children's privacy rights or cause specific harm to children may trigger additional regulatory scrutiny.

The lawsuit against OpenAI is the first significant legal test of the webscraping and data collection practices that underlie many generative AI models. For this reason alone, the lawsuit will be closely watched by AI developers, regulators, and attorneys.

While unlikely on its own to halt the rapid advancement of generative AI, the outcome of this lawsuit may influence the development of such technologies in the future. We will continue monitoring this case as it develops.

Copyright © 2023 by Ballard Spahr LLP.
www.ballardspahr.com
(No claim to original U.S. government material.)

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, including electronic, mechanical, photocopying, recording, or otherwise, without prior written permission of the author and publisher.

This alert is a periodic publication of Ballard Spahr LLP and is intended to notify recipients of new developments in the law. It should not be construed as legal advice or legal opinion on any specific facts or circumstances. The contents are intended for general informational purposes only, and you are urged to consult your own attorney concerning your situation and specific legal questions you have.

Contacts

Philip N. Yannella

PARTNER

yannellap@ballardspahr.com

215.864.8180

Timothy W. Dickens

ASSOCIATE

dickenst@ballardspahr.com

215.864.8126