Fair use standards for generative AI

Introduction

In today's business landscape, integrating artificial intelligence (AI) brings numerous benefits, risks, and challenges. Understanding and addressing these risks, from data privacy concerns to algorithmic biases, is crucial for responsible and sustainable AI implementation.

Guidelines

1. As a data and technology driven business, we embrace new applications that leverage advanced data analytics and machine learning technology.

2. To ensure that we are trustworthy, responsible, and reasonably transparent in our uses of data analytics and machine learning, we will apply our (separate) Fair Use of AI and Personal Data Standards.

3. Generative AI (**GenAI**) tools (text generating ChatGPT, Google Bard, Microsoft CoPilot, image generating Midjourney and DALL-E, and many other GenAI tools that are now emerging) provide/create a myriad of opportunities to improve our working day and the quality of the services that we provide. Use of Generative AI also brings new and unfamiliar risks that we must manage. That is why, in addition to our Fair Use of AI and Personal Data Standards, we will apply these GenAI Fair Use Standards whenever we use GenAI.

4. These GenAI Fair Use Standards provides a framework for humans to make responsible decisions about whether and how to use GenAI for any task in conduct of our business. Because Generative AI will evolve rapidly, we will continue to review and update these GenAI Fair Use Standards to respond to the latest developments and consider the latest tools.

5. We will also adjust the Standards to reflect evolving global best practice, law and regulatory requirements. In developing the Standards , we have taken into account the Australian Government's interim statement of how the government will ensure AI is designed, developed and deployed safely and responsibly: https://consult.industry.gov.au/supporting-responsible-ai

6. By making this policy publicly available, we are also providing transparency to our clients and other stakeholders about how and when we will use Generative AI.

Commitments

We will use GenAI as follows:

1. **We won't incorporate confidential, sensitive or personal information into a publicly available GenAI tool.**

   Many tools in the current generation of publicly available GenAI tools do not provide sufficient security or privacy assurances relating to confidential, sensitive or personal information, particularly as to how input data and prompt queries may be used to train the tools or used or reproduced in generating further content.

   We will take a risk-based approach to inputting data, including through prompt queries.

   a) We will use publicly available GenAI tools only if our use is consistent with this policy and appropriate in the circumstances.

   b) We won't enter into a publicly available GenAI tool any personal, sensitive or confidential information relating to a client or other person or entity unless we have the relevant person or entity's express and informed consent.

   c) We will make sure the settings of any GenAI tools we use are managed to safeguard our information (including clients' information) to the extent appropriate. This may include (where appropriate) disabling training from our data. It will always be appropriate to disable training on confidential, sensitive or personal information. If information is anonymous, de-identified or publicly available, we will consider any risks before permitting training: we may permit training unless any risks are identified.

   d) When we buy or develop any enterprise/private GenAI tools, we will undertake an appropriate assessment (including security and privacy impact assessments where relevant) to ensure that the tools as deployed and used are secure, legally compliant and ethically sound.

   e) We will make sure our people are trained on the use of GenAI tools, so that we only use them in a secure, legally compliant and ethical way.

2. **We will check the outputs of GenAI before using them.**

   We know the current generation of GenAI tools can "hallucinate". This means they can fabricate, or present incorrect information relating to, facts, occurrences or people, and convincingly present erroneous or incorrect outputs as fact.

   A lot is still unknown or uncertain about the way GenAI tools create content.

   There are reasonable concerns that GenAI tools may be built or develop in a way that creates biased, discriminatory or misleading content, or infringes the legal or other rights (including copyright) of others.

   a) We will always check the outputs of GenAI before using them and keep a record of these checks as appropriate.

   b) We will also ensure the GenAI outputs are not the only source we use for any work product.

c)   Examples of checking the outputs include:

- Checking any assertion of fact or law produced by GenAI to ensure that it is accurate

- Keeping a record of any prompts used

- If using a substantial part of the output of GenAI in our work, we will: check the terms of the GenAI tool to make sure this is permitted

    o   check any knowledge cut-off dates for the GenAI tool

    o   check the content itself, to ensure it has not been copied from another source

    o   ensure the use of GenAI is appropriately disclosed/referenced in our work.

3.   We will be transparent about our use of GenAI – with each other and with clients and stakeholders.

a)   Each of our personnel will let their manager know if, and how, that person wants to use a GenAI tool to generate any output intended for use for any external exposure. Each of us will do this in advance of any use, so we can discuss the implications and agree an approach consistent with these Standards

b)   We will proactively disclose our use of GenAI tools to our clients, in advance of any use, where we believe there is additional risk involved in the particular use of a GenAI tool.

c)   We will also let any client know the specifics of our use of GenAI, if they ask.

4.   We will communicate in our own voice.

GenAI tools can produce content that is clear, concise and well-written.

However, the content is based on patterns and probabilities derived from data, so is inherently generic. It is relatively easy to check for "hallucinations" (clear errors of fact or law). It is harder to check for "generalisations" which may be imprecise or omit key details.

a)   We can use GenAI as a starting point for developing concepts but won't solely rely on it to produce detailed content or advice.

b)   Our humanity is one of our key assets. We won't outsource our voice, empathy, or strategic thinking to GenAI.

c)   We will always stand behind what we say. Our use of GenAI tools won't affect this.