

# AI, Machine Learning & Big Data 2024

Sixth Edition

---

Contributing Editor:

**Charles Kerrigan**

CMS Cameron McKenna Nabarro Olswang LLP

# CONTENTS

<b>Preface</b>	Charles Kerrigan, <i>CMS Cameron McKenna Nabarro Olswang LLP</i>	
<b>Expert analysis chapters</b>	<i>Ethical AI</i> , Erica Stanford & Charles Kerrigan, <i>CMS Cameron McKenna Nabarro Olswang LLP</i>	1
	<i>Practical guidelines for the use of generative AI</i> , David V. Sanker, <i>SankerIP</i>	8
	<i>HAL, the Terminator, and Agent Smith entered a bar; Regulation of Artificial Intelligence in the United States</i> , Richard B. Levin, Kevin Tran & Bobby Wenner, <i>Nelson Mullins Riley &amp; Scarborough LLP</i>	16
<b>Jurisdiction chapters</b>		
<b>Australia</b>	Jordan Cox & Francesca Mendoza, <i>Webb Henderson</i>	29
<b>Austria</b>	Günther Leissler & Thomas Kulnigg, <i>Schoenherr Attorneys at Law</i>	43
<b>China</b>	Peng Cai, <i>Zhong Lun Law Firm</i>	46
<b>France</b>	Boriana Guimberteau, <i>Stephenson Harwood</i>	55
<b>Germany</b>	Moritz Mehner, Martin Böttger & Christoph Krück, <i>SKW Schwarz</i>	69
<b>Greece</b>	Marios D. Sioufas, <i>Sioufas &amp; Associates Law Firm</i>	82
<b>India</b>	Aprajita Rana, Navdeep Baidwan & Shubham Parkhi, <i>AZB &amp; Partners</i>	95
<b>Ireland</b>	Jane O’Grady, <i>LK Shields</i>	106
<b>Italy</b>	Massimo Donna & Chiara Bianchi, <i>Paradigma – Law &amp; Strategy</i>	118
<b>Japan</b>	Akira Matsuda, Ryohei Kudo & Taiki Matsuda, <i>Iwata Godo</i>	131
<b>Lithuania</b>	Asta Macijauskienė, Viktorija Stančikė & Renata Jankauskytė, <i>WIDEN</i>	143
<b>Malta</b>	Ron Galea Cavallazzi, Alexia Valenzia & Veronica Campbell, <i>Camilleri Preziosi</i>	151
<b>Netherlands</b>	Joris Willems, Sarah Zadeh & Danique Knibbeler, <i>NautaDutilh</i>	161
<b>Poland</b>	Natalia Kotłowska-Wochna, Monika Maćkowska-Morytz & Tomasz Szambelan, <i>Kochański &amp; Partners</i>	175
<b>Portugal</b>	Magda Cocco, Sofia Barata & Iakovina Kindylidi, <i>Vieira de Almeida</i>	187
<b>Singapore</b>	Lim Chong Kin, Anastasia Su-Anne Chen & Cheryl Seah, <i>Drew &amp; Napier LLC</i>	197

<b>South Africa</b>	Simone Dickson, <i>Independent Consultant</i>	211
<b>Switzerland</b>	Jürg Schneider & David Vasella, <i>Walder Wyss Ltd.</i>	214
<b>Taiwan</b>	Robin Chang & Eddie Hsiung, <i>Lee and Li, Attorneys-at-Law</i>	225
<b>United Arab Emirates</b>	Sandeep Bhalothia & Aman Garg, <i>Node.Law</i>	236
<b>United Kingdom</b>	Charles Kerrigan, Erica Stanford & Hannah Francis, <i>CMS Cameron McKenna Nabarro Olswang LLP</i>	247
<b>USA</b>	Chuck Hollis & Sean Christy, <i>Norton Rose Fulbright</i>	258

# Practical guidelines for the use of generative AI

David V. Sanker  
SankerIP

Historically, “creativity” has been the realm of humans. There are many tools to assist a creative process, but one or more actual people control both the tools and the overall process. As a result, existing laws have evolved based on the assumption that inventors and authors are people. But now AI inventors and AI authors are stretching the existing laws, so users of generative AI systems should be cautious. There are several distinct scenarios to consider.

