GENERATIVE ARTIFICIAL INTELLIGENCE (GENAI) OVERVIEW

Version 1.0 June 2024

Co-created by

Amanda McKenzie, MA
Karen Lochead, PhD
Kari D. Weaver, MLIS, EdD
Lauren Byl, MLIS
Clare Bermingham, PhD
Julie Piatek
Trevor Holmes, PhD
Scott Anderson, MES

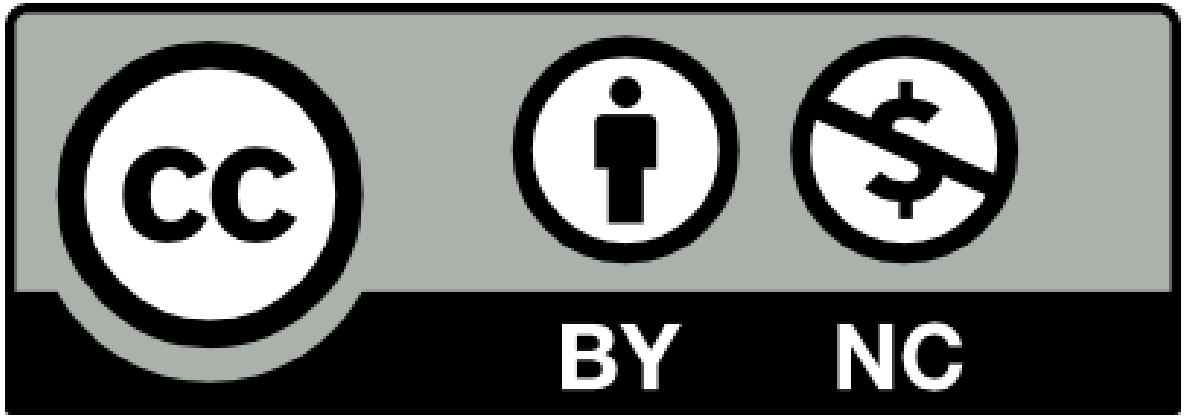
UNIVERSITY OF WATERLOO

INTRODUCTION

This overview was developed based on a comprehensive slide deck from University of Massachusetts Amherst, Dr. Torrey Trust: "AI & Ethics" slide deck by Torrey Trust, Ph.D. is licensed under CC BY NC 4.0.

This adapted version aims to provide the reader with a foundational understanding of GenAI from a University of Waterloo perspective. This version of the slide deck retains the CC-BY-NC license, and can be credited to:

McKenzie, A., Lochead, K., Weaver, K. D., Byl, L., Bermingham, C., Piatek, J., Holmes, T., & Anderson, S. (2024). *Generative artificial intelligence (GenAI) overview* [Slide deck]. University of Waterloo Associate Vice-President, Academic Standing Committee on New Technologies, Pedagogies, and Academic Integrity. Adapted from T. Trust (2024). *AI & ethics* [slide deck]. University of Massachusetts Amherst as allowed under CC BY-NC 4.0.



https://creativecommons.org/licenses/by-nc/4.0/

TOPICS COVERED

WHAT IS AI?

UNIVERSITY OF WATERLOO

WHAT IS ARTIFICIAL INTELLIGENCE (AI)?

- Broadly speaking, **artificial intelligence (AI)** refers to the ability of computer systems to simulate human-like cognitive capabilities such as learning, decision-making, and problem solving. The term was coined by Stanford University computer scientist John McCarthy in 1955.
- Although AI can include a wide variety of methods and tools, **machine learning (ML)** is most commonly associated with traditional AI.
- In simple terms, **machine learning (ML) is a method of data analysis** that automates pattern identification giving a computer system the ability to make decisions and predictions based on what it has “learned” from a dataset.
- Machine learning can be **unsupervised** (training from labeled data), **supervised** (training from unlabeled data), or **reinforcement based** (training from mistakes).

EXAMPLES OF AI IN EVERYDAY LIFE

- spam filters
- search engines
- chatbots (computer programs, software applications, or web interfaces that simulate and process conversations with users through text and/or audio interactions)
- unlocking your phone with facial recognition
- unlocking your computer with fingerprint scanning
- digital voice assistants like Siri and Alexa
- recommendations on video streaming services
- social media feed content including ads

WHAT IS GENAI?

AN INTRODUCTION TO GENAI

- **Generative artificial intelligence (GenAI)** is a type of artificial intelligence that can create new content such as text, code, images, audio, and video by extrapolating from its training data.
- Unlike traditional AI that are limited to a set of predefined rules and templates, GenAI are “blank brains (neural networks) that are exposed to the world via training on real-world data. They then independently develop intelligence—a representative model of how that world works—that they use to generate novel content in response to prompts [i.e. queries from users].”[1]
- **Traditional AI** excels at **pattern recognition**. It processes data to perform preestablished tasks (like providing analyses or make predictions) based on predetermined algorithms and predefined rules.
- **Generative AI**, by contrast, excels at **pattern creation**. It creates or ‘generates’ new data that resembles, rather than replicates, its training dataset.

