



**WORKSHOP 1A
AI and the Law:
Navigating Compliance and
Responsibility**

**TLOMA 2025 Conference and
Trade Show**

**18 September 2025
Niagara-on-the-Lake, ON**

Maura R. Grossman, JD, PhD

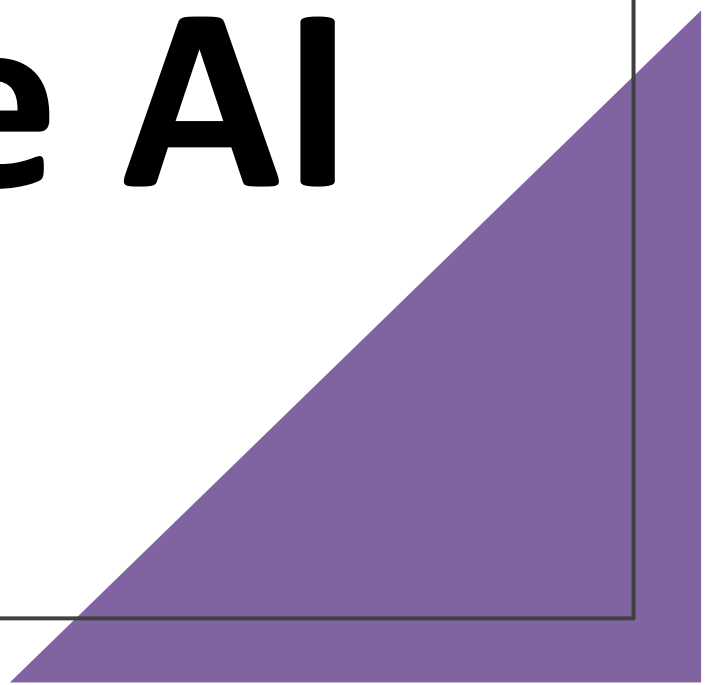
Agenda for Today

- **A Quick Tutorial on Generative AI: What It Is, How It Works, What It Can and Can't Do**
- **Examples of Gen AI in Action**
- **Applications and Risks of Gen AI**
- **Challenges in Detecting Gen AI Content**
- **Ethical and Professional Responsibility Issues Implicated by Gen AI**
- **Gen AI Policies, Procedures, and Training**
- **[Bonus Content: Best Practices in Prompt Engineering]**
- **Q&A**





A Quick Tutorial on Generative AI



WHAT IS “GENERATIVE AI” (“GEN AI”)?



- A subset of AI that uses training on **massive** data sources – primarily from the **Internet and large proprietary databases** – to generate **new content** in response to a **user prompt**. It can converse, replicate specific styles, and excels at creative tasks and synthesizing or summarizing content.
- Gen AI falls under the broad categories of **machine learning** and **natural language processing**.
- It leverages neural networks (*i.e.*, **deep learning**) to analyze the underlying patterns and structures of data, enabling it to **predict what should come next**, or to **generate fresh and unique content**. (This explains why it **doesn't give the same answer twice** and “**hallucinates!**”)

BRIEF HISTORY OF GEN AI: PRECURSORS TO LARGE LANGUAGE MODELS (“LLMS”)



- New or Not?
 - **Claude Shannon** (one of the founders of AI) wrote *Prediction and Entropy of Printed English* in 1951.
 - He would erase fragments of text and have humans guess what was erased (**Mad Libs!**).
 - **Language models are statistical models applied to Shannon’s prediction task.** Historically used for things like data compression, information retrieval, author and spam detection.
 - In comes massive computing power + massive data + neural networks, convolutional neural nets, deep learning . . .

A BRIEF HISTORY OF GEN AI: 2010 – 2022



- In 2014, **Generative Adversarial Networks (“GANs”)** took a huge leap forward in their ability to create authentic-looking content.
- GANs introduce a new way for algorithms to learn: One algorithm (the **“generator”**) **creates content**, and the other algorithm (the **“discriminator”**) **evaluates it against real data in an effort to distinguish them**. This approach creates more and more realistic-looking content (and also explains why detection of Gen AI content is so difficult).
- GANs revolutionized image, audio, and video generation.

