Getting Started with Generative AI for Research: Microsoft Copilot Hands-on Workshop

June 24th, 2024 | 1:30 pm – 3:00 pm

Presented by:

- **Jordan Holmes**, Senior Manager, Teaching, Learning & Technology | Centre for Teaching Support & Innovation (CTSI) and Academic, Research & Collaborative (ARC) Technologies

- **Tori Sheldon**, Faculty Liaison Coordinator, Generative AI Pedagogies, Centre for Teaching Support & Innovation (CTSI) and Academic, Research & Collaborative (ARC) Technologies

- **Derek Hunt**, Faculty Liaison, Teaching, Learning & Technology, Centre for Teaching Support & Innovation (CTSI) and Academic, Research & Collaborative (ARC) Technologies

Land Acknowledgement

We wish to acknowledge this land on which the University of Toronto operates.

For thousands of years it has been the traditional land of the Huron-Wendat, the Seneca, and the Mississaugas of the Credit.

Today, this meeting place is still the home to many Indigenous people from across Turtle Island and we are grateful to have the opportunity to work on this land.

Presenters

Jordan Holmes

Senior Manager, Teaching, Learning & Technology | Centre for Teaching Support & Innovation (CTSI) and Academic, Research & Collaborative (ARC) Technologies

Tori Sheldon

Faculty Liaison Coordinator, Generative AI Pedagogies, Centre for Teaching Support & Innovation (CTSI) and Academic, Research & Collaborative (ARC) Technologies

Derek Hunt

Faculty Liaison, Teaching, Learning & Technology, Centre for Teaching Support & Innovation (CTSI) and Academic, Research & Collaborative (ARC) Technologies

Stephen Beauregard

Research Tools & Technology Liaison, Centre for Research & Innovation Support (CRIS)

Agenda

- What is Generative AI and How Does it Work?

- Promise & Pitfalls of Generative AI

- Applications, Challenges, and Considerations for Research

- Introduction to Microsoft Copilot

- Bio/stretch break

- Hands-on Copilot Activities

- Discussion & Q&A

Image generated using Midjourney

What is Generative AI?

Generative AI

- Refers to **artificial intelligence** that can generate new content, such as text, computer code, images, and audio, based on patterns learned from data

- It works by using machine learning algorithms, particularly neural networks, to analyze and learn from large datasets

- Large language models (LLMs) are a class of generative AI capable of contextualizing and generating human-like text

How does it work?

- To grasp a word's meaning LLMs first observe it in context using enormous sets of training data

Words into Numbers

- When the LLM processes text, each word is encoded as a "vector"

- A vector in an LLM represents a word or phrase as a series of numbers, capturing its meaning and relationship with other words

Generation

- Once trained, the AI can generate responses that resemble its training data but are original creations

- Generation is **probabilistic**: each successive word in a series of text has a certain probability of being placed there, based on the words that have come before

- This process also applies to images, code, and audio

UNIVERSITY OF TORONTO | Centre for Research & Innovation Support

Let's gamble…

"It was the best of times, it was the [BLANK]"

- How does generative AI select the next word?

- It rolls a *weighted* die:

  - 95%: "worst"

  - 4%: "blurst" (thank you, The Simpsons)

  - 1%: all other words in the training data

- The model does NOT just choose the word with the highest probability!

- This approach has strengths and limitations

Promise & Pitfalls

Generative AI *sounds* convincing…

… but can "hallucinate"

Here's What Happens When Your Lawyer Uses ChatGPT

A lawyer representing a man who sued an airline relied on artificial intelligence to help prepare a court filing. It did not go well.

A Man Sued Avianca Airline. His Lawyer Used ChatGPT. - The New York Times (nytimes.com)

OpenAI Hit With First Defamation Suit Over ChatGPT Hallucination

OpenAI Hit With First Defamation Suit Over ChatGPT Hallucination (bloomberglaw.com)

Hallucination in Images

UNIVERSITY OF TORONTO | Centre for Research & Innovation Support

13

Inherent Bias

- Generative AI reflects the biases in its training dataset:

Lighter skin: I, II, III — Darker skin: IV, V, VI

LAWYER

SOCIAL WORKER

Generative AI Takes Stereotypes and Bias From Bad to Worse (bloomberg.com)

14

Inherent Bias - Gender

Perceived Gender: ■ Man ■ Woman ■ Ambiguous

LAWYER

SOCIAL WORKER

Generative AI Takes Stereotypes and Bias From Bad to Worse (bloomberg.com)

University of TORONTO | Centre for Research & Innovation Support

"Average" Lawyer and Social Worker

LAWYER

SOCIAL WORKER

AI bias can be worse than reality

Sources: Bureau of Labor Statistics, American Medical Association, National Association of Women Judges, Federal Judicial Center, Bloomberg analysis of Stable Diffusion

Bias and Hallucination Mitigation Strategies

- Pre-training phase: data over-sampling, under-sampling

- Model fine-tuning: Human feedback on generated content

- Retrieval-Augmented Generation (RAG): Use of external knowledge stores (e.g., course materials) to weigh responses towards relevant/current information, "sources of truth"