AI MODELS AND THE DEVELOPMENT OF GENAI

There are three models of AI particularly important to the development of GenAI:

- **Large language models (LLMs**) are deep learning algorithms trained on massive collections of text data. This training process allows LLMs to perform a variety of language processing tasks including recognizing, translating, predicting, or generating text and other content including code.
- **Generative adversarial networks (GANs)** are paired sets of deep learning algorithms called neural networks that are simultaneously trained. One, the generator, is trained to produce a novel output based on labels (e.g. ‘dog’, ‘brown’, ‘friendly’), while the second, the discriminator, tries to classify examples as either real (from the domain) or fake (produced by the generator) until the discriminator is regularly ‘tricked’ by the generator. GANs can output audio, video, and 3D models as well as images.
- **Variational autoencoders (VAEs**) are unsupervised neural networks that learn to ignore ‘noise’ in data through a process of first compressing and encoding data, and then reconstructing the compressed and encoded data to a representation that is as close to the original input as possible. VAEs can be used for file compression, image de-noising, image transformation, and audio synthesis.

WHAT IS CHATGPT?

UNIVERSITY OF WATERLOO

CHATGPT BASICS

GenAI deployments or applications are commonly called tools so, in the simplest of terms, ChatGPT is a **GenAI tool**.

- More precisely, ChatGPT is a **text-generation GenAI tool** directed toward holding conversations ('**Chat**s') with users by responding to user questions, directions, and instructions (i.e. 'prompts') very quickly and clearly.
- Explained in a bit more detail, ChatGPT is a **large language model (LLM) of artificial intelligence** that is **Generative**, meaning it produces outputs by way of **Pretrained** algorithms. Pretrained algorithms have been fed all the data they need to carry out tasks using a **Transformer**, which is a type of neural network that is particularly good at processing sequential data = **Generative Pretrained Transformer (GPT).**
- ChatGPT isn't built on particularly new, innovative, or breakthrough discoveries or technology.[2,3] What makes it notable is that it's the first tool of its kind to be released to the general public for free and open use (by OpenAI in late November 2022).

HOW DOES CHATGPT WORK?

In simple terms, ChatGPT uses its training data to iteratively predict a reasonable continuation of whatever text it has so far, but with an element of randomness. This randomness (a 'temperature parameter') is responsible for ChatGPT's ability to produce unique responses to the same prompt as well as the illusion of human-like creativity.

- "[L]et's say we've got the text "*The best thing about AI is its ability to".* Imagine scanning billions of pages of human-written text (say on the web and in digitized books) and finding all instances of this text—then seeing what word comes next what fraction of the time …

  [A]t each step [ChatGPT] gets a list of words with probabilities. But which word should it actually pick to add to the essay (or whatever) that it's writing? If we always pick the highest ranked word, we'll typically get a very "flat" essay …

  But if sometimes we pick lower-ranked words, we get a 'more interesting' essay. The fact there's randomness here mean that if we use the same prompt multiple times, we're likely to get different essays [or whatever] each time."[4]

POPULAR GENAI TOOLS

UNIVERSITY OF WATERLOO

GENAI TOOLS

What follows is **not a definitive list**. GenAI Tools are being created, updated, rebranded, and shuttered daily.

This is **not an endorsement**. Individual GenAI tools have unique features and applications, as well as specific and generic strengths, weakness, merits, demerits, biases, and harms, as well as ethics-, security-, and privacy-related considerations.

Looking for other GenAI tools? Consult one of these GenAI compendiums:

- There's an AI for That
- AIcylopedia

TEXT-GENERATION TOOLS (aka CHATBOTS)

GenAI Tool	Description
ChatGPT	A GenAI chatbot developed by OpenAI. It uses Natural Language Processing (NLP) to answer questions, generate content, summarize information, and more. It runs on a large language model (currently GPT-4o; reduced to GPT-3.5 when usage limit is exceeded) and has access to the internet. UWaterloo does not provide Team or Enterprise licenses. While personal accounts can be used for experimentation and to boost productivity, University data should never be entered into this system. Free to use.
ChatGPT Plus	This paid version of ChatGPT provides access to the latest updates (such as improved models), the latest model of DALL-E (OpenAI's GenAI image generator), and expanded token use (meaning more units of data inputs). This model currently accepts text and image inputs. University data should never be entered into this system. Paid subscription required.
Claude	Developed by Anthropic, this is a chatbot based on the Claude 3 large language model. It uses Constitutional AI, integrating AI and legal frameworks to self-govern and eliminate human data labeling, to support operations within ethical and legal limits. Claude is designed to be self-contained but can access the internet. University data should never be entered into this system. Free (with limits) and paid (Pro) versions. Not currently available in Canada.