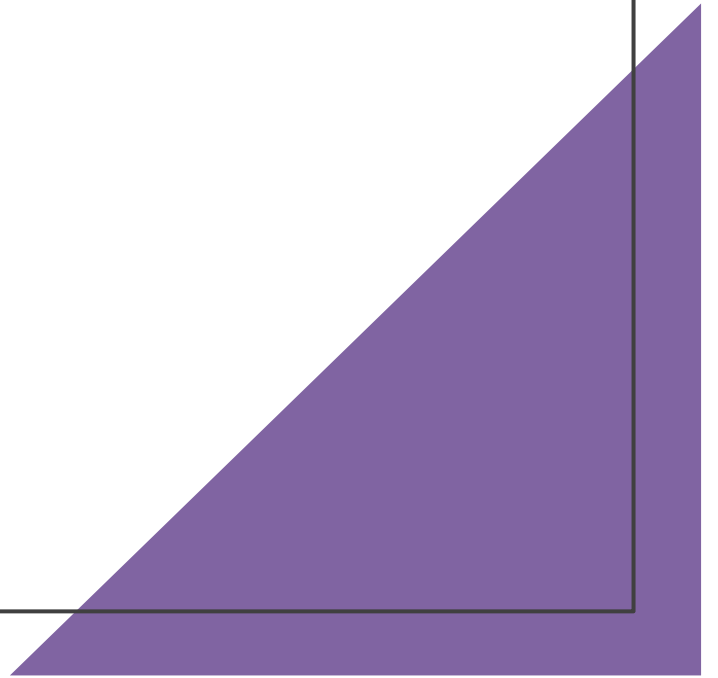
A BRIEF HISTORY OF GEN AI: 2010 – 2022

(CONT'D)



- In 2015, Jascha Sohl-Dickstein introduced **diffusion models**, which used advanced ML to generate extremely high-quality images by **progressively introducing noise to the dataset and then reversing the process** (*i.e.*, gradually degrading data quality only to reconstruct it into something new and better).
- In 2017, Google researchers introduced the **transformer architecture**, a significant breakthrough in processing natural language which **no longer required pre-labelled training data** and **allowed processing to occur in parallel** (which is much faster).
- Another major change introduced with GPT-3 was the use of **reinforcement learning**, in which external (*i.e.*, human) **feedback is used to modify and improve the output of the model**.

Examples of Gen AI in Action



GEN AI IN ACTION: TEXT-TO-TEXT (#1A)



- **MRG's Prompt:**

- **“Please write a short Shakespearean sonnet about the Ontario judiciary”**

- **ChatGPT's Response:**

Upon Ontario's bench, the judges sit,
Dispensing justice with a learned grace,
In robes of black, they ponder and commit,
To navigate the law's intricate space.

With wisdom vast they weigh each legal claim,
Their gavels sound, decisions fair and just,
In courtrooms' hallowed halls, they play the game,
Defending rights, in them, we place our trust.

From Kingston's shores to Ottawa's great height,
Through Toronto's bustling streets, they hold their sway,
The judiciary's role, a beacon's light,
Guiding the law through each and every day

Oh, Ontario's judges, noble and true,
In justice's name, we salute all you do.

GEN AI IN ACTION: TEXT-TO-TEXT (#1B)



- **MRG's Prompt:**

- **“Please write a short Eminem-style rap about the Ontario judiciary”**

- **ChatGPT's Response:**

Yo, let me tell you 'bout the Ontario crew,
Judiciary game strong, yeah, they stay true
In the courtroom battles, they wear the robe,
Dispensing justice, man, they're in control.

From Toronto to Ottawa, they hold it down,
Keeping order in the legal battleground,
Ruling with wisdom, in that judge's chair,
They make decisions, no one can compare.