- Prompt augmentation: Automatically (and invisibly) appending text to users' prompts to increase diversity of output, e.g.,

  o "Draw a picture of students in a lecture hall [include a variety of genders, ethnic origins, and cultural representations]"

UNIVERSITY OF TORONTO | Centre for Research & Innovation Support

APPLICATIONS FOR RESEARCH

Specific Applications in Research

Most applications are ways of speeding up or enriching key research stages and activities:

Planning

- **Literature Search:** Gen-AI powered discovery tools mitigate information overload and direct users to relevant content

- **Synthesis:** GenAI tools can quickly summarize and synthesize large volumes of material

- **Ideation**: GenAI tools can serve as brainstorming partners, helping to generate ideas, perspectives, and potential solutions

Creating

- **Data Analysis/Visualization:** GenAI applications can extract, prepare, analyze, and interpret datasets

- **Coding/Programming**: Tools can help users write, improve, and debug code, as well as create low-/no-code apps to support projects

Sharing

- **Writing & Presenting:** GenAI applications can provide writing and editing assistance, as well as create presentations of research

UNIVERSITY OF TORONTO | Centre for Research & Innovation Support

Challenges

The rapid development and use of GenAI in research poses many important risks, including to:

Copyright & Intellectual Property

- **Training Data:** The legality of content used to train AI models is unknown in many cases; if and how far the fair-use doctrine can be applied remains unclear

- **Output**: Authorship and ownership of works created by AI is unclear; <u>major scholarly publishers do not allow AI to be attributed authors</u>

- **Library-licensed** e-resources (e.g., journal articles) should not be uploaded to public-facing (unprotected) GenAI tools, as this could violate vendor agreements

Data Privacy / Research Security & Ethics

- Any cloud-based GenAI tool presents data privacy risks, as users cannot be certain what happens to their input data and if it will be retained for future training of the AI model

- Confidential, sensitive, or proprietary information should not be shared without understanding the policies governing the use of data in GenAI tools
    - **Example**: Level-3 Data can be shared while using Microsoft Pilot in 'Protected Mode'

Methodological Rigor

- Extensive use of GenAI tools as "shortcuts" could lead to a degradation of research skills and general methodological rigor

UNIVERSITY OF TORONTO | Centre for Research & Innovation Support

Research Integrity Considerations

The use of GenAI in research should always be guided by the highest ethical standards and corresponding duties

- **Honesty & Accuracy - Duty of Verification**
  - Always verify the accuracy and validity of GenAI outputs – the responsibility for research accuracy remains with researchers

- **Due Acknowledgement - Duty of Disclosure**
  - Always document and provide disclosure of any GenAI use in all aspects of research, in accordance with principles of reproducibility, transparency, authorship, and inventorship

- **Confidentiality - Duty of Discretion**
  - Do not share confidential, sensitive, or proprietary information without understanding the policies governing the use of data in GenAI tools
  - Do not assume GenAI output is already considered part of the public domain (i.e., not restricted by copyright)

Always adhere to any restrictions or limitations on the use of GenAI put in place by the institution, research sponsors, collaborators, publishers, etc.

UNIVERSITY OF TORONTO | Centre for Research & Innovation Support

Microsoft Copilot

uoft.me/copilot

University of Toronto "Protected" Copilot

- Microsoft Copilot is mostly ChatGPT
    - Can be used to replace what you might do in ChatGPT
        - Can use OpenAI 4.0
        - Uses DALL-E 3 to generate images
        - Can access the web to find references
    - Does more than ChatGPT:
        - It does not use your data for training.
        - Data is not saved or used by Microsoft.
        - Use the "sidebar" of the Edge browser to access files/content

- The Microsoft Copilot "Protected" version has been evaluated by the University's Information Security team, and it has been deemed safe to use for up to Level 3 (Three) data avoiding the use of any private, sensitive, personally identifiable or confidential information.

- Bing Chat
- Bing Chat Enterprise
- **Microsoft Copilot**
- Microsoft Copilot with commercial data protection
- Commercial data protection for Microsoft Copilot
- ~~Microsoft 365 Copilot~~

University "Protected" Copilot:

Navigate to **copilot.microsoft.com**
select "Sign in" with a work or school account.

Edge browser, Copilot "sidebar" experience

Copilot sidebar experience by selecting the Copilot icon at the top right corner of the Edge browser.

Optional: Create a profile in the Edge browser.

UNIVERSITY OF TORONTO | Centre for Research & Innovation Support

https://teaching.utoronto.ca/tool-guides/microsoft-copilot/

Chat   Compose   Protected

UNIVERSITY OF
TORONTO

CENTRE FOR TEACHING SUPPORT & INNOVATION

Home   Contact   What We Offer ▾   Events   Resources ▾   Search ▾ ⌕

Copilot

Your everyday AI companion

with b Bing

Home » Educational Technology » Tool Guides » Microsoft Copilot

Choose a conversation style

Microsoft Copilot

| More Creative | More Balanced | More Precise |

What Can I Use It For?

Here are some things Copilot can help you do.

https://copilot.microsoft.com

Work

Show me all my favorites

Copilot

Chat   Compose

How can teachers make lectures more engaging and effective?