## **Risk #1: Possible loss of confidential information or corporate trade secrets**

Where does your information go when you converse with an AI system? In many cases, any text entered into a generative AI system is treated as public. This may seem counter-intuitive because it looks like a private conversation on the computer. Whether the conversation is private or not depends on the AI tool and the contractual arrangement (if any) between the user and the provider of the AI tool. In particular, when using a free public version of any generative AI tool, it is best to assume that the information will become public.

For intellectual property protection, any public disclosure has important consequences. If the disclosed information is a corporate trade secret, that protection is now lost. If the disclosed information relates to a patentable invention that is not yet filed with a patent office, the disclosure commences a one-year grace period for patent filing in the United States and precludes patent filing in all other countries.

Although most people would not intentionally consider disclosing protected information, accidental disclosure is more likely than it might appear. For example, consider a new invention that has three inventors. The inventors use generative AI regularly in their work (e.g., generating software code) and they have come up with a new idea that could be novel. To determine whether the invention is actually novel, one of the inventors queries a generative AI system, asking the system if there is anyone else developing an ABC device using a combination of a DEF unit and a GHI unit with a JKL controller. That alone could be enough to create an unintentional public disclosure of the invention.

In purely human-to-human interactions, the parties frequently use non-disclosure agreements to limit dissemination of the disclosure. A contract with the provider of a generative AI system may include similar language, but this should not be assumed.

The simple rule: don’t enter any information into a generative AI system that you want to keep secret.

### Create, monitor and review a workable corporate policy

First of all, creating a workable corporate policy for the usage of generative AI is essential. To make sure the policy *is* workable, it’s important to involve the people to whom the policy will apply. That is, find out how workers are already using generative AI and learn

how generative AI is making them more productive. It would be difficult, and perhaps undesirable, to impose a blanket prohibition against using a valuable tool.

Second, you need to monitor the gap between the policy objectives for generative AI and what workers are actually doing. As an analogy, consider the difference between speed limits in the United States and the speed people actually drive. Without enforcement, the disparity can be quite large. In the context of generative AI policy, the mechanism to monitor and enforce compliance will determine whether the policy is successful. For small organisations, human-based monitoring may be adequate, but for larger organisations, having IT monitor network traffic is probably needed. The need for monitoring and enforcement is also proportional to limits specified in the policy; the more draconian the policy, the greater the need to monitor and enforce.

Third, it's important to review and update the policy regularly. Generative AI tools are evolving quickly and many new tools are being released, so even a well-designed generative AI policy could become obsolete quickly. The rapid evolution of generative AI also imposes a practical limit for the complexity of a policy. A policy that focuses on the big issues and is concise enough to encourage workers to read it has a better chance of success. It is also helpful for the policy to explain the reasoning.

#### Review the provisions in your contract with the AI system provider

In the absence of a specific contract with an AI system provider, any use of the AI system should be treated as a public disclosure. The public disclosure extinguishes any trade secret protection and precludes patent protection outside the United States for any information that is entered. (In the United States, the public disclosure starts the one-year patent filing grace period if that period has not already started.)

Even when there is a contract with an AI system provider, there are many issues to consider:

#### *Security for data logged by an AI system provider*

Interactions with generative AI systems are generally logged in a database. One issue is finding out how much data is stored (e.g., the entire prompt) and evaluating how trustworthy the provider is. In this context, there are several facets to trustworthiness. First, is the provider being truthful about what information is stored? Because the data is stored in a location users can't access, it's useful to have either a reasonable level of trust in the company and/or certification by an independent third party. Second, is the provider being truthful about who has access to the data and how it's used? Again, it comes down to trust or third-party certification. Third, even if a provider is completely truthful about the data it stores and how the data is used, how good is the provider's IT infrastructure at preventing hackers and other bad actors from accessing and stealing data? In particular, is the security of the provider as good as the security provided by your own IT infrastructure?

#### *Public cloud/private cloud infrastructure*

If an organisation has a private cloud, and the AI system can run inside the private cloud without communicating to the outside, the configuration mitigates many of the risks. In this case, the data is usually as secure as any other data the organisation stores in the private cloud.

