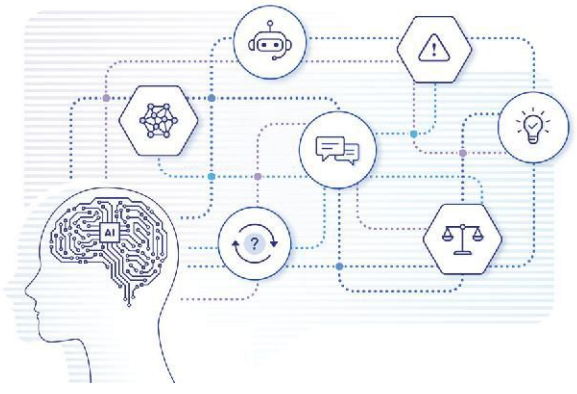
The case for vertical AI models in F&I

TOM OSCHERWITZ
Informed.IQ



AI researchers from Google, Berkeley, and other universities recently engineered what they described as a "silly attack" on Chat-GPT. They asked the model to repeat the word "poem" forever, which caused the model to emit large quantities of its training data, including names, email, physical addresses, phone numbers and other personal information (PII).

Though superficially amusing, the attack has troubling implications. It vividly exposes how generic multimodal LLMs like ChatGPT, Bing, or Claude can expose training information. Custom-GPT models also can leak data as demonstrated by Researchers at Northwestern University, who used prompt injection attacks (malicious use of prompts) to retrieve user-uploaded files. Given the capacity of generic and custom GPT models to leak data, prudent risk managers at lenders should be careful about using models built on data whose provenance is unknown.

Independent of this most recent research, Open AI and other generic model developers face litigation and regulatory challenges from all quarters. Copyright holders have filed multiple lawsuits against these models. Domestically, the Federal Trade Commission has launched a probe of OpenAI for violation of consumer protection laws. From a Supervisory perspective, these models implicate compliance risk and reputational risk, two of the nine risk categories monitored by the Office of the Comptroller of the Currency (OCC). And internationally, Italian, Japanese, and other regulators have also begun investigations.

In light of this legal uncertainty and regulatory scrutiny, our view is that currently, generic multimodal models are not viable for all use cases. For industries with sensitive data like lending, we use AI models that produce verifiable outputs with the following characteristics:

- The training data is properly licensed and permissioned: Models should only train on data explicitly permissioned by the data rights holders.
- Models are transparent and subject to model governance: Constantly monitor models, assuring that we meet model governance and regulatory requirements.
- Models are trained on high-quality data: Training data comes from highly scrutinized and vetted data sources, like customer applications or income documents.
- Training data is fit for purpose: Have large quantities of the precise information historically used to solve the lending problems faced by lenders, whether it's auto lending or other lending verticals. Rather than using someone else's generic LLMs, create and use "vertical" models.
- The models operate behind security walls: Ensure that models are not publicly available. There are security virtues with non-public data sets, because they minimize data exposure and opportunities for hostile attacks.

This approach is called Vertical AI. This approach meets lenders' business challenges from the initial ingestion of a loan application, all the way through loan approval. Vertical AI leverages many of the benefits of large multi-modal models without the troubling baggage. These models do highly accurate linkages and analyses involving images, words, and text from lender documents that customers depend upon in real-time. Like generic models, Vertical AI models show continuous improvement as they grow. But, unlike generic publicly available models, this growth occurs in a safe, secure and protected data environment.

It's worth stressing that this Vertical AI minimizes intellectual property and privacy risks by using explicitly permissioned data. Vertical AI data comes from the data rights holders and is governed by mutual expectations and obligations defined in contracts, including commitments to complying with regulatory requirements. In financial services, our Vertical AI models comply with Gramm-Leach Bliley, the Equal Credit Opportunity Act, the Dodd-Frank Act and other financial laws.

Lenders should be extremely bullish on multimodal LLMs. But, there is a right (and careful) way to adopt them. The industry should only use these models when there is sufficient confidence in their safety and reliability, and have visibility into how they were trained. If the model lacks transparency and appropriate permissioning, it bears extra scrutiny. And presumptively is not a good fit for lenders or others who must comply with rigorous regulations and meet consumer privacy expectations.

AI has extraordinary promise for lending and other industries. Let's build it out the right way.

Tom Oscherwitz is vice president of legal and regulatory advisor at Informed.IQ, an AI software company with a specialty in auto lending. He has over 25 years of experience as a senior government regulator (CFPB, U.S. Senate) and as a fintech legal executive working at the intersection of consumer data, analytics, and regulatory policy.