TEXT-GENERATION TOOLS (aka CHATBOTS)

GenAI Tool	Description
Copilot (not the same as Copilot for Microsoft 365)	IST's recommended GenAI tool at UWaterloo. This Microsoft GenAI chatbot currently runs on OpenAI's GPT-4 large language model. Accessing this tool with your UWaterloo credentials will automatically provide you with personal and company data protection. Free to use with your UWaterloo credentials.
Copilot for Microsoft 365 (not the same as Copilot)	An AI 'assistant' that works within Microsoft applications such as Teams, Word, Excel, and PowerPoint. UWaterloo does not currently provide Copilot for Microsoft 365 licenses but is investigating a secure and equitable access strategy. University data should never be entered into this system. Requires a Microsoft 365 licence and a paid subscription.
Gemini (formerly called Bard)	Google's AI chatbot. It runs on a large language model Google also calls Gemini. Integrated with Google applications and can be connected to other Google services. Free (basic) and paid (advanced) versions available. University data should never be entered into this system. Requires a personal Google account.

IMAGE-GENERATION TOOLS

GenAI Tool	Description
Craiyon (previously known as DALL-E mini)	Enables users to create images from text inputs in natural language. Free and paid versions.
DALLE-3	OpenAI's latest AI-powered image generator. Enables users to create images from text inputs in natural language. Requires ChatGPT login. Free to use.
Firefly	Adobe's AI image generator. Available as a standalone web application and integrated into Adobe's 'Creative Cloud' applications including Photoshop, Illustrator, Adobe Express, and Adobe Stock. Trained on licensed content, such as Adobe Stock, and public domain content where copyright has expired to support responsible GenAI image creation. Credit-based pricing (free version includes 25 credits per month; 1 credit = approx. 1 image)
Midjourney	An AI image generator that works through the Discord app. Enables users to create images from text inputs in natural language. Paid plan required.
Stable Diffusion	A GenAI-powered open source image generator by Stability.ai. Enables users to create images with a prompt (text-to-image generation) or an image + a prompt (image-to-image generation). Credit-based pricing (first 25 credits are free).

SPECIFIC PURPOSE TOOLS

GenAI Tool	Description
AskYourPDF	A GenAI tool designed to turn documents into chat partners. Supported formats: PDF, TXT, PPT, PPTX, EPUB, RTF. Free and paid plans.
Consensus	A GenAI tool designed to help users find evidence-based answers/insights in research papers. Free and paid versions.
Elicit	A GenAI research assistant. It can find relevant research papers and connected papers, generate abstracts, find themes and concepts across papers, and more. Free (5000 credits) trial and then use requires a paid subscription.
GitHub Copilot	A GenAI coding assistant. It translates natural human language into programming code. University data should never be entered into this system. Free for verified students, teachers, and maintainers of popular open source projects. Free and paid versions.
Goblin Tools	A collection of small, simple, single-task tools, mostly designed to help neurodivergent people with tasks they find overwhelming or difficult, for example: Magic ToDo helps to breakdown tasks (like writing a 10-page research paper) into steps. Free.

SPECIFIC PURPOSE TOOLS

GenAI Tool	Description
Grammarly	A GenAI writing assistant that now generates alternative written sections as part of recommending revisions and suggesting changes for clarity, style, grammar, spelling, and punctuation. Used by students for over a decade, its functionalities have expanded considerably since 2022 and new capabilities are similar to ChatGPT and other text-generating AI tools. Free and paid (premium) versions.
Jenni AI	An academic paper writing companion that can generate essay outlines, relevant literature suggestions, citation assistance, tone and style variations, summaries, paraphrasing, and feedback on writing. Claims a commitment to academic integrity - “requires [user] input and critical thinking to produce the final product.” Free (limited) and paid pricing options.
Otter AI	A meeting assistant designed to record, transcribe, caption, and summarize meetings (including in person and online lectures). In virtual sessions, Otter AI automatically captures lecture slides and adds them to the notes. Can be embedded in Zoom, Teams, and Google Meet.
Perplexity	Like ChatGPT in many respects but specifically designed for research; a chatbot-style search engine and content generator. No log-in or account required. Free.

SPECIFIC PURPOSE TOOLS

GenAI Tool	Description
Research Rabbit	A scholarly publication discovery tool. Helps users discover publications related to one or more seed publications with the help of visualization maps and lists of earlier, later, and similar publications. Free.
Smodin Omni	A ‘homework solver’ and AI tutor geared predominantly toward STEM fields. Free.
Sora	An AI media tool developed by OpenAI that allows users to create video from natural language text inputs. Expected public release: Aug 2024. Pricing not yet available.
Speechify	A text-to-speech tool that allows users to listen to documents, articles, PDFs, email, and more. Free (limited) and paid versions.
Suno	A music-generation tool that allows users to create songs in any musical style from text inputs. Credit based pricing. Free version includes 50 credits per day (= approx. 5 generations)
WolframAlpha	A GenAI chatbot geared toward providing “expert-level” answers for queries in a wide range of fields including: arts & media, linguistics, food & nutrition, political geography, history, demographics and social statistics, economic data, institutions and organizations. Free and paid (pro) versions.