#### *Consider indemnification*

With the known and unknown risks of using any AI system, some AI systems include varying degrees of indemnification. If there is an indemnification clause in a contract, look at what it protects. It may not cover everything you want (e.g., it may not cover monetary damages in case of lost patent or other IP rights). The value of indemnification also depends

on the financial resources of the entity providing it. For example, a Fortune 100 company providing indemnification will have the resources to back up its promise if the need arises.

### *Cybersecurity*

If an AI system is running locally (e.g., on a user's laptop) or within a corporate firewall, existing security may be adequate. However, if critical data is transmitted outside of a secure firewall, the data may be intercepted. Data should be encrypted both *in motion* and *when stored*. And because quantum computing is coming soon, encryption should use a post-quantum algorithm (i.e., an algorithm that will not be easily broken as soon as quantum computers are available). Bad actors can intercept data now and decrypt it later when quantum computing becomes available, so it is better to employ post-quantum encryption sooner rather than later.

### *Use of your data for training the AI system*

There's an important distinction to make about training. If you're using a shared instance of an AI system, you should generally not permit the AI system to use your data for training. If you allowed it to do so, your proprietary data could be used to directly benefit others and potentially compete with your own research and development. Even in a purely academic or non-profit setting, allowing training based on your own work is undesirable because it lacks attribution to you as the author or inventor.

On the other hand, if you have a private instance of an AI system, then it *is* beneficial to you for the AI system to be fine-tuned using your data. Training an AI system based on your own data can increase efficiency and potentially lead to faster innovation. For example, if you have multiple engineering teams, and you train the AI system based on all their work, the integrated training may be mutually beneficial or lead to synergistic innovation that might not occur otherwise.

When using a private instance of an AI system, you need to be aware of potential issues when the core system is upgraded. If the AI system is designed well, upgrading the core system should include retraining based on your data so that you continue to have the benefit of your fine tuning (and the upgraded core system may use your data even more effectively than the earlier version). However, because some AI system providers are building systems quickly, the upgrade path may not be properly designed, and the fine tuning from your data may be lost. When evaluating whether to use a private instance of a particular AI system, it's wise to confirm how the system handles core upgrades.

Note that a supposedly private instance of an AI system might not necessarily be running in a private cloud, meaning that your instance could be "co-mingling" with any other instances. And, even when an instance of an AI system is running in a private cloud, it is important to understand what happens during an upgrade (e.g., is any data copied outside of the firewall?).

### *Summary*

Know how your information can be used and accessed by an AI system and determine whether you can entrust your data with the AI provider. If in doubt, it is better not to input anything confidential.

## **Risk #2: Possible lawsuits for copyright infringement**

Human content creators generally have copyrights for their work. These copyrights provide protection against the creation of "derivative works". A derivative work is "a work based upon one or more preexisting works, such as a translation, musical arrangement,

dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which a work may be recast, transformed or adapted”. See Chapter 1 of U.S. Copyright Law.

Generative AI systems are trained on a substantial *corpus* of existing content (e.g., scraped from the Internet), and much of that content is subject to copyright. Therefore, when anyone uses the output of a generative AI system, there is an argument that the output is a “derivative work” under copyright law. Some content creators have already filed lawsuits to enforce their rights.

An important countervailing argument is “fair use”, which is a complex legal doctrine that “promotes freedom of expression by permitting the unlicensed use of copyright-protected works” in certain circumstances. See Section 107 of the U.S. Copyright Act. There are four primary factors for fair use analysis, including:

- the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
- the effect of the use upon the potential market for or value of the copyrighted work.

These two factors are intuitively reasonable. If someone copies an entire work, it clearly constitutes copyright infringement. If nothing is copied, then there is no infringement. As long as the amount copied is a small portion of a work, it would favour classification as “fair use”; however, if the amount copied exceeds a certain threshold, it looks more like copyright infringement. The second factor here calls for looking at markets. Whenever a portion of an original work is copied, and that copied portion reduces the market for the original work, a court would most likely rule that the copy was not fair use. On the other hand, if a copied portion of an original work has no effect on the market for the original work, then it is likely to be considered fair use.

