Davis Polk

AI traps and strategies: Seven practical pointers for businesses

April 12, 2024 | Client Update | 9-minute read

As the use of generative artificial intelligence becomes ubiquitous, it is critical for companies to learn to recognize and avoid common traps associated with AI. The hazards range from uncertain intellectual-property protections and privacy considerations to the risk of missing out on the significant benefits of AI. Here, we highlight seven key traps and associated strategies for businesses.

Trap No. 1: A company uses AI tools to develop a valuable product, idea, or work

Risk

The availability of intellectual property protections for AI-generated material is unclear.

The U.S. Copyright Office has denied copyright registration for AI-generated images, but it has also stated that copyright protection may be available for works that combine such material with "human authorship." Similarly, U.S. courts have concluded that an inventor of patent must be a human, and the U.S. Patent Office recently issued guidance stating that a human inventor must "significantly" contribute to an invention (as discussed in our client update). Other jurisdictions, including China and Europe, continue to reach varying conclusions about IP protection when AI is involved.

Recommendations

Take steps to demonstrate human involvement and exercise caution when using AI where protection of output is important.

- **Incorporate human modification:** Where feasible, humans should review and modify AI outputs by, for example, editing and adding text, or selecting and arranging components into a new whole.
- **Compile evidence of human efforts:** Because the prompts fed into AI represent a human's original intellectual investment in the final product, users should save records of all the prompts they provide to document their contribution. While the U.S. Copyright Office has cautioned that simply prompting AI may not necessarily confer copyright protection on the resulting work, human creativity in the inputs will increase the chances of IP protection.
- **Avoid use of AI where IP ownership over output is crucial:** If it is not feasible to track inputs or have humans review and modify outputs and IP protection is vital, companies should minimize the use of AI where IP ownership of the work is important.

Trap No. 2: Employees input personally identifiable information (PII) or other private information into generative AI tools

Recent studies have found that 15% of workers are using generative AI tools and that personally identifiable information comprises 12% of what they paste into those tools.

Risk

Inputting PII into generative AI tools without the PII owner's consent may violate data-privacy laws.

Most leading generative AI tools do not currently permit the deletion of such information once inputted. Without sufficient oversight of employees' inputs, companies may violate federal and state data-privacy laws, as well as international data?privacy laws such as the EU's General Data Protection Regulation and the new AI Act.

Recommendations

Negotiate protections with generative AI providers and educate and monitor employees.

- **Negotiate protections with providers:** Companies should negotiate the terms of use and any data-processing agreements with generative AI providers to make clear that any PII inputted into the tool continues to belong solely to the company. The terms should also restrict the provider from disclosing such information and prevent the provider from incorporating the information into its model, especially if that model may be used by other customers. Companies should also seek to strike language that would limit the provider's liability for breaches of data-privacy laws.
- **Update employee materials:** Companies should update their employment agreements, trainings, handbooks, and related materials with express guidance on what types of information may be inputted into generative AI tools and how employees can anonymize data to avoid feeding PII into those tools.
- **Automatically filter and de-identify inputs:** Companies should consider content?filtering technology that identifies PII, stores it in a data-privacy "vault," and "de-identifies" it by anonymizing any sensitive information before sending the filtered data to a generative AI tool. Although such technology is not foolproof, proper deployment in a company's generative AI dataflow can help reduce the risk of a data-privacy breach.

Trap No. 3: A company relies on AI outputs for crucial decisions or products

Risk

AI models can "hallucinate" and produce inaccurate output that appears trustworthy.

Relying on inaccurate AI output can present business and legal risks, particularly if the false or inaccurate information shows up in public or client-facing materials. Last year, for example, lawyers submitted a legal brief that contained made-up case citations from an AI tool—and were sanctioned for doing so. Indeed, leading generative AI companies have started to add disclaimers to their product page explaining that their tool can make mistakes and that users should consider checking important information.

Recommendations

Build in human supervision and restrict AI usage where output cannot be confirmed.

- **Iterate different prompts and ask the AI model for sources of information:** Employees should use different formulations of the same prompt and use different AI models to cross-check answers. Where possible, they should also conduct human fact checking to confirm the AI's outputs, both on a per-output basis and on general trends.

- **Add disclaimers to AI-generated content:** Companies should add disclaimers regarding the accuracy of AI-generated content where the output cannot be confirmed.
- **Avoid use of AI where accuracy of output is necessary but cannot be confirmed:** Companies should require human supervision where AI is used to produce crucial outputs, and prohibit its use where risks that accompany unconfirmable outputs are particularly high.

Trap No. 4: A company utilizes generative AI outputs to recommend and perform actions on a large scale

Risk

Generative AI outputs may be algorithmically biased, exposing the company to reputational and business risks as well as liability under anti-discrimination, fair housing, and employment laws.

AI systems are only as accurate as the datasets they are trained on, and when that data includes biased information, they can output biased answers. Such algorithmic bias can affect the quality of a company's products and services as well as its reputation. In addition, several federal agencies and the Department of Justice have affirmed that AI systems are subject to anti-discrimination laws and have committed to enforcement against systems that violate those laws.

Recommendations

Monitor AI inputs and outputs to prevent discriminatory effects.