DATA & PRIVACY

UNIVERSITY OF WATERLOO

DATA PRIVACY AND SECURITY MATTERS

Before engaging with any AI/GenAI, please consult IST's Guidance on Artificial Intelligence Use.

- The only GenAI tool at UWaterloo that will not track or retain your data is **Copilot** (not to be confused with Copilot for Microsoft 365).
- To ensure you are protected, login to **Copilot** using your **UWaterloo credentials**.

WHERE DOES THE DATA GO & WHO SEES IT?

When using any AI or GenAI tool, it's important to **read the terms of use and privacy policy**. Pay close attention to information about the data that may be collected such as: device information, geolocation, IP address, user inputs (including prompts and uploads), AI/GenAI outputs, conversations with the system, etc.

- **Never** enter student or institutional information into a system that hasn't been approved for use by IST at UWaterloo.
- **Avoid** entering personal or sensitive information into any AI/GenAI system to protect your privacy and ensure data security. These tools may not guarantee the confidentiality of information inputted, outputted, deliberately shared, or inadvertently/unintentionally shared.

RISKS

- **Privacy of data:** Users may unknowingly provide sensitive institutional data or personally identifiable information (PII) in their queries, prompts, and inputs. Since GenAI is trained on text that is entered into it, the tool may disclose sensitive and/or personally identifiable information in an output.
- **Loss of intellectual property (IP):** Entering sensitive or proprietary information into a GenAI tool opens you up to the loss of confidentiality and IP (e.g., business information, research data, patented or copyrighted data, software code, or drug trial data) which can devastate your personal, research, or institutional reputation, revenue, and future growth.
- **Threat or "bad" actors** have been known to harvest sensitive information to impersonate individuals or spread false information. They can also steal IP data quickly and in bulk. Threat actors can overcome restrictions within GenAI tools to create malware for use in a targeted cyberattack and can inject malicious code into the dataset used to train a GenAI system. This could undermine the accuracy and quality of the generated data. It could also increase the potential for large-scale supply-chain

BIAS

UNIVERSITY OF WATERLOO

BIAS IN DATA

- AI systems rely on large sets of training data and their outputs reflect the quality of those data.
- As a result, outputs may be racist, sexist, ageist, ableist, homophobic, transphobic, antisemitic, Islamophobic, xenophobic, deceitful, derogatory, culturally insensitive, and/or hostile. Due to intersectionality, it is very difficult to effectively remove hateful speech from existing data sets.[6]
- Additionally, the lack of metrics to support testing, evaluation, validation, and verification coupled with existing social, historical, and cultural biases present concerns for data training that can result in wrong or inaccurate outputs.[7,8]

TYPES OF BIASES IN AI SYSTEMS[8,9,10]

- Machine/algorithmic bias
- Availability bias
- Representation bias
- Historical bias
- Sample/selection bias
- Group attribution bias
- Contextual bias
- Linguistic bias
- Anchoring bias
- Automation bias
- Confirmation bias
- Cultural bias

HALLUCINATIONS

UNIVERSITY OF WATERLOO

IS IT REAL OR FAKE?

- A "**hallucination**" created by GenAI "refers to mistakes in the generated text that are semantically or syntactically plausible but are in fact incorrect or nonsensical".[11]
- Hallucinations appear frequently when attributing information to a source. For example, AI might state the wrong author or the wrong year published. It has also made-up authors, references, and URLS that do not exist, and has conflated research.[12,13]
- **What Causes AI Hallucinations?**[14]
  - Outdated or low-quality training data;
  - Incorrectly classified or labeled data;
  - Factual errors, inconsistencies, or biases in the training data;
  - Insufficient programming to interpret information correctly;
  - Lack of context provided by the user;
  - Struggle to infer the intent of colloquialisms, slang expressions, or sarcasm.

RECOGNIZING THE ISSUE

GenAI tools have limitations and potential hallucination rates from 4.8 to 8.5 and as high as 27 percent.[15,16,17,18,19]

Hallucination Examples

ChatGPT Legal Quotes and Citations: Earlier this year [2023], ChatGPT (developed by OpenAI) famously generated fake legal quotes and citations that were used in an actual court case. Unfortunately, the lawyers who submitted this false information were fined $5,000.[20]

Google's Bard [now Gemini] Chatbot and the James Webb Space Telescope: Google's Bard [now Gemini] chatbot inadvertently provided inaccurate details about the James Webb Space Telescope, leading to confusion among users. It incorrectly claimed that the telescope had captured the world's first images of a planet outside our solar system.[21]

As AI advances, hallucinations will decrease; regardless, material generated by AI should be closely scrutinized, fact-checked and the sources validated.

