

GenAI Content: who owns it and how to protect it?

15 December 2024, 07:19

Katrien Meuwis

Generative artificial intelligence (GenAI) is revolutionising the tech industry. Large language models (LLM) like ChatGPT’s OpenAI have made it accessible to almost anyone.

For tech companies, GenAI offers unparalleled opportunities for innovation, by autonomously creating content, designs, and solutions. This makes it easier and more cost-effective to develop products and services that previously required extensive human effort and creativity. But who owns AI-generated content? And how to protect this ownership? As in many cases, it’s a matter of preparation and making the right agreements.

What is Generative AI?

GenAI is a type of AI that can create new content from different inputs, such as text, images, audio, video and software code. The GenAI ecosystem has three main groups: LLM (Large Language Model) Developers, Tool Developers and Users. LLM Developers, like OpenAI, create the initial models and gather data to train them. GenAI Tool Developers create apps that expand the models' capabilities. Users, from students to professionals, use the models to obtain information, create content and perform tasks.

When new content is created from existing content, IP specialists pay close attention. This is especially true when it comes to data used to train models and the content they produce. Understanding these issues is crucial for protecting innovations, ensuring legal compliance and maintaining a competitive edge in the tech landscape.

Data protection and confidentiality

GenAI models and tools rely on large datasets. These may include copyrighted material or personal data, so businesses must ensure they are legitimate and secure.

3 types of data sets could be distinguished:

- **Training Data:** the data used to train or fine-tune the GenAI model or GenAI tool, helping it learn and generate output.
- **User Input Data:** The user provides content in various forms, including prompts, images, texts, videos, code and more.
- **Generated Output Data:** produced by the GenAI tool based on the user input, which the user can then use for various purposes.

Beware: your AI tool might not be treating your private data as such

All stakeholders in the ecosystem should be aware of how data and datasets are used. When users input confidential data into GenAI systems, the system may reuse that data in future interactions. Input or output data can leak through the AI system, risking public exposure of confidential information. This happened to Samsung in early 2024, after an engineer uploaded sensitive code into ChatGPT, leading to unintended exposure of this information. Four guidance are crucial in this context:

1. It is important to choose GenAI systems that tell you if user input or output data will be used for further training. Some versions of GenAI tools do not use user input or generated output for training.
2. GenAI tool developers must also check that the data used by the GenAI model is properly licensed to avoid potential copyright infringement claims from 3rd parties. Some companies offer protection to tool developers for claims of copyright infringement arising from the use of their models.
   Although the risk is lower, it's also advised for end-users to review those protection clauses.

When subscribing to a Gen AI tool, you can find clauses regarding ownership, licenses, warranties and indemnification in The Service level Agreement or Term of Use of the GenAI tool.

3. Furthermore, the tool must be built on high-quality, unbiased datasets to achieve **reliable and accurate output**.
4. It's important for developers of GenAI models and tools to be transparent about the datasets they use, so users can trust them and so they comply with regulations like the AI Act and data privacy laws. Many datasets contain personal data, so they must comply with the General Data Protection Regulation (GDPR).
5. It's also crucial to assess the security protocols of the GenAI system to protect against data breaches and unauthorised access.

Ownership and IPR Protection for GenAI Output

Determining who own AI-generated works largely depends on the terms & conditions of the GenAI tool. Further, protecting outputs like creative works, inventions, and designs present unique challenges.

Copyright protection

Copyright laws usually require human authorship, so purely AI-generated works are not copyright protected. However, AI-assisted creations can be protected if they clearly show human creative input.

For example, an author uses an AI tool to help write a novel by providing detailed prompts, selecting and refining the AI-generated text, and integrating her unique style and creativity in the work. This means the work may be eligible for copyright protection under the current EU copyright framework.

It is important to credit the human contributors and keep detailed records of all steps to create the work to support authorship. Without copyright protection, it can be hard to enforce IP rights, even if you own the work.

Patent protection

Patenting inventions that include generative AI features is feasible, but you may face challenges similar to those encountered when patenting software.

The European Patent Convention (EPC, Art. 52(2)) states that computer programs as such cannot be patented. But software can be if it provides a technical solution to a technical problem. Additionally, an ‘invention’ must be sufficiently disclosed. A person skilled in the art should be able to replicate it. AI systems often operate as "black boxes", making it very hard to discover their decision-making processes. This is what is required in patent applications to meet the "sufficiency of disclosure" requirements.

When it comes to patenting the output generated by generative AI, a significant challenge is that current laws do not recognise AI as an inventor. A recent patent application naming an AI system “DABUS” as an inventor was rejected in multiple countries. Currently, only human developers can be credited as inventors.

As patenting AI inventions is complex, it is advisable to consult experts in this field.

Trade Secret protection

AI-generated outputs can be protected as trade secrets if they are kept confidential have a commercial value and if the owner takes reasonable steps to maintain the secrecy. Therefor,

Users should ensure that the GenAI system’s contractual terms do not allow for the use of generated outputs as training data, which could compromise trade secret protection.

Legal terms and contractual strategies when using AI tools

The diversity in contractual terms for AI tools poses challenges, particularly regarding ownership of AI-generated outputs, usage rights, and obligations on confidentiality and data security. To further limit risks, companies should adopt clear contractual strategies that address key areas of concern:

- Contracts should state how training data can be used, including any copyright restrictions.
- Contracts should also say who owns the rights to AI-generated content.
- IP infringement claims arising from the use of AI tools and their outputs can be protected against with indemnification clauses

Paid versions of GenAI systems, typically offer stronger protections, including data confidentiality, liability coverage, and indemnification. They ensure user data isn't used for further training and comply with GDPR, unlike some free versions which may lack these safeguards. These features make professional versions often more suitable for businesses needing to protect proprietary information.

Standard terms from major generative AI providers are often non-negotiable for standard users, although enterprise clients may access customized agreements, while smaller or alternative tool providers may offer more flexible contractual terms.

Regulatory Developments

All actor sin the Gen AI ecosystem, whether LLM developers, tool developers and users sumust comply with regulations such as the AI Act, which established obligations across the Ai value chain. These include requirements for transparency, accountability, data governance and risk management, tailored to the risk level of the AI system.

Complex landscapes require competent guides

Generative AI presents significant opportunities and complex IP challenges for tech companies. Implementing robust IP strategies, securing appropriate licenses, and developing clear internal policies are essential steps in navigating the dynamic landscape of AI and intellectual property.

The Sirris Patent Cell, founded with the support of FOD Economy, is your contact point for all your questions related to this issue. Katrien Meuwis is Sirris' expert on this matter. She combines a scientific background with an extensive knowledge on intellectual property. You can contact her to set up an IP strategy or getting the right answers to your questions on IP asset management, valuation of IP, licensing, technology transfers, and collaboration contracts.

Want to know more on this topic?

Our expert Katrien Meuwis is just one click away.

Get in touch

Authors



Katrien Meuwis