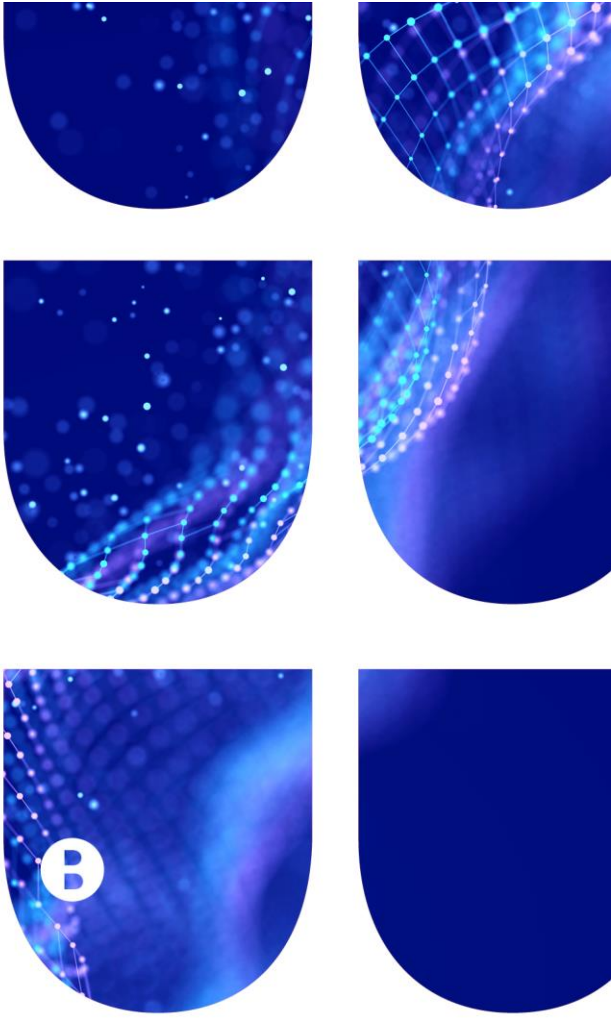
Bristows



Generative AI

Principles for employee and supplier policies.

Your employees and suppliers are already using ChatGPT and other generative AI tools.

Here's what to do about it.

Do you know how many of your employees are using ChatGPT? According to a Business Today survey, 70% of employees who use ChatGPT and other generative AI tools at work do so without their employers' knowledge.

And who can blame them? These products promise to be so useful that they'll be irresistible to employees. They will help your teams analyse complex documents, transcribe and summarise video calls and meetings, compose emails, write marketing content and software code, generate stunning imagery for campaigns, research complex topics and even brainstorm plans and strategies. These use cases have real potential to save a lot of time and massively boost productivity.

However, there are legal risks associated with their use, which is why some organisations including investment banks and the likes of Accenture, Verizon, Samsung and Apple have reportedly issued complete or partial bans until the products can be vetted and the risks better understood.

It's clear that the decision of whether, when and how to adopt generative AI should not be left to individual employees to figure out on their own. Instead, it's important for organisations to adopt a principled, policy-driven approach to how their people should engage with AI, and to do so sooner rather than later.

This is something we've been helping clients with in recent weeks, creating GenAI Acceptable Use Policies for employees and potentially also their third party suppliers involved in creating content on their behalf, tailored to business and sector. While the risks need to be considered on a case-by-case basis, there are some general principles which we think any policy should cover. Below we set out the top five issues we think any employee or supplier policy should address.

© Bristows LLP 2023

1. Commercially sensitive/confidential information

It can be very tempting for an employee to upload commercially sensitive and/or confidential information, including documents protected by legal privilege, into text-based generative AI tools. Imagine an account manager uploading client sales data into ChatGPT and asking it to produce a breakdown or analysis, or a sales person asking Microsoft Co-pilot to turn a detailed pitch document into a short slide deck. This issue is particularly important given the ability of these products to handle ever longer documents. For example, Anthropic's Claude can now ingest up to 75,000 words in a single context window, roughly the length of a short novel like The Great Gatsby.

Generally, the uploading of commercially sensitive and/or confidential information should be prohibited. Otherwise, there is a risk of breaching confidentiality obligations or legal protections that rely on confidentiality or trade secrets, as the upload could be deemed to be a disclosure to a third party or inconsistent with the need to take steps to maintain confidentiality. Note that most AI chatbots' default settings will allow them to use the data uploaded by a user for training and improving their underlying models, so a user would need to proactively opt out of this type of use of the data that they upload.