ACADEMIC INTEGRITY

UNIVERSITY OF WATERLOO

MAINTAINING INTEGRITY IN ERA OF GENAI

For students and instructors who are wondering where the line should be drawn for using GenAI in the completion of assessments or their own scholarly work, the answer is simple. The material produced must be, and has always been required to be, your *own* work.

What does your "own" work look like in age of GenAI?

Ask yourself:

- Am I confident that I understand the topic from the work that I have put in?
- Can I show how I got to the final version of my work with drafts and sources?
- Can I speak to the information and arguments I have written?
- Does my writing reflect my voice, my experience, and my level of understanding?
- Did the technology I used for checking errors re-write any of my work?
- Is the information accurate and can I verify it?

VALIDATING STUDENTS' WORK & KNOWLEDGE

In the GenAI era, instructors may need to rethink traditional written assignments as well as the ways knowledge, comprehension, and skills are assessed. **No assignment can ever be cheat proof or GenAI proof**. To reduce students' temptation to use GenAI in unauthorized ways instructors might:

- discuss the importance of academic integrity and credibility with students;
- be explicit about expectations, what is/isn't allowed, and deadlines;
- have a generative AI use policy, include it in the syllabus/course outline, discuss it with students, and draw students' attention to it at multiple points during the term (see Course Outline Suggestions for GenAI);
- point out the application of content, skills, and learning outcomes to the program and to a future career(s);
- ask students to submit drafts, outlines, details, and rough work that demonstrates how they proceeded to complete their assignments, and give them credit/grades for doing so;
- recommend students meet with Writing and Communication Centre advisors for support early in their assignments or with the Student Success Office coaches for time management and other learning strategies.

When in doubt – ask the student to speak to their work, replicate it, or answer questions about it to validate their knowledge and understanding.

AI DETECTION

UWaterloo subscribes to Turnitin for both text matching and GenAI detection. The product excels at identifying text matches; however, **it is not reliable at detecting GenAI produced content**.

A high score for AI generated text does not automatically equate to academic misconduct. A high score is only *one* potential piece of evidence of misconduct. UWaterloo encourages an educative approach – a high GenAI detection score is an opportunity to discuss and explore ethical GenAI use with a student.

Be aware that tools like Grammarly now have some GenAI-supported features, the outputs of which can be flagged by Turnitin. Students may not understand how Grammarly has been augmented by GenAI, how GenAI tool use impacts their work, or how to use GenAI tools productively, responsibly, and with integrity.

Be clear about appropriate and inappropriate use of GenAI in your course and assignments by having a **GenAI use policy** in your course outline and **discussing it** with your students.

COPYRIGHT

UNIVERSITY OF WATERLOO

COPYRIGHT CONCERNS WITH GENAI – SOURCE MATERIAL

Dubious Sources of Training Material

- Content used to train GenAI models is either unknown (i.e., the content source cannot be verified) or known to be copyright infringing. This means it is difficult to make a fair dealing argument for use of the materials GenAI tools output – you can't make legal use of an infringing work.

Absence of Attribution

- Many GenAI services can't attribute the sources of the underlying works used to generate their outputs. This means that if a particular output relied on a certain source in a substantial way and that source was protected by copyright, not citing that source would infringe on the rights of its creator.

Uncertainty Surrounding Outputs

- Given the way GenAI tools provide outputs to users, there is no way for a user to know whether a GenAI output based on their prompt(s) constitutes a substantial use of another creator's work.

COPYRIGHT CONCERNS WITH GENAI – LEGAL LANDSCAPE

- Copyright ownership of AI-generated content is unclear in Canada (and elsewhere).
- The rights of creators whose copyright material was used to train AI systems, and thus used in AI-supported content generation, remains unsettled.
- Because AI-generated content may not be protected by copyright, its creators (i.e. GenAI users) assume risks if they benefit from it financially or professionally
- There is no AI/GenAI copyright legislation or case law in Canada. The lack of developments in these areas means that there is considerable legal uncertainty for users and developers of AI/GenAI.

UWaterloo recommends exercising caution, by making principled, risk-informed use of GenAI tools and their outputs.

HUMAN LABOUR

EXPLOITATIVE EXTRACTIVE LABOUR

As is the case with "fast fashion" and “fast food” systems, tech companies often exploit workers (especially in the Global South) to reduce the costs of developing AI.

- Shannon Tse (2023, Nov 27). Technological Disparities: Exploiting the Global South and AI’s Hidden Human Labour. Catalyst. (accessed June 7, 2024)
- Niamh Rowe (2023, Oct 16). Millions of Workers are Training AI Models for Pennies. Wired. (accessed June 7, 2024)
- Billy Perrigo (2023, Jan 18). OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic. Time. (accessed June 7, 2024)
- Karen Hao & Andrea Paola Hernández (2022, Apr 20). Part 2: How the AI industry profits from catastrophe. AI Colonialism series, MIT Technology Review. (accessed June 7, 2024)
- Aitor Jiménez González (2021). Law, Code and Exploitation: How Corporations Regulate the Working Conditions of the Digital Proletariat. Critical Sociology. 48(2). (accessed June 7, 2024)
- Moritz Altenried (2021). The platform as factory: Crowdwork and the hidden labour behind artificial intelligence. Capital and Class, 44(2). (accessed June 7, 2024)

FREE LABOUR

Be aware of privacy and licensing agreements when you create an account, sign in to, or use any GenAI tool. In many cases, agreeing to the terms of service results in free labour for the company in question, meaning the company can use your inputs to improve, train, and/or enhance their product(s).

