MLex Ltd. | www.mlex.com
Phone: +44 800 999 3237 | +1 800 356 6547 | +852 2965 1424 | customerservices@mlex.com

Proprietary AI data labels are often unprotected trade secrets

By **Nick Robertson**

January 2, 2026, 21:24 GMT | Comment

As companies increasingly customize AI models with proprietary software, agents and data-labeling frameworks, Saul Ewing partner Matthew Kohel told MLex that IP practitioners must remain keenly aware of the commercially valuable trade secrets in data filtering and data set creation.

While the trade secret implications of AI training data sets and models have received significant attention from IP practitioners, custom tools used to filter and manage data within proprietary models represent an overlooked trade secret protection opportunity, Saul Ewing partner Matthew Kohel told MLex.

As AI technology has become commonplace, more companies are layering their own software and agents onto “out-of-the-box” AI models to best tailor them to their specific needs. The proprietary data labels and other data filtering information used to enable those in-house solutions may be valuable trade secrets.

Kohel said that custom AI models and agents, specific to a company’s data and needs, are driving the next wave of AI innovation and commercial development. The bespoke tools enable companies to filter and pare down massive public data sets into actionable insights.

“There's just so much information. You can sign up for API feeds and get data. You can buy data right from data brokers. There's tons of publicly-available data sets that are out there,” Kohel said. “The hard part is curating it, cleaning it, making it really useful, but just the access to information and the accessibility of LLMs advances things so much, because that's where the insights come from.”

Proprietary data labels can transform publicly-available data into custom solutions for a company’s in-house AI model. For example, satellite imagery combined with a company’s private data can be labeled and interpreted to predict supply chain vulnerabilities. The same principle can be applied to numerous industries.

Kohel said that many companies are unaware of the trade secret risks posed by AI, especially from custom models and agents.

“The real challenge is understanding the scope of what can be a trade secret,” Kohel said. “The shift [towards using AI] is going to really push people to innovate, but they need to understand the security concerns of publicly available platforms.”

Security vulnerabilities arise from gaps in employee training and human error, Kohel said. Specific to AI agents, a mistake as simple as granting a custom agent permissions to access more files than necessary, often out of convenience, could unintentionally expose a company’s trade secrets.

“A lot of trade secret loss or damage isn't necessarily bad actors, a lot of it is due to sloppy employees, or employees who may just not know,” Kohel said.

Even as AI models and agents become more advanced and autonomous — including the use of AI to help track and protect confidential data (see here) — human oversight remains the cornerstone of trade secret protection, he said.

“If you don't have the perspective of lots of people internally that have different expertise — information security people, HR people, data privacy — you're going to have a blind spot,” Kohel said. “And it's going to either result in you giving away something by inadvertently disclosing it, or by not managing a threat, whether internally or externally.”

“With the technology becoming more complicated, it really is much more of a partnership with your human employees, understanding the blind spots, following protocols that are put in place after the technology is explored and you do all the use case analysis,” he continued. “Having those open lines of communication between information security people and IP professionals just makes sure that even your employees aren't doing things that are giving away the game inadvertently.”

Kohel recommended a two-prong approach to protecting trade secrets while using AI agents and custom data labeling: contractual and operational defense.

Robust IP clauses in agreements with contractors and AI software providers must go beyond simply taking ownership of data, Kohel said. Companies should explicitly claim ownership of derivative data, including proprietary labels, annotations and the framework for data filtering and labeling to ensure ownership and protection of potential trade secrets.

In day-to-day operation, practitioners should be focused on creating a system that can meet the “reasonable measures” test in trade secret litigation. That includes employee confidentiality — implementing non-disclosure agreements and thorough training — limiting employees’ access to data they don’t need, and setting up monitoring systems to track who accesses sensitive information, Kohel said.

Ultimately, the key to protecting trade secrets amid the rise of AI is still human.

“I think it really gets started with buy-in from your people,” he said.

Please email editors@mlex.com to contact the editorial staff regarding this story, or to submit the names of lawyers and advisers.

Tags

Sections: Artificial Intelligence, Intellectual Property
Industries: Computing & Information Technology
Geographies: North America, United States
Topics: Trade Secrets