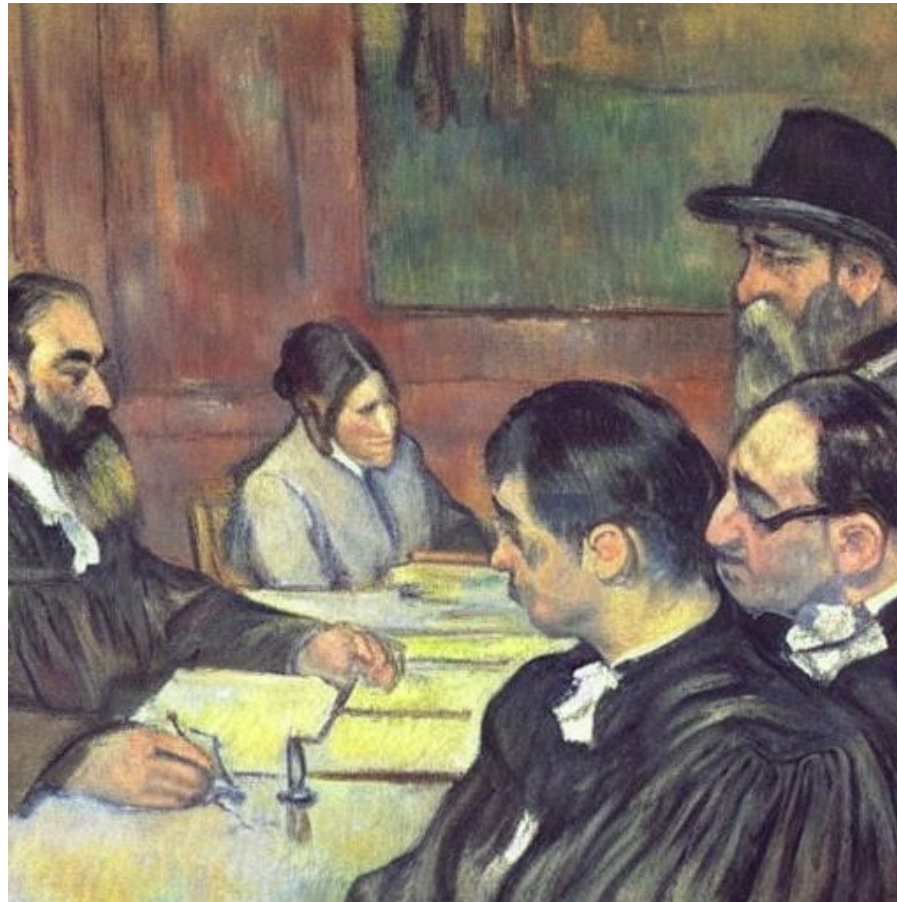
Legal minds sharp, like a razor blade,
Analyzing cases, every move is weighed,
No bias, just facts, they play it clean,
In Ontario's courtrooms, it's a righteous scene.

So shout out to the judges, respect's due,
In the land of Ontario, they're the crew,
Making sure justice always prevails
In the province's heart, their legacy sails.

GEN AI IN ACTION: TEXT-TO-IMAGE (#2A)

- **MRG's Prompt:**
 - “Picture of the Ontario judiciary in the style of Degas”

- **DeepAI's Response (based on Stable Diffusion):**



GEN AI IN ACTION: TEXT-TO-IMAGE (#2B)

- **MRG's Prompt:**
 - **"Picture of the Ontario judiciary in the style of Picasso"**

- **DeepAI's Response (based on Stable Diffusion):**



GEN AI IN ACTION: TEXT-TO-IMAGE (#3A & #3B)

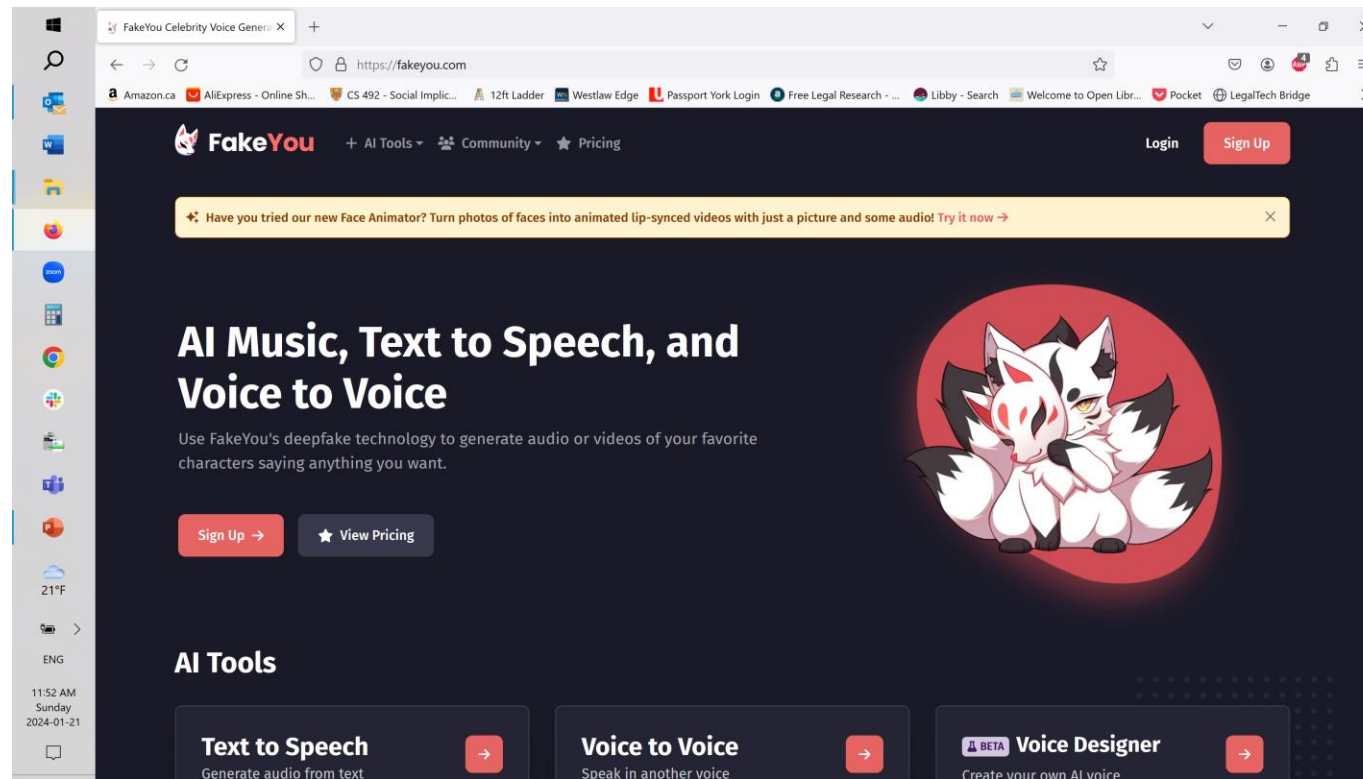
- MRG's Prompt:
 - "Picture of [more] diverse Ontario judges"