- **Build diverse and multidisciplinary teams working on AI systems:** Diverse teams working on AI systems will help minimize, identify, and mitigate bias issues in the development phase. It will also promote a culture in which algorithmic bias issues may be identified and addressed before they become acute.
- **Conduct human review of training datasets and outputs:** Users must understand where a model's training dataset came from and account for the types of biases inherent in the source material. They must also carefully review the system's outputs for bias before they are disseminated or relied on. If review of every output is impractical, users should at least review random samples to assess performance.
- **Conduct bias audits:** Because AI tools are constantly learning, a company should conduct periodic audits to ensure the systems are not developing biases over time. Audits will enable the company to identify emerging issues and reprogram the tool or limit its use.

Trap No. 5: A company uses a generative AI tool trained using copyrighted content

Risk

A company could unwittingly create content that infringes copyrighted material, and the tool's creator could also be liable for copyright infringement.

Third-party training datasets may create liability for copyright infringement, not only for companies using such datasets to build generative AI tools, but also for companies using AI tools trained on such datasets. Over 70% of the most commonly used datasets do not include licensing information for the underlying content, despite the fact that many of the content originators *did* specify how their content should be licensed if used by others. Courts across the United States are now working through how copyright in that underlying content applies to both creators and users of AI tools.

Recommendations

When building or using AI tools, consider the source of the training data and minimize infringement risks.

- **Confirm licensing information and alignment when using third-party datasets to train AI tools:** Confirm that the licensing information for the content is clearly available and verify (if possible) that the licensing information in the consolidated dataset accurately reflects the policies specified by the original content creators.
- **Consider using synthetic data:** Consider if synthetic data (i.e., machine generated data based on the properties of real-world data) can adequately train your tool. Use of synthetic data has a much lower risk of liability than using third-party datasets or creating your own datasets.
- **Seek indemnification when using third-party AI tools:** Companies should negotiate protections that require the provider of an AI tool to indemnify for any potential IP claims arising from its tool or training dataset.

Trap No. 6: A company does not accurately disclose its use of generative AI

Risk

Hiding or mischaracterizing generative AI use may elicit regulatory investigations or decrease consumer confidence in a company's product or service.

As generative AI rapidly attracts the attention of companies eager to maximize business efficiencies, many companies may be tempted to understate or exaggerate their use of generative AI. The Securities and Exchange Commission has warned against mischaracterizing generative AI use and recently investigated the compliance of numerous investment advisers. The Federal Trade Commission has similarly warned against companies misrepresenting or exaggerating their use of AI.

Recommendations

Companies should make only accurate and low-risk disclosures in their terms of use, corporate filings, and privacy policies.

- **Avoid so-called "AI-washing":** Companies should consider defining what they mean when referring to generative AI. A definition can helpfully ensure their corporate disclosures do not falsely claim a product or service uses generative AI when it does not. SEC Chair Gensler has cautioned companies to define their understanding of generative AI, accurately represent their use of AI tools, provide the basis for such representations, and identify the material risks associated with AI technology.
- **Obtain customer consent:** Companies should communicate to customers the material risks associated with the company's use of generative AI technology. The use of generative AI is not always readily apparent to consumers, so many consumers will not independently research the associated risks.
- **Disclose policies on employee AI use:** Companies should consider disclosing whether they monitor internal use of generative AI products and services, and provide high-level descriptions of safeguards against unintended use cases. In particular, companies may disclose limitations on the scope of their generative AI use, as well as their policies to enforce those boundaries.
- **Disclose privacy mechanisms:** Finally, companies should disclose whether they adopt privacy mechanisms for protecting sensitive input information (as discussed in Trap. No. 2). In particular, companies should consider disclosing various privacy mechanisms for children and other high-risk populations.

Trap No. 7: A company worried about the potential risks of AI prohibits its use altogether

Risk

Companies that entirely prohibit using AI risk falling behind their competitors and wasting time on mundane operational tasks while their competitors focus on skilled, innovative work.

The company's products and services may fall behind competitors, its marketing and communications may be less timely or effective, and its workforce may become less productive and face retention issues.

Recommendations

Create an AI governance team that can approve appropriate uses while establishing safeguards to mitigate the risk.

- **Organize a strong and diverse AI governance team:** Establish a cross-organizational AI governance team comprising business, engineering, and legal leaders, including the AI evangelists within the company.
- **Ask questions and conduct due diligence:** Investigate which tools may be appropriate and consider how they were developed and trained, how robust their security practices are, and their terms of use. Refresh this analysis on a regular basis so that the team stays up to date in a quickly evolving environment.
- **Institute policies to govern use of AI:** Develop internal policies, procedures, and risk-mitigation rules and strategies, including establishing permitted tools, inputs and uses. Then educate employees to enable them to avoid the traps while taking advantage of what AI has to offer.

If you have any questions regarding the matters covered in this publication, please reach out to any of the lawyers listed below or your usual Davis Polk contact.

David Lisson
+1 650 752 2013
david.lisson@davispolk.com

Matthew J. Bacal
+1 212 450 4790
matthew.bacal@davispolk.com

Elaine M. Andersen
+1 650 752 2083
elaine.andersen@davispolk.com

James Y. Park
+1 650 752 2044
james.park@davispolk.com

Serge A. Voronov
+1 650 752 2055
serge.voronov@davispolk.com