Your everyday AI companion

Protected

with b Bing

Bing Chat
Enterprise

Make a syllabus template for a college-level course

Microsoft Copilot is powered by AI, so surprises and mistakes are possible. Please share feedback so we can improve! Terms · Privacy

AI-powered copilot for the web   Preview

Copilot is powered by AI, so surprises and mistakes are possible. Terms | Privacy | Ads

Choose a conversation style

Choose a conversation style

| More Creative | More Balanced | More Precise |

| More Creative | More Balanced | More Precise |

This Academic Toolbox tool helps you... [?]
Organize content /
Connect & communicate /
Teach from a distance

Your personal and company data are protected in this chat

Your personal and company data are protected in this chat

Your personal and company data are protected in this chat

Typical course activity format: [?]
Asynchronous

Ask me anything...

Ask me anything...

Quercus integration
Non-integrated tool

0/2000

0/4000

Where can I get more support?
- Vendor (Microsoft)

- Email CTSI or book a consultation

With **Microsoft Copilot**, you can ask the AI chatbot questions and get detailed responses with footnotes that link back to original sources. Because it is connected to Microsoft's search engine, it has the ability to provide users with up-to-date information and real links, which may make it a better research and teaching tool than ChatGPT.

Cost
Centrally funded

Your personal and company data are protected in this chat

In addition to its utility as a research tool, Copilot also has creative modes, which would allow users to leverage its writing capabilities and code generation, in addition to other basic tasks, such as solving math

Ask me anything...

0/2000

copilot.microsoft.com

- Open a local file (pdf) with the Edge browser.

On a PC – Open Windows Explorer - find your pdf file first, then right-click and select "Open with"

On a MAC –

Open Edge, then select "File" and "Open File…"

Copilot modes

- Creative: whenever you need to generate new content, longer output, more expressive, slower (GPT 4)

- Balanced: best for the most common tasks, like search, maximum speed (GPT 3.5)

- Precise: most factual, minimizing conjectures (GPT 4)

As you are working with Generative AI

- **Use it as a tool, not as a source of information**
    (at least not a sole source of information)

- It can be "confidently wrong"
    - Check its work!

ai.utoronto.ca

Prompt Strategy – 3 P's

- Provide a **persona**
  - "You are a university instructor in the department of anthropology."

- Provide a **purpose**
  - "You are developing a 12-week graduate course on ethnographic research methods."

- Outline your **parameters**
  - "Generate a syllabus for this course. Start with a description, followed by five learning outcomes. The learning outcomes should follow Bloom's taxonomy. Generate topics for each of the 12 weeks. The topics should build on each-other sequentially. This is a seminar format."

uoft.me/copilot

UNIVERSITY OF TORONTO | Centre for Research & Innovation Support

BREAK

- Stretch / Bio Break

- We will start again in 5 minutes

Hands-on Activities with Copilot

Use the "3P's" strategy (i.e., provide a persona, purpose, and parameters) when working through the following use cases with Copilot. Experiment with different levels of detail for each of the persona, purpose, and/or parameter variables in your prompts – how does this affect the output? Try Copilot's different 'conversational style' settings (creative; balanced; precise) – how does this affect the output?

Use Case #1 – Idea Conception

- Use Copilot as a brainstorming partner by inputting a given research concept and asking the tool to suggest experimental approaches, potential methodologies, and/or alternative research questions.

Example Prompt:

"You are a [persona], and you are brainstorming possible topics for a research paper. Analyze recent research on memory consolidation and the influence of emotions on learning and recall. Based on this analysis, generate new hypotheses for potential studies investigating neurobiological mechanisms."

Use Case #2 – Literature Search

- Use Copilot to help identify relevant literature for your research concept. For any given research topic, this may consist of asking the tool to identify key studies/findings, or top journals and publications to explore.

Example Prompt:

"You are a [persona], and you are researching the influence of emotions on learning and recall. Generate a list of the top journals publishing research on this topic."

Use Case #3 – Research Presentation

- Use Copilot to create a summary or presentation of your research, adjusting the parameters of your prompt to create presentations that suit different audiences (e.g., discipline, knowledge levels, etc.). As research material, browse to an online article or upload a local copy to the tool.

Example Prompt:

"You are a [persona], and you will be presenting your research at an academic conference. Create a PowerPoint presentation outline that summarizes this research article. The presentation should be 10 slides or less and should be suitable for university graduates."

UNIVERSITY OF TORONTO | Centre for Research & Innovation Support

Discussion

- What were your top takeaways from your work with Copilot today?

- What are the most important implications you see for the use of GenAI in your own research context?

- What research activities can you see yourself using GenAI for?

- What would you like to learn more about?

UNIVERSITY OF TORONTO | Centre for Research & Innovation Support

Q&A

- Please raise your hand and we will bring over a mic.

Thank You!

Thank you to our presenters and to you for attending!

- Links for the workshop slides, support resources, and a feedback form will be sent out after the session.

- Follow-up questions can be addressed to cris@utoronto.ca

UNIVERSITY OF TORONTO | Centre for Research & Innovation Support