- DeepAI's Response (based on Stable Diffusion):



GEN AI IN ACTION: TEXT-TO-SPEECH (“TTS”) (#4)

- **MRG’s Prompt:**
 - **President Obama, can you say hello to today’s webinar attendees?**
- **FakeYou’s Response:**



GEN AI IN ACTION: TTS (#5A)

Can we make a decent angry voice clone of Maura Grossman for free using Speechify?

- **Input Recording:** 

GEN AI IN ACTION: TTS (#5B)

Can we make a decent angry voice clone of Maura Grossman for free using Speechify (cont'd)?

- **Output Recording:** 

GEN AI IN ACTION: IMAGE-TO-VIDEO (#6)

Replicate

Explore Pricing Docs Blog Changelog Sign in **Get started**

Want to run this model with an API? **Get started**

yoyo-nb / thin-plate-spline-motion-model

PUBLIC Thin-Plate Spline Motion Model for Image Animation **41.6K runs** **GitHub** **Paper** **License**

Demo **API** **Examples** **Versions (382ceb8)**

Input

source_image

Drop a file or click to select
<https://replicate.delivery/mgxm/b8efcd01-da81-42c7-8cd6-0a820084a983/source.png>

Take a photo with your webcam

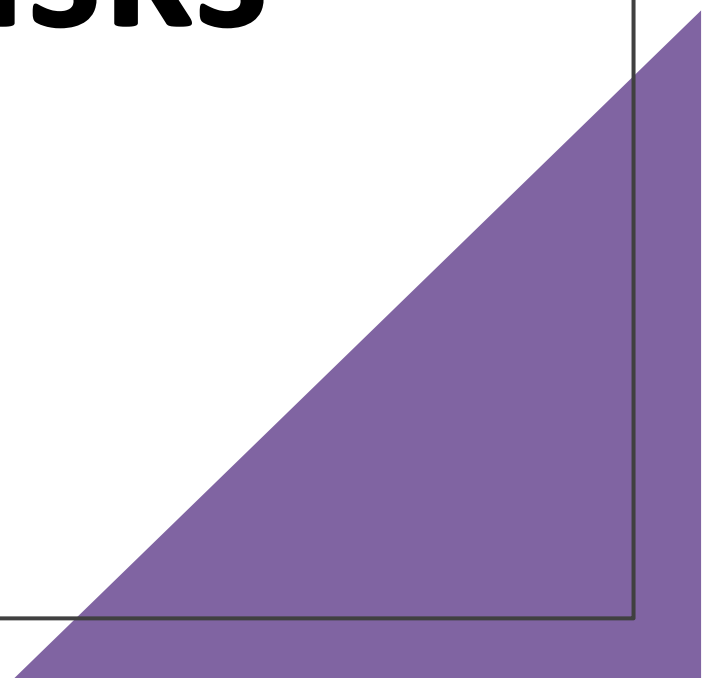
Input source image.

Output

Video frame showing a woman's face.