In general, training an AI system uses millions or even billions of training inputs, so any connection between a specific input and a specific output is tenuous. Because of the tenuous connection between the training inputs and the generated output, both of the factors discussed above support the fair use argument. Because any generated output uses very little from any one specific input, the generated output is unlikely to have any effect on the market for any of the inputs. However, if a specific output is close enough to a specific input that was used for training, the fair use argument is weaker. Courts in the United States will probably rule on this soon. Of course, any analysis of fair use is highly fact-dependent, so subsequent decisions could render opposite results.

Some plaintiffs may also argue that a trained AI model itself constitutes copyright infringement. The two fair use factors discussed above lean in opposite directions. For the first factor, the training may use the entirety of individual works instead of small portions. But, for the second factor, a trained AI model does not compete with any of the original works. In fact, the large number of stored parameters of an AI model are unlikely to be considered a “creative” work that is eligible for copyright.

Because the issue of copyright infringement for AI-generated works is unresolved, how can users of generative AI minimise their risk of lawsuits from human content creators? Consider:

- For small, generated works, the risk is rather low, particularly if distribution of the generated work is limited and/or internal to an organisation.
- Use existing tools to compare the generated output to known content. For example, use a generative AI system and ask if the generated output is similar to any other work. That is, use a second generative AI tool to evaluate the output of the first generative AI tool. There are also software systems specifically designed to identify plagiarism (e.g.,

in academia). These tools can quickly provide reasonable assurance that the generated work is not too similar to any other specific work.

- If in doubt, have one or more people modify the output created by generative AI and document the modifications. If arguments of copyright infringement arise later, the documented changes can bolster a fair use argument.

### **Risk #3: Possible inability to secure copyright or patent protection**

Both the US Copyright Office and the US Patent Office have ruled that works created solely by AI are not eligible for intellectual property protection. For works partially created by an AI system, the Copyright Office has held that the AI-generated portions are not eligible for protection. The USPTO announced its guidance for AI on February 12, 2024, and it bypasses the issue of AI inventors by focusing on the inventive contributions of human inventors.

#### Copyright protection for AI-generated content

Under current U.S. copyright law, there is no protection for AI-generated content. When an AI system generates an entire work, there is no protection at all; when an AI system generates portions of a work, the overall work and the portions not generated by AI can be protected by copyright, but the AI-generated portions are not protectable. See “*Zarya of the Dawn*” and the decision by the U.S. Copyright Office on February 21, 2023. In this example, the human author wrote the text and used an AI system to generate most of the images. Based on the Copyright Office decision, there is no protection for the individual images, so anyone can freely copy them.

An AI system does not spontaneously generate content out of thin air. Such systems generate content (e.g., text or images) in response to user prompts (e.g., text). This process is almost always iterative, particularly for generated images. In each iteration the user updates the prompt to generate output that is closer to what is desired. The U.S. Copyright Office currently does not consider the construction of the prompt to add to the creativity of the work, even when there are many iterations and many changes to the prompt by the user.

The decision to ignore any creativity in the input prompt is illustrated by the work “Theatre D’opera Spatial”, which won an art contest at a Colorado fair. According to the artist, Jason Allen, he envisioned the artwork beforehand, and it took “at least 624” iterations to get the final generated image. Despite the extent of human input required to tweak the output over 624 iterations, the U.S. Copyright Office focused solely on the fact that the final image was generated based on one final prompt.

In contrast to the U.S. Copyright Office, the Beijing Internet Court held on November 27, 2023, that the human artist, Mr. Li, “made a certain degree of intellectual investment in selecting prompt texts, setting up parameters, and designing the presentation”. According to the Court, Mr. Li “continuously added prompts and repeatedly adjusted the parameters to come up with a picture that reflected his aesthetic choice and personalized judgment”. The Court also noted that “to encourage creation is the essential purpose of the copyright system”.

Although the ruling by the Beijing Internet Court seems to better align with the objective of copyright law (“to promote the Progress of Science and useful Arts” according to the U.S. Constitution), it is important for now to work within the existing copyright framework: content generated by an AI system is not protectable by copyright.

#### *Recommendations for now and the future*

What can you do to get copyright protection for AI-generated work? Because the Copyright Office does not account for human creativity *before* the AI-generation step, the best current



solution is to apply human creativity *after* the AI-generation step to modify the output. It is not clear yet how much human-based modification is necessary to turn the work into something that is eligible for copyright protection, but it should be non-trivial. For example, changing a single word in generated text or changing a single pixel in a generated image would not be enough. And it is important to document the changes in case copyright eligibility is challenged later.

There is also an important legal analogy that may be pursued to change the way copyright law is applied to AI-generated works. Section 101 of the United States Copyright Act already provides for “Works Made for Hire”, enabling people or corporations with zero creative input to be considered the author for works created by others. This exception to core copyright law was created to account for the reality of how some works are created. By analogy, there is a plausible argument that a user of an AI system is entitled to a “work made for hire” under a similar exception. In fact, because the user of an AI system provides an appropriate prompt to the AI system (as in the examples above), there is arguably greater reason to grant copyright protection. Courts or Congress will need to address this argument.

#### Patent protection for AI-generated inventions

Courts throughout the world have held that inventions created entirely by AI systems are not patentable. *See, e.g., Thaler v. Vidal*, 43 F.4th 1207 (Fed. Cir. 2022). According to the filed patent application in the *Thaler* case, the sole inventor was the AI system called “DABUS”. But the *Thaler* case is not typical, because most inventions involve some human input. Therefore, the bigger question is whether an invention is eligible for patent protection when there are both human and AI inventors.

Prior to the release of AI guidance by the USPTO on February 12, 2024, the USPTO sought input on a variety of specific questions about AI inventors. The present author drew attention to the importance of addressing hybrid (human and AI) inventorship, pointed out three distinct ways to address this issue, and explained why the third of the three options best aligns with the goals of promoting innovation. *See* <https://www.regulations.gov/comment/PTO-P-2022-0045-0060>. It is useful to understand the February 12, 2024, AI guidance in the context of the possible options.

The first option would be to deny patent protection to any invention that has any AI-invented aspects. It would be hard to justify such a drastic policy, and it would be difficult to imagine that such a policy would “promote the Progress of Science”, as specified in the U.S. Constitution.

A second option would be to evaluate patentability of an invention based solely on the claim features that were invented by humans. Under this option, any features that were created by AI would be ignored. Initially this seems appealing, but it has some substantial implementation issues. This option would be analogous to the recent copyright decision for the novel “*Zarya of the Dawn*”, as discussed above. However, a copyright on portions not generated by AI is meaningful, whereas patents are granted on claims *as a whole*. Eliminating any claim elements could prevent patentability of an entire invention.

This second option is also similar to the analysis of “technical character” by patent examiners in Europe. Any claim elements that lack technical character are not included when evaluating patentability.

Trying to apply a “technical character” type analysis to differentiate between human- and AI-invented features would be impractical. Unlike “technical character”, which can be evaluated just by reading the words, there is no way to distinguish between human- and AI-invented

claim features without requiring substantial additional disclosure by patent applicants. Not only is there no statutory basis to require such additional disclosure, but neither patent applicants nor patent practitioners have an incentive to spend substantial additional time and money for the disclosure. Further, even if such a requirement were imposed, it would be both difficult and burdensome for applicants to comply because claim features do not align neatly with inventors, and claim features evolve during patent prosecution.

The third option is to require at least one human inventor and to apply the usual patentability requirements.

### *The USPTO guidance for AI takes a better path for hybrid inventorship*

The USPTO guidance from February 12, 2024, endorses the third option, focusing on the human inventors, and bypassing the question of whether any AI system might qualify as an inventor. (According to Guidance at Section II: “The *Thaler* decisions around “inventorship” are not a recognition of any limits on the current or future state of AI, but rather are an acknowledgment that the statutory language clearly limits inventorship on U.S. patents and patent applications to natural persons.”)

With AI-assisted inventions, it is possible to have human contributors whose contributions are insufficient to classify the contributors as inventors. For example, if a user asks an AI system to “build a better mousetrap” and it does build a better mousetrap, that user has not contributed enough to be an inventor. In fact, there may be no human “inventor” if the substantive inventive work is performed by an AI system. Therefore, an essential part of the AI guidance involves reviewing the caselaw that defines what constitutes a sufficiently significant contribution for a person to be named as an inventor. See Guidance at Section IV, subsection A.

In *Pannu v. Iolab Corp.*, the Federal Circuit provided three factors to evaluate what constitutes a significant contribution. *Pannu v. Iolab Corp.* 155 F.3d 1344 (Fed. Cir. 1998). The three factors are:

- significant contribution to the conception of the invention;
- the contribution is significant as a portion of the full invention; and
- the contribution is more than explaining well-known concepts or current state of the art.

The word “significant” appears in the first two factors, so they are not particularly helpful when faced with the practical question of which contributors are actually inventors. Future litigation, as well as future publications by the USPTO and patent practitioners, will need to elaborate on this.

### *Key takeaways for AI-assisted inventions*

First, make sure that every invention has at least one person who has made a significant contribution to the invention. Because the word “significant” is somewhat vague, aim higher rather than lower. Others may challenge your assessment later.

Second, when there is any doubt, document the contributions of the human inventors. It is useful to imagine future litigation in which an opposing party argues that the human contributions are insignificant. It could be very valuable to have documents or emails contemporaneous with the invention that describe the human contributions and describe how the human inventors used AI tools to assist in the inventive process.

Third, monitor internal hype about the role of AI in inventions. For example, because AI is a leading buzzword, a sales or marketing team might want to overstate the role of AI in a product. Such overstatements could be utilised by opposing parties in future litigation to invalidate patents.

**David V. Sanker****Tel: +1 510 714 5196 / Email: david@sankerIP.com**

David's path to becoming a patent attorney was atypical, but each step informed the next one. He earned a PhD in Mathematics from UC Berkeley in 1989 and then spent three years as an associate professor of mathematics and 12 years in production software development before law school. His experience as a software engineer was also unusual, as he took on several roles, including creating detailed technical requirements, writing and quality testing code, providing technical support to users, and designing and implementing an SQL database schema with nearly 500 highly interrelated tables.

David's career as a patent attorney began after he returned to Berkeley for a degree in law, graduating in 2007. A partner at Morgan Lewis for almost 20 years, David began in patent litigation, representing clients at the US Trade Commission, the US Federal Court and in federal district courts. He moved on to patent prosecution in a wide variety of technology areas, including software, AI, cybersecurity, semiconductor devices, database architecture, data visualisation, medical devices, artificial reality, virtual reality and identify verification. Being a patent litigator for five years turned out to be useful: when you see firsthand how patent claims are torn apart in litigation, you learn how to draft better patent claims.

In February 2024, David founded SankerIP, specialising in intellectual property and AI and backed by a team of scientists and engineers with advanced degrees and practical experience working in science and technology. David works with client companies, inventors, the US Patent Office and associates throughout the world to build strong IP portfolios, informed by his years of real experience as a software developer and database architect. David is also a thought leader in AI. In the past few years, he has frequently written and spoken publicly on the use of generative AI and how AI influences IP protection. In May 2023, he was asked to speak before the US Patent Office on the topic of AI.

Very early in life, David liked the number  $\pi$ , an irrational number with an infinite number of non-repeating digits. In the early 1970s, the Guinness Book of World Records started tracking a record for the most digits of  $\pi$  memorised. In 1978, while in junior high school, he twice broke the record: first at 6,350 digits and then at 10,000 digits after a summer family vacation gave him plenty of free time for study.

## SankerIP

4632 Queen Anne Ct, Union City, CA 94587, USA

Tel: +1 510 714 5196 / URL: [www.sankerip.com](http://www.sankerip.com)



**Global Legal Insights – AI, Machine Learning & Big Data** provides analysis, insight and intelligence across 22 jurisdictions, covering:

- Ownership/protection
- Antitrust/competition laws
- Board of directors/governance
- Regulations/government intervention
- Generative AI/foundation models
- AI in the workplace
- Implementation of AI/big data/machine learning into businesses
- Civil liability
- Criminal issues
- Discrimination and bias
- National security and military

---

[globallegalinsights.com](https://globallegalinsights.com)