- Melissa Heikkilä (2023, Jun 13). We are all AI's free data workers. MIT Technology Review. (accessed June 7, 2024)
- Fabio Morreale, Elham Bahmanteymouri, Brent Burmester, Andrew Chen, and Michelle Thorp (2023, May 17). The unwitting labourer: extracting humanness in AI training. AI & Society. (accessed June 7, 2024)

THE REAL COST

"So-called AI systems are fueled by millions of underpaid workers around the world, performing repetitive tasks under precarious labor conditions."[22]

- Content moderators, for example, are responsible for finding and flagging content deemed inappropriate for a given platform.
- Every murder, suicide, sexual assault or child abuse video that does not make it onto a platform has been viewed and flagged by a content moderator or an automated system trained by data most likely supplied by a content moderator.
- Employees performing these tasks suffer from anxiety, depression and post-traumatic stress disorder due to constant exposure to horrific content.

ENVIRONMENTAL IMPACT

UNIVERSITY OF WATERLOO

"While AI offers the promise of transformational leaps in efficiency, productivity, and innovation, it is not without **significant costs, including to the environment.** The world is in the midst of a climate emergency, making it imperative that we **temper our excitement around AI with a sober reflection on its impact on climate change**."[23]

Dr. Mary Wells - Dean of Engineering, UWaterloo

ENERGY AND WATER USE

Generative AI applications use enormous amounts of energy and water compared to traditional search engines and computer applications.[24, 25, 26, 27] It's easy for this to be invisible to users, but there have been attempts to quantify. For example:

- a 20-25 question interaction with ChatGPT requires about 500mL of clean drinking water to cool its data centres,[28]
- the generation of 1000 images is equal to driving 4 miles (6.5km) in a gasoline-powered car,[29]
- researchers have estimated that creating GPT-3 consumed 1,287 megawatt hours of electricity and generated 552 tons of carbon dioxide equivalent, the equivalent of 123 gasoline-powered passenger vehicles driven for one year,[30]
- e-waste is also a concern as AI developers churn through hardware upgrades in pursuit of larger and faster models.[31]

WHAT TO DO

- Consider reusing sample text or images rather than performing frequent live demonstrations of a tool
- Be mindful of how many AI apps you are using; consider using fewer
- Watch for "thinner" apps that use fewer bits of information for the task at hand (when appropriate)
  - e.g., Gemini Flash, Microsoft's Phi3-mini, Meta's LLaMa 3 and other Small Language Models (SLMs)
- At a minimum, ask for and generate smaller amounts of text, image, and other GenAI outputs

SPREADING MISINFORMATION

UNIVERSITY OF WATERLOO

MISINFORMATION SPREAD AND CONCERNS

Hundreds of fake news sites have already been created with generative AI. They cover topics including politics, technology, entertainment, and travel.[32, 33]

Voice and image generation technology is also being used to generate inaccurate TikToks, Instagram Shorts, and other video content.[32, 33, 34]

In addition to personal, political, and social implications, there are also widespread concerns about the economic impacts of global misinformation spread. These include concerns about:

- damage to property by riots;
- damage to production through strikes;
- damage to trade through limitations imposed on movement of people and goods by authoritarian governments;
- threats to democracy. [35, 36]

POLICY & REGULATION

- Generating false or misleading content is against the user agreements of GenAI tools. See, for example:
  - Gemini's Generative AI Prohibited Use Policy
  - ChatGPT's parent company OpenAI's Usage Policies
  - Microsoft's Copilot Code of Conduct within their Services Agreement
- But the **use of generative AI tools is unregulated**. This places the responsibility for evaluating and discerning AI-generated content, misinformation, and disinformation on the individual. This set of abilities is part of **AI Literacy**.
- **AI literacy** is considered to be "**one of the most important skills** we learn in the 21st century." [37]

DIGITAL DIVIDE

UNIVERSITY OF WATERLOO

TYPES OF DIGITAL DIVIDES

"There's a major **gap between people who can access and use digital technology and those who can't**. This is called the **digital divide**, and it's getting worse as 3.7 billion people across the globe remain unconnected."[38]

Access Divide refers to the difference between those who have access to technology and those who do not.

- For example, students who can afford upgraded versions of GenAI tools (e.g., ChatGPT Plus) will have access to better features and functionality than those who cannot.

Usage Divide refers to the difference between those who know how to use technology and those who do not.

- For example, students with family members or teachers who can show them how to use laptops, phones, applications, and services to access GenAI tools for thinking, communication, and learning will be at more of an advantage than those who do not.

GENAI PERPETUATES INEQUITY

- Students from wealthy or privileged backgrounds are more likely than others to use GenAI [39]
- Higher performing students are more likely to use GenAI than lower performing students [40]
- Male students use GenAI tools for data analysis and coding, female students for edits, translation, or transcription [41]
- GenAI is linguistically English and culturally American [42]

SUMMARY

"With great power comes great responsibility." [43]

Rapid technological changes will continue to bring new challenges and opportunities.

With new advances happening daily, it is imperative to:

- keep learning about and exploring GenAI tools;
- consider how GenAI tools can help or harm at personal, profession, and societal levels;
- discuss the merits and demerits of GenAI and of specific GenAI tools with colleagues and students;
- remain focused on the purpose of education and preparing students as citizens of the future.

CAMPUS RESOURCES

Centre for Teaching Excelence (CTE)

Website: CTE

Email: cte@uwaterloo.ca

Phone: 519-888-4567, ext. 43353

GenAI resource: Conversations with Students about GenAI Tools

Office of Academic integrity (OAI)

Website: OAI

Email: academic.integrity@uwaterloo.ca

GenAI Resource: Artificial intelligence and ChatGPT

Centre for Extended Learning (CEL)

Website: CEL

Email: extendedlearning@uwaterloo.ca

Phone: 519-888-4567, ext 44050

GenAI Support: Agile Development - AI Help Portal

Copyright

Website: Copyright

Email: copyright@uwaterloo.ca

GenAI resource: Copyright & Generative AI

Writing and Communication Centre (WCC)

Website: WCC

Email: wcc@uwaterloo.ca

GenAI resource: GenAI and the Writing Process Resources

Information Systems and Technology (IST)

Website: IST

Email: helpdesk@uwaterloo.ca

Help & Support: start here

GenAI resource: Guidance on Artificial Intelligence Use

UWaterloo Library

Website: Library (faculty, instructors, postdoc, and visiting scholars page)

Email: libaskus@uwaterloo.ca

GenAI resource: ChatGPT and Generative Artificial Intelligence Research Guide

REFERENCES

1. Pavlik, G. (2023, Sept 15). What is Generative AI? How Does It Work? Oracle. Accessed April 29, 2024.
2. Bastian, M. (2023, Jan 25). Meta's AI Chief Yann LeCun calls ChatGPT "not particularly innovative". The Decoder. Accessed April 29, 2024.
3. Legoux, G. (2023. Apr 16). History of Generative AI. Medium. Accessed April 29, 2024.
4. Wolfram, S. (2023, Feb 14).). What Is ChatGPT Doing … and Why Does it Work? Stephen Wolfram Writings. Accessed April 29, 2024.
5. Government of Canada (2023, July). Generative Artificial Intelligence (AI) – ITSAP.00.041. Awareness Series. Accessed April 29, 2024.
6. Zehui, Y.. et al. (2024). The Unseen Targets of Hate—A Systematic Review of Hateful Communications Datasets. ArXiv. Accessed April 29, 2024.
7. Kleiman, G. (2023, Jan 5). Teaching Students to Write with AI: The SPACE Framework. Medium. Accessed April 29, 2024
8. Schwartz, R. (2022). Towards a Standard for Identifying and Managing Bias in Artificial Intelligence. National Institute of Standards and Technology (NIST), U.S. Department of Commerce. Accessed April 29, 2024
9. Scheuer-Larsen, C. (2023). A nuanced view of bias in language models. Viden.AI. Accessed April 29, 2024
10. van Niekerk, D. et al. (2024). Challenging systematic prejudices: An investigation into bias against women and girls in large language models. UNESCO. Accessed April 29, 2024
11. Smith, C. (2023, Mar 13). Hallucinations Could Blunt ChatGPT's Success. IEEE Spectrum. Accessed April 24, 2024.
12. Wu, K., Wu, E., Ho, D. E., and Zou, J. (2023, Feb 12). Generating Medical Errors: GenAI and Erroneous Medical References. News - Stanford University (stanford.com). Accessed May 7, 2024.

REFERENCES CONT'D

13. University Libraries and Teaching and Learning Transformation Centre (n.d.). What does AI get wrong? - Artificial Intelligence (AI) and Information Literacy. Research Guides at University of Maryland Libraries (umd.edu). Accessed May 7, 2024.

14. Keary, T. (2024, Jan 15). What Is AI Hallucination? Examples, Causes & How To Spot Them. Techopedia. Accessed April 24, 2024.

15. OpenAI (n.d.) GPT-4. OpenAI (openai.com). Accessed April 24, 2024.

16. Diaz, N. (2024, Feb 23). Google explains what went wrong with Gemini's AI image generation. AndroidCentral. Accessed April 24, 2024.