<https://grossman.uwaterloo.ca/deepfakes.mp4>

Applications and Risks of Gen AI in Legal



APPLICATIONS OF GEN AI IN LEGAL

- **Gen AI will:**
 - **Enhance delivery of legal services** by providing lawyers with tools to increase their productivity.
 - **Enhance access to justice** by providing tools to litigants unable to afford legal services or navigate the legal system.
- **Gen AI will not:**
 - **Replace a lawyer's or judge's reasoning, critical thinking, compassion, empathy, etc.**
- **Gen AI can:**
 - **Analyze, translate, and summarize lengthy documents, e.g., complex statutes or regulatory codes; witness transcripts to identify key people, events, or inconsistencies**
 - **Brainstorm ideas or (counter)arguments**
 - **Help with marketing and creative copy**
 - **Create outlines and draft or edit documents and presentations**
 - **Conduct research???**
 - **Perform electronic discovery???**
 - **Respond to emails???**

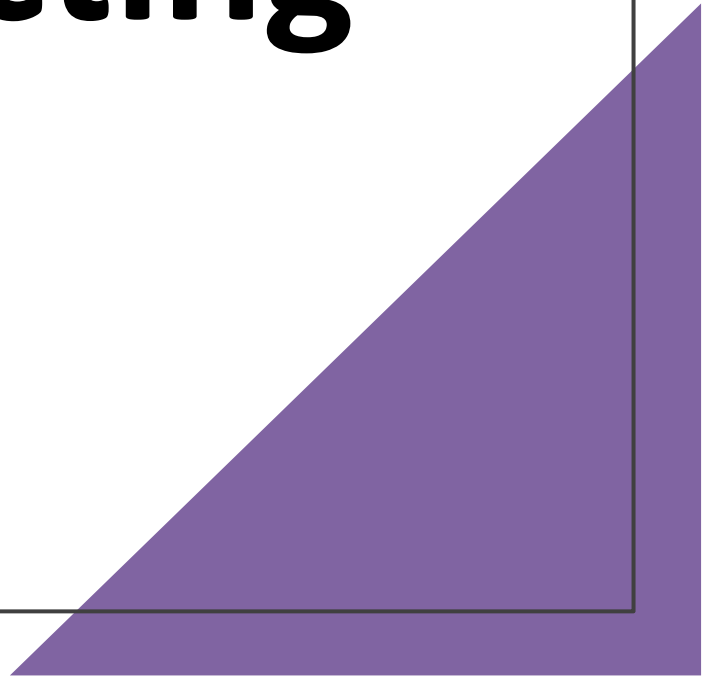


RISKS OF GEN AI IN LEGAL

- Gen AI **does not respect confidentiality or privacy**; anything you enter may be used for training or other purposes unless you contract otherwise
- Gen AI **does not guarantee the accuracy of its output**
 - It sounds very confident and compelling
 - But, . . . **it hallucinates**
 - **It reinforces stereotypes** (as you saw!)
 - It is predicting things based on Internet content; your mileage may vary
 - It can be biased, toxic, and defamatory
- Gen AI **is not secure and is subject to jailbreaking and other adversarial attacks** (e.g., prompt injections)
- Gen AI **content is likely not subject to copyright protection and may infringe on others' IP**



Challenges in Detecting Gen AI Content





DISTINGUISHING AI FROM HUMAN-GENERATED TEXT

- Tools look for the use of **formal, common, predictable language**, versus the use of **surprising or emotional language**.
- Detectors **often mistake text written by people whose native language is not English** as written by an LLM.
- Detectors have very **high false-positive rates**.
- OpenAI itself says:
 - “Do AI detectors work? In short, no. While some (including [us]) have released tools that purport to reliably distinguish between AI-generated content, **none of these have proved to reliably distinguish between AI-generated and human-generated content.**”

WHAT ABOUT “WATERMARKING”?



- Tools that embed a **stealth signal in the text** to signify that it was generated by an LLM rather than a human.
- Typically, it is a **hidden pattern that is inserted at random** into the textual content. It is **imperceptible to the human eye, but readily detectable by an algorithm**.
- To work, the method has to be **widely accessible but permanently maintained as a secret**.
- Must be **highly robust to modification or removal**; currently, skilled attackers can remove watermarks.
- **All LLMs would have to use them from ground zero** because the second you have a single LLM that does not, you can take the output of a watermarking LLM and ask the non-watermarking LLM to paraphrase it (or a human).

What makes deepfakes different from other forms of past forgery?

- Basically, it boils down to scope and scale
 - Cheap and easy to make; no talent required
 - All you need is a device and access to the Internet → democratizes fraud
 - Ubiquitous and growing exponentially
- **But there's more . . .**

Spotting deepfakes has become a lot harder!

- Early tell-tale signs:
 - May not blink properly
 - Lip synching may be slightly off
 - Skin tone may be patchy
 - Flickering around the edges of transposed faces
 - Fine details such as hair strands, jewelry, teeth, ears, and fingers are off
 - Strange lighting or shadow effects *e.g.*, inconsistent illumination and reflections on the iris
- **Has gotten much much harder as the technology has improved and the distinctions are subtle and often exceed human perception; experts are typically needed**

Approaches to detection: human and automated

- Inference-Based

- Looks for **signals in the actual media content** that do not line up or match properly with what one would expect in the real world, such as fabric textures, reflections on the cornea, lines on the floor not extending backwards in the proper perspective, mispronunciations of words, irregular breathing patterns, peculiar background noises, etc. [**Think: Princess Kate photo**]

- Provenance-Based

- Involves the review of **metadata, such as time stamps, GPS coordinates, or pixels** to look for signals suggesting the use of AI in generating or manipulating the media that are not content-based per se.

- **Both involve a search for inconsistencies or signals suggesting departure from what one would expect in the real world, but neither approach is certain or foolproof.**

They merely allow an expert to opine with some degree of confidence that the media appears to be consistent or inconsistent with authentic media, or that the metadata or pixels (or other digital artifacts) suggest that the media was created or altered using AI.

Challenges with human detection of deepfakes

- **Humans are simply not very good at making judgements about authenticity, regardless of whether the media is visual or auditory**
- **In terms of images**, multiple studies show that **humans do not perform much better than chance and are overconfident in their deepfake-detection abilities**
 - In a 2024 study of 244 US participants, more than half could not reliably distinguish between AI-generated and genuine content;
 - In a 2024 study of 3,002 US, German, and Chinese subjects, most could not distinguish AI-generated media from real media; AI-generated media was rated as more likely to be human-generated for all media types (text, image, and audio) in all three countries
 - **Average detection accuracy was below 50% for images and never exceeded 60% for text or audio**

Challenges with human detection (cont'd)

- Fewer studies have looked at speech detection, but the available studies suggest that **humans do not fare all that much better with audio media**
 - **At best, humans made correct classifications about 70-85% of the time depending on how many times they could listen to the audio before rendering a decision**
 - Familiarizing listeners with examples of deepfakes in advance of detection exercises does appear to boost detection ability by a small degree (~3.84%)

Challenges with automated detection tools

- Developer A can develop a discriminator that works well in detecting the use of Generator A (developed by Developer A) but the same tool will not work well in distinguishing the use of Generator B (developed by Developer B)
 - **Open AI's image-detection tool can correctly flag 98% of the images generated by its own DALL·E image generator, but it only flagged 5-10% of images generated by competitors' image generators**
 - **ElevenLabs' AI audio detector reported a high (>90%) detection rate for its own synthetic audio, but failed to report how well it detected synthetic audio generated by other audio generators**

The two-fold risk of deepfakes on the legal system

- **Undermining of public trust** → leading to a zero-trust society where **people either cannot or no longer bother to distinguish truth from falsehood**
- **Makes it easier to raise doubts about real events** → the “liar’s dividend,” where **actual reality becomes plausibly deniable**

Some scary statistics

- In 2023, there was a 10-fold increase in deepfakes detected globally across all industries: in North America alone, there was an 18-fold increase on deepfake fraud cases. In 2024, 49% of companies in the US, UAE, Mexico, Singapore, and Germany reported encounters with deepfake scam attempts. 60% of Americans have expressed a significant concern about deepfakes—more than any other AI-related risk.
- In a recent Deloitte poll, nearly 26% of polled executives reported that their organization had experienced at least one or two deepfake attacks in the previous 12 months and 51.6% expected the number and size of such deepfake fraud attacks to increase over the next 12 months. Deloitte's Centre for Financial Services estimated that AI-generated fraud will reach \$40 Billion in damages in the US by 2027.
- **Finance, banking, law, and healthcare are key targets because of their ready access to either huge amounts of money or data**

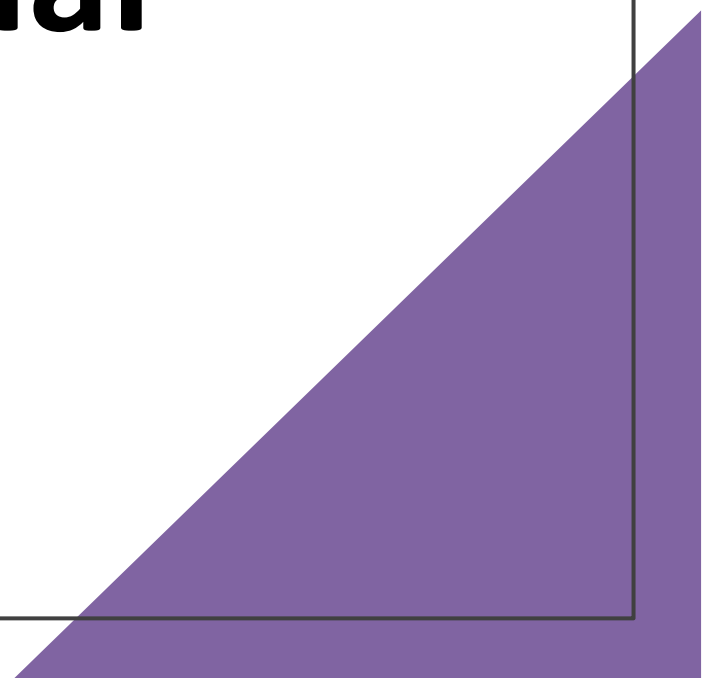
Steps organizations can take

- **Assess your organization's particular vulnerabilities and risk profile**
- **Enhance anti-fraud training and a skeptical culture**
 - Train employees to identify newly emerging fraud types; you cannot get complacent in this environment
 - Be particularly suspicious of **urgent or non-standard requests**; always verify these using an **independent communication method**
 - Encourage employees to trust their instincts, maintain a skeptical mindset, and **always seek backup/confirmation in any suspicious situations**
 - **Test how employees respond to mock fraud attempts** involving

Steps organizations can take (cont'd)

- **Reassess policies and practices for large monetary and data transfers**
 - Examine internal controls over who can authorize/approve or implement major financial or data transfers and how
 - Set policies for communication including the use of codewords or other methods for validating authenticity
- **Use 2FA whenever possible**
- **Embrace but do not overly rely on AI-detection technologies**

Ethical and Professional Responsibility Issues



APPLICABLE CANADIAN RULES OF PROFESSIONAL CONDUCT



- **Rule 3.1-2 – Duty of Competence** (understand the **limitations of the tech** rather than tech itself)
- **Rule 3.2-2 – Duty of Honesty and Candour to Clients**
- **Rule 3.2-7 & 2.1 – Duty to Comply with the Law** (includes AI-specific, privacy, cybersecurity, IP, etc.)
- **Rule 3.3-1 – Duty of Confidentiality**
- **Rule 3.2-1 & 3.2-2 – Duty to Communicate with Clients**
- **Rule 3.6-1 – Duty to Charge Reasonable Fees and Disbursements**
- **Rule 5.1-1 – Duty Not to Mislead the Tribunal**
- **Rules 6.1-1 & 6.2-2 – Duty to Supervise Subordinates and Students**
- **Rule 6.3.1-1 – Duty of Non-Discrimination** (*i.e.*, to guard against discrimination, harassment, and bias)
- **Rule 7.6-1 – Duty to Prevent Unauthorized Practice**

AMERICAN BAR ASSOCIATION FORMAL OPINION 512 (JULY 29, 2024)



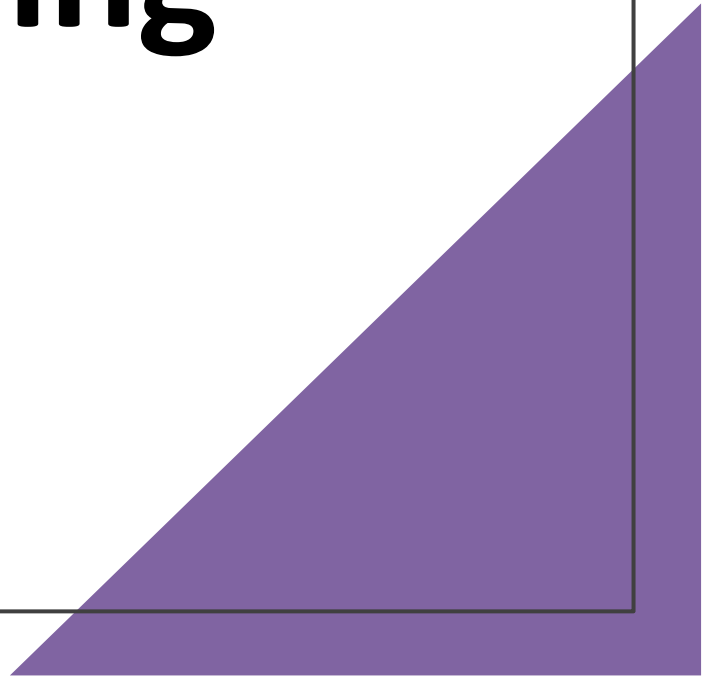
- To ensure clients are protected, lawyers using generative artificial intelligence tools must fully consider their applicable ethical obligations, including duties to:
 - provide competent legal representation,
 - protect client information,
 - communicate with clients,
 - supervise their employees and agents,
 - advance only meritorious claims and contentions,
 - ensure candor toward the tribunal, and
 - charge reasonable fees.
- Full Opinion available at:
https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/ethics-opinions/aba-formal-opinion-512.pdf

CANADIAN BAR ASSOCIATION AI ETHICS TOOLKIT



- Released Nov. 28, 2024
- Has five parts:
 - Definitions
 - Use of AI in Legal Practice: An Overview
 - Guidelines Relating to Use
 - Choosing AI Tools
 - Conclusion
- Website available at: <https://www.cba.org/resources/practice-tools/ethics-of-artificial-intelligence-for-the-legal-practitioner/?lang=en-ca>

Gen AI Policies, Procedures, and Training



DEVELOPING A GEN AI USE POLICY



- **Check with your insurance carrier and your local law society** regarding any applicable rules, regulations, or guidance so you can build them into your policy
 - Also check the Court Notices and Practice Directives in the jurisdictions you practice in!
- **Define the scope of the policy**
 - Which Gen AI tools are covered (*i.e.*, permitted) and which are not?
 - Who may use them and who may not?
 - Under what circumstances may they be used or not (*i.e.*, for internal purposes only; not for external communications or work product)?
 - On what devices may they be used for work purposes?
- **Develop and deliver a training program**
 - Mandatory
 - Must precede any work-related use

DEVELOPING A GEN AI USE POLICY (CONT'D)



- Should cover, at minimum, the following topics:
 - The basics of the technology to be used
 - Its capabilities and limitations
 - How to use the technology properly (*e.g.*, structuring a good prompt)
 - Ethical and legal considerations, including but not limited to, best practices for secure data handling, protecting privacy and confidentiality, intellectual property issues, etc.
 - Examples of appropriate and inappropriate uses
- **Establish a list of appropriate or permissible uses**, for example:
 - Summarization and translation of lengthy documents
 - First drafts of routine internal correspondence or internal memoranda
 - Brainstorming deposition questions, identifying legal issues and arguments, etc.
 - Developing marketing or creative copy
 - Verification and validation are required in all cases

DEVELOPING A GEN AI USE POLICY (CONT'D)



- **Establish a list of inappropriate or impermissible uses**, for example:
 - **Generation of any false or misleading information; creation of any deepfake or other content that impersonates someone, even as a joke**
 - **Violation of any intellectual property laws or regulations or infringement on any third-party rights or interests**
 - **Facilitating any form of harassment or discrimination**
 - **Unauthorized research on any client, counsel, or judicial officer**
 - **Failure to exercise professional judgment, ongoing supervision, and/or professional judgment**
 - **Any other conduct that violates the Rules of Professional Conduct, or any ethical or firm guidelines**
- **Implement regular monitoring and compliance activities**, including:
 - **Periodic audits and reports on user activity, access controls, QC, etc.**
 - **Updates to policies, procedures, and training as the technology evolves**

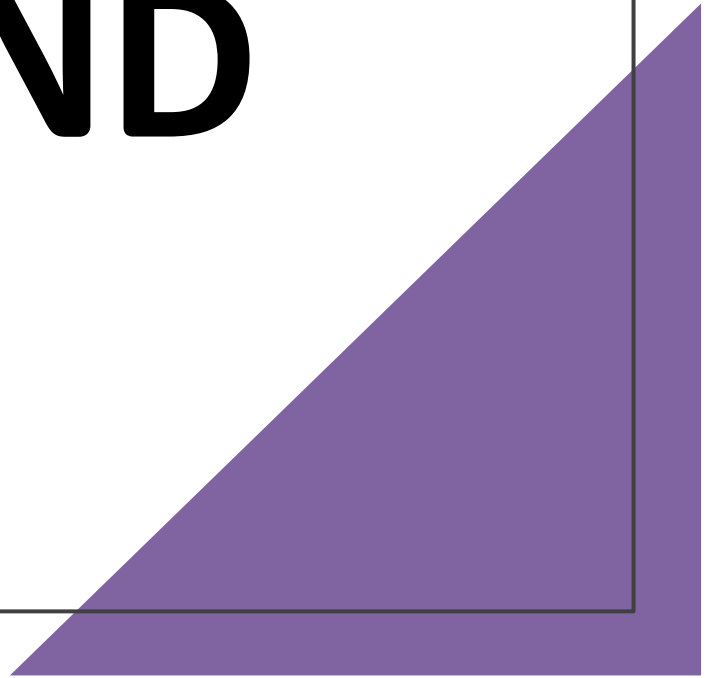
DEVELOPING A GEN AI USE POLICY (CONT'D)



- **Other considerations**

- **Who should personnel contact with questions or concerns?**
- **What records should be maintained and for how long (*e.g.*, proof of review and verification?)**
- **Any disciplinary action that may be taken for violations of the policy**
- **Does the firm's engagement letter need to be amended?**
- **What discussions need to be held with clients?**
- **How will you handle billing for work done using GenAI?**

**ONE MORE ISSUE
TO BEAR IN MIND**