2. Personal data

Information uploaded to generative AI tools may well include personal data, for example of the organisation's employees or customers (think of a Slack message history uploaded to ChatGPT to produce a summary). Therefore, a policy should typically prohibit the uploading of personal data, and particularly special category data (e.g. data relating to someone's health, sexuality, ethnicity, etc.), or ensure this is limited to very specific scenarios depending on how the model is being used. Otherwise, this may result in a breach of data protection rules as personal data will likely be processed without the data subjects' knowledge and, in the case of special category data, without their consent.

Even if a user does not upload personal data themselves as an input to a model, the content generated by such tools could inadvertently include personal data from outside the organisation that formed part of the data on which the language model was trained. There have been cases of patients identifying their own medical information in generated content when they were not aware that the information was even available online, let alone part of a model's training dataset. An organisation using such content could, again, find itself in breach of data protection law despite having taken upload precautions.

3. Reliance on generated content

Text-to-text models often "hallucinate" and present false information that appears plausible, sometimes even accompanied by fabricated citations. Reliance on such content could lead to dissemination of misinformation, inaccurate or wrong advice being given to clients or misinformed business decisions. This could cause harm to an organisation's credibility and reputation, and potentially false advertising claims or breach of consumer protection laws or stock exchange rules, putting organisations at risk of regulatory enforcement action. Therefore, a generative AI policy should set guardrails around employees' reliance on ostensibly factual information generated by AI models, and encourage or require them to fact-check the output against reliable alternative sources. Practical guidance could be given, for example, to use these tools to produce a first pass or first draft of a desired output (as this is where much of the productivity benefit is) which should then be thoroughly reviewed and checked before it is finalised.

The risk of inaccurate or deficient outputs is also relevant for text-to-code models as, without appropriate human review, generated code that is buggy or of low quality could affect functionality of critical software. Poor quality generated code can also create security

© Bristows LLP 2023

threats, vulnerabilities or exploits. These potential issues mean that for technology sector organisations, text-to-code models may present the most tangible risks and legal teams may want to make them the focus of a generative AI policy.

4. Intellectual property ownership

Employees should ensure that when using text-to-image models, like Stable Diffusion or Midjourney, that appropriate intellectual property rights are granted to their employer from the model provider to ensure that their employer owns and has the necessary rights to commercially exploit the images. These tools are relatively new products and as such their terms of use take a variety of approaches to dealing with output material. For example, Midjourney has a tiered commercial model, and the free tier only grants the user a licence to use its outputs for non-commercial purposes and subject to attribution. Paying subscribers receive an assignment of rights in output material, subject to several exclusions. On the other hand, Stable Diffusion (as provided by Stability AI) grants all rights in its output content to its users across the board. However, its users still bear all the risk of third party claims of IP infringement relating to the outputs and to that end give a broad indemnity to Stability AI.

Furthermore, whether the outputs of these models even benefit from protection as copyright works may vary from country to country. For example, the US does not typically recognise computer-generated works as being protected by copyright, whereas in the UK they may be. Even in the UK there are some uncertainties around how copyright will apply to computer-generated works, given that the purpose of copyright law is to ensure the creativity and originality of human authors is protected. Ultimately, if outputs are not protected by copyright, an employer would not own any copyright in them and they could be copied and re-used by anyone for any purpose without fear of infringement. As such, particular care should be taken if images are going to be used for external marketing campaigns, or at all.

5. A white list of approved AI tools

Finally, given the ongoing proliferation of generative AI tools in the market, employers should consider creating a white list of approved and vetted products.

Most of the use we're currently aware of is of the public or free-to-use versions of chatbot AIs, while there are of course other types of AI tools that could deliver huge benefits and paid-for subscription versions which may be more reliable for use in the enterprise. As more third party vendors use large language models to make their own AI products more accurate, and offer protections against the types of risks described above, we can expect more organisations to use different types of AIs for particular use cases. To ensure such use is consistent with the type of AI governance an organisation has set out, both the policy and this list should be kept under regular review, particularly given the pace of change in this field.

We're seeing a real demand for help in this area from our clients over recent weeks, so if you'd like to know more, please do get in touch with us.