City of Birmingham Interim Guidelines For Using Generative Artificial Intelligence (AI)

Purpose

Generative AI is a new technology that uses Large Language Models (LLM) along with some basic machine learning to produce content, based on questions from users. This content can be written or visual. These tools are very sophisticated models that predict what content satisfies the user's questions. **The impact, usefulness, as well as potential risks and dangers, are still the subjects of research by the AI committee. We believe, however, that employees of the City of Birmingham should follow some basic interim guidelines.**

Note: Generative AI is a tool. We are responsible for the outcomes of our tools. For example, if autocorrect unintentionally changes a word, or changes the meaning of something we write, the author is still responsible for the text. Technology enables our work; it does not excuse or substitute for our judgment or our accountability.

These interim guidelines will be replaced with permanent policies, standards, and training. However, we do want to encourage responsible experimentation. We encourage you to try these tools for yourselves to understand their potential. In doing so, we ask that you **use generative AI only as a jumping-off point** to give you general ideas or templates for your work, but not the finished work product.

Guidelines for Using Generative AI

1. Please **fact-check and review** all content generated by AI, especially if it is for public communication or decision-making.

    o <u>Why:</u> While generative AI can rapidly produce clear text or images, the information and content might be inaccurate, outdated, or simply made up. **It is your responsibility to verify that the information is accurate** by independently researching claims made by the AI.

    o <u>What to look for:</u>
      - Inaccurate information including references to events or facts.
      - Bias in the positions or information.
      - Copyrighted material that has been included.
    o <u>Prohibited</u>:
      - The verbatim copy and paste of content produced by generative AI without credit.

- The submittal of unedited work; work must have human interaction before submittal.
- The use of macros, links, scripts, QR codes, or other potentially dangerous material without the express consent of the IT Department.

2. **Consider limiting the way you use Generative AI.** Some people find it less complicated to only use it for certain functions, for example, specialized searches, data collection, etc., but not to summarize or analyze content. Be aware that you can still get skewed information and need to check it for completeness and accuracy.
   Examples of acceptable use include:
   - Idea starters
   - Document outline
   - Changing voice or target of existing work
   - Comparing researched content,
   - Organizing your thoughts or ideas on a subject
   - Generating graphic representation of (fact-checked) data
   - Organizing researched data

3. **We want to make sure that vulnerable populations are not harmed by these technologies.** Think about how racial and ethnic minorities, different genders, people with disabilities or language barriers, or others could be negatively portrayed or impacted by the content.

4. **Disclose that you have used AI to generate the content,** and to what extent. You should also include the version and type of model you used (e.g., Open AI's GPT 3.5 vs Google's Bard). You should include a reference as a footer to the fact that you used generative AI:
   - Sample credit line: "Outline was generated by ChatGPT 3.5 and content edited by Eric Brunk"
   - Sample credit line: "This text was summarized using Google Bard"
   - Sample credit line: "Written by Eric Brunk with assistance from Generative AI."

5. We ask that you **do not share sensitive or private information** in the prompts or questions.
   - Why: data including prompts used in generative AI might be accessible by others using the LLMs that power these systems. Any information that includes personally identifying information about our residents, other public servants, etc. could inadvertently be shared with others. If you wouldn't share the data with other people or put the data in a public place, avoid sharing the information in the prompt/question

To further support and improve these guidelines, the Information Technology (IT) department is looking into training and workshops that can support users and departments interested in learning more about the technology and how to use it to augment your daily tasks. If you have questions please contact the IT department for more information.

Acknowledgments

The development of these guidelines has benefited from the contributions of generative AI, the Birmingham AI committee, and from reviewing guidelines of the City of Boston.