17. Rees, K. (2023, Dec 24). What Is AI Hallucination? Can ChatGPT Hallucinate? How-To Geek. Accessed April 24, 2024..

18. Metz, C. (2023, Nov 6). Chatbots May 'Hallucinate' More Often Than Many Realize. The New York Times (nytimes.com). Accessed May 7, 2024.

19. Williams, A. Milagros, M., and Gebru, T. (2022, Oct 13). The Exploited Labor Behind Artificial Intelligence. NOĒMA. Accessed May 8, 2024.

20. Dooley, J. (2023, Dec 21). AI Hallucinations: Does Your GenAI Need a Reality Check? Coveo. Accessed May 7, 2024.

21. Quthor (2024, Apr 26). The Impact of Generative Artificial Intelligence on Hallucinations: A Data Analysis. QuickCreator. Accessed May 7, 2024.

22. Williams, A. Milagros, M., and Gebru, T. (2022, Oct 13). The Exploited Labor Behind Artificial Intelligence. NOĒMA. Accessed May 8, 2024.

23. Wells, M. and Mehlenbacher, A. R. (2024, May 8). We must temper hype around AI with sober reflection on its climate impacts. The Hill Times. Accessed May 9, 2024.

REFERENCES CONT'D

24. Clarke Gray, B. (2024, Feb 14). Toxic Tech #5: Generative AI: Sustainability, Equity, and Living our Values (or not). Digital Detox. Accessed June 12, 2024.

25. Saenko, K. (2023, May 23). Is generative AI bad for the environment? A computer scientist explains the carbon footprint of ChatGPT and its cousins. The Conversation. Accessed June 12, 2024.

26. Luccioni, S. (2023, Mar 12). The mounting human and environmental costs of generative AI. Ars Technica. Accessed June 12, 2024.

27. Leffer, L. (2023, Oct 13). The AI Boom Could Use a Shocking Amount of Electricity. Scientific American. Accessed June 12, 2024.

28. Li, P., Yang, J., Islan, M. A., and Ren, S. (2023, v3 Oct 23). Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models. arXiv. Accessed June 12, 2024.

29. Clarke Gray, B. (2024, Feb 14). Toxic Tech #5: Generative AI: Sustainability, Equity, and Living our Values (or not). Digital Detox. Accessed June 12, 2024.

30. Saenko, K. (2023, May 23). Is generative AI bad for the environment? A computer scientist explains the carbon footprint of ChatGPT and its cousins. The Conversation. Accessed June 12, 2024.

31. Craig, L. (2023, Sept 6). Designing systems that reduce the environmental impact of AI. TechTarget. Accessed June 12, 2024.

32. Sadeghi, M. et al (2024). Overview: Tracking AI Enabled Misinformation. NewsGuard. Accessed April 29, 2024..

33. Satariano, A. and Mozur, P. (2023, Feb 7). The People Onscreen Are Fake. The Disinformation Is Real. The New York Times (nytimes.com). Accessed June 12, 2024.

34. Safi, M., Atack, A. and Kelly, J. (2024, Feb 29). Revealed: the names linked to ClothOff, the deepfake pornography app. The Guardian. Accessed June 12, 2024. AI Driven Misinformation is "Biggest Short-term Threat to Global Economy". The Guardian (theguardian.com). Accessed April 29, 2024.

REFERENCES CONT'D

36. Filibeli, T. (2019). Big Data, Artificial Intelligence, and Machine Learning Algorithms: A Descriptive Analysis of the Digital Threats in the Post-truth Era. Galatasaray University Journal of Communication, 31, pp. 91-110. Accessed April 29, 2024.

37. LaChance, A. M. (2024, Feb 18). Accurate Reporting? That'll be $5/Month, Please. On AI Slop, the Death of Journalism, and Class. That Anna Marie Newsletter. Accessed April 29, 2024.

38. Connecting the Unconnected (n.d.). What's the Digital Divide?. IEEE.org. Accessed May 1, 2024.

39. MacGregor, K. (2024, Feb 2). Most Students use AI for Studies, Digital Divide Emerges – Survey. University World News. Accessed May 1, 2024.

40. Schiel, J., Bobek, B. L. and Schnieders, J. Z. (2023, Dec). High School Students' Use and Impressions of AI Tools. ACT Research. Accessed May 1, 2024.

41. MacGregor, K. (2024, Feb 2). Most Students use AI for Studies, Digital Divide Emerges – Survey. University World News. Accessed May 1, 2024.

42. Ta, R. and Lee, N. T. (2023, Oct 24). How Language Gaps Constrain Generative AI Development. The Brookings Institution (brookings.edu). Accessed May 1, 2024.

43. Gutoskey, E. (2024, Mar 29). Who First Said "With Great Power Comes Great Responsibility"? Mental Floss (mentalfloss.com), Accessed May 8, 2024

Canada’s Innovation University