Proskauer

Key Issues to Consider When Investing In or Contracting With AI Businesses

June 2023

While many types of AI have existed in some form for several years, the new generation of sophisticated AI solutions, especially generative AI ("GenAI"), has driven a rise in AI use cases, including tools to create and debug software code, mint collections of blockchain-based NFTs and produce other content, automate responses to customer queries via chatbots (and create help center scripts), summarize customer reviews and datasets, screen resumes and analyze video interviews, and produce real-time translations, among others. This session, "Key Issues to Consider When Investing In or Contracting With AI Businesses," addresses legal and practical concerns for businesses that are seeking to procure AI systems or invest in AI business, or, on the other hand, AI businesses seeking to provide their software to the market and prepare for further investment. The key question for organizations can be boiled down to: Is there anything to keep in mind when contracting for AI systems or investing in AI ventures? The answer is "Yes," with the presenters outlining important, practical issues during the webinar, including: (1) the key risk issues when contracting for GenAI and how such issues can be managed; and (2) investment due diligence and transactional agreement issues when considering M&A and AI-related business targets.

The terms governing use of an AI system and the legal issues raised can vary depending on the service's particular attributes, which might range from:

- Free vs. paid
- Shared infrastructure vs. dedicated infrastructure
- Provider's own user interface vs. API access
- Sophistication of architecture
- Ability of user to train using their own datasets or ability to customize
- Cloud hosted (SaaS) vs. locally deployed (licensed)
- Open source vs. proprietary

- Size and diversity of training dataset; quality of outputs and languages recognized

Contracting with AI Businesses

The presenters discussed how companies should manage the risks that come with procuring and using a GenAI system and what risks AI developers should expect to be raised by their customers.

The presenters pointed to five general risks of using GenAI under agreements with AI providers and mechanics to mitigate and manage such risks (noting that such management will include contractual protections, but that such provisions "only go so far" and negotiations must be targeted in approach). Thus, the presenters stated that practical, operational mechanics are also needed (a/k/a "guardrails") and that risks and protections will change depending on the GenAI tool and how it is being used.

- **Risk 1: Accuracy of output and bias**. As previously discussed in prior webinars in this series, no business should assume that GenAI output is 100% accurate, as systems are not designed to always produce the "right answer" and may fabricate facts and sources (and GenAI training data may be inaccurate, incomplete and not jurisdiction-specific and may produce discriminatory output). As the presenters noted, GenAI product terms will sometimes disclaim all responsibility for the quality of outputs. Depending on one's bargaining position, a user might negotiate for acceptable data standards and controls focusing on accuracy of output products and removal of bias. However, as noted by the presenters, such a position may be a "non-starter" for many GenAI providers as the most popular GenAI tools right now are not bespoke products and likely cannot be technically tailored (outside of certain GenAI tools that allow users to fine-tune using their own datasets). Thus, the presenters stated that if a contractual protection of this sort is not available, users will need to resort to practical guardrails to mitigate inaccuracy and bias, including: designing queries with focus on jurisdictions and supplementing outputs with information from other reliable sources; conducting human review for more risky outputs; and considering insurance policies to cover such risks.
- **Risk 2: Confidentiality.** Generally speaking, inputs and outputs from publicly-available GenAI tools may be used by the provider for training purposes. Thus, there is the potential for breaches of confidentiality, IP leakage, loss of legal privilege and loss of trade secret protection when such information is inputted into a public-facing GenAI interface. Moreover, the presenters also stated that inputting third party licensed data or software into a public GenAI tool might breach contractual obligations. To mitigate this risk, the presenters stated that users should engage AI providers as to terms of confidentiality and non-sharing

restrictions, particularly for protecting prompts and other inputs. Some providers, such as OpenAI, offer such protections to its paid, enterprise users; on the other hand, ChatGPT's public version offers no such standard protections and it is up to the user to take practical steps to mitigate this risk, such as affirmatively opting-out of any data sharing through privacy setting controls; limiting submission of confidential information or rewording or disaggregating important elements of confidential information (e.g., submitting snippets of code without context).

- **Risk 3: Security.** The presenters stated that businesses need to be cautious when integrating GenAI solutions into their systems, as this rapidly developing technology may have unknown vulnerabilities. GenAI products may either be hosted and managed by a service provider or downloaded and deployed on the customer's own infrastructure. From a contractual standpoint, the presenters stated that users can seek contractual protections and security assurances from AI providers, though, again, this may be difficult to negotiate outside of a customized GenAI arrangement and users deploying both enterprise and standardized products should examine the terms related to data security (with the former likely offering additional security assurances than the latter) or look to existing products that offer GenAI tools that already promise certain levels of security. Practically speaking, the presenters stated that GenAI tools should be pre-checked before downloading, as with any software product, and that protections should be put in place to monitor data flow against "poisoning" of training data or related AI-related "inference" attacks that might allow an attacker to infer what training data was used to train the model.
- **Risk 4: Privacy.** The presenters stated that submitting personal data into an AI model could infringe privacy rights of data subjects; the user of personal data in outputs could be similarly infringing of the rights of individuals. Thus, from a contractual standpoint, the presenters stated that a data processing agreement should be entered into by the user and the AI provider that processes personal data on the user's behalf, with such agreement covering various aspects of privacy and security. While the presenters pointed out that a data processing agreement might be fairly standard in nature, users might scrutinize provisions related to privacy liability and how GenAI providers characterize their role under data privacy regulations (i.e., whether they are data processors or data controllers). Other practical considerations when protecting against privacy risk include: using available settings to opt out of any data sharing; avoiding sharing any personal data unless truly required; and reviewing privacy policies.
- **Risk 5 – Intellectual Property.** The presenters pointed to an earlier webinar in this series that covered the IP issues, but wanted to briefly cover some general IP concerns surrounding the use of GenAI, such as the fact that the training of the model could potentially be considered an infringing use of copyrighted works and

certain output could be deemed to be an unauthorized derivative work of third party content (not to mention that registration of GenAI output is on unclear legal footing). As for contractual protections, the presenters suggested that users examine terms concerning the statement of ownership of inputs and outputs and indemnities as to infringement, as different GenAI tools handle these issues differently.

Overall, the presenters stressed that, on the contractual side, it helps to have counsel with experience in this area that negotiate specific risk items rather than academic points. To further mitigate risk, the presenters suggested companies produce GenAI-focused internal documents, policies and training materials that outline Do's and Don'ts and explain risks for specific GenAI tools.

Investment in AI Companies

In the second part of this webinar, the presenters went over some issues to consider when investing in an AI business (which might be described as one that hosts and offers an AI service or uses AI as an integral part of its business functions). Investors might first seek to uncover the core value in an AI service provider as well as the primary revenue drivers for a provider, whether it concerns rights to datasets (including third party licenses), proprietary data collected by the company, proprietary models used to ingest and analyze the data, key personnel and engineers, or rights to use user inputs and outputs for training purposes. AI-related deals raise some similar issues to other technology deals related to software and SaaS, but the presenters stressed that there are a host of due diligence issues unique to AI that are relevant to investment in or acquisition of an AI company and knowledge of these issues can help buyers structure deals and tailor representations and warranties in purchase agreements and sellers manage operations to attract further investment.

The presenters noted that the terms of use or license agreement are foundational documents, and investors should understand what positions the target company is taking on key issues of potential liability, data sharing, allowance for using user prompts for training purposes, or potential over-commitments for accuracy or non-infringement. Investors might also ask additional questions, including:

- What level of negotiability does the provider have with customers?

- What exposure does the company have to risk areas previously discussed in this webinar?
- How is the AI trained and how are models and outputs validated?
- What are the company's practices with compliance with IP and other laws and relevant contracts with third parties?
- Does the provider use only lawfully-procured data according to data procurement standards and controls?
- Does the AI company have any gaps with respect to warranties and indemnities with their data providers?
- Does the AI company have any contracts with third party data providers that limit its flexibility or compel it to give up rights in the underlying technology?
- Are there any issues related to UK IP issues related to text and data mining?
- As to cybersecurity, what protocols has the AI company put in place, and how is the data stored, and where?
- Some additional questions might include operational queries, including: How will the target's AI business be integrated into the buyer's existing business? Are there any implications from open source elements of the target's platform?

The presenters also brought up some US-related due diligence issues with respect to data scraping and data mining, and the variety of potential legal claims that a website operator could make against a data scraper, such as liability under the Computer Fraud and Abuse Act (CFAA), which prohibits access to a protected computer "without authorizations" or that "exceeds authorized access," most pertinently related to website pages that are not publicly accessible. Additionally, the webinar discussed additional U.S. legal issues investors should consider, including copyright infringement liability (subject to a fair use defense vs. the UK's narrower "fair dealing" defense), and breach of contract for violating website terms of service, which may or may not be enforceable in all cases and instances. Overall, the presenters stressed that investors should assess potential infringement risks in training data and outputs: What rights does the AI provider have to user prompts and outputs? If AI company users its own AI to develop outputs, is ownership of output clear? What IP protection is available for the company's algorithms and models?

As to due diligence related to compliance with laws and regulations, the presenters noted that investors should consider a host of relevant laws, including U.S. and EU privacy laws, biometric privacy laws, industry-specific laws (e.g., healthcare), and employment laws (particularly with respect to employment discrimination, including local laws such as a recently-enacted New York City AI-related employment regulation). Rules on foreign investment and national security export regulations may also be relevant, depending on the nature of the AI system.

Looking ahead, it is important for an investor to also look long-term, and closely examine how the AI provider is positioning itself with respect to future regulations and other emerging AI principles that have captured regulators' attentions. These might include: potential for outputs to influence undesirable human behavior, discrimination through bias, exploitation of vulnerable groups, and real-time biometric profiling, as well as use of AI processing of personal information to use in automated decision technology. As for the EU, for example, the AI Regulation is forthcoming and will place the most regulation upon high-risk AI tools. It is also important to note that the AI Regulation will have extraterritorial effect, thus impacting UK and U.S. entities with global operations. The AI Regulation is expected to place additional regulatory burdens on AI companies that may increase exposure and cost for investors in such companies (particularly those that handle so-called high-risk tools). Investors should examine how AI companies are incorporating policies and procedures to comply with existing and upcoming AI regulations – otherwise, if an AI company must later re-engineer their technology, this will have an adverse impact on an investment.

The presenters posited additional due diligence questions that might be asked when a buyer is acquiring a company that uses GenAI, a fact that may not be obvious without performing some investigation. Questions include: How does the target use GenAI? What is the nature of information the company includes in this their prompts? What are the governing terms of service with the GenAI provider? How is the GenAI hosted? Is the company using its own data to train the GenAI system? Are there any ambiguities in ownership in AI outputs?

In all, this is an exciting time for users, developers, investors, but also presents legal risks that need to be carefully managed. While the diligence approach is similar in some respects to other tech deals, the presenters closed with a reminder to viewers that AI-related issues need to be considered through a special lens and ideally by a practitioner or team with experience in this emerging area.

Related Professionals

- **Wai L. Choy**
  Partner
- **Oliver R. Howley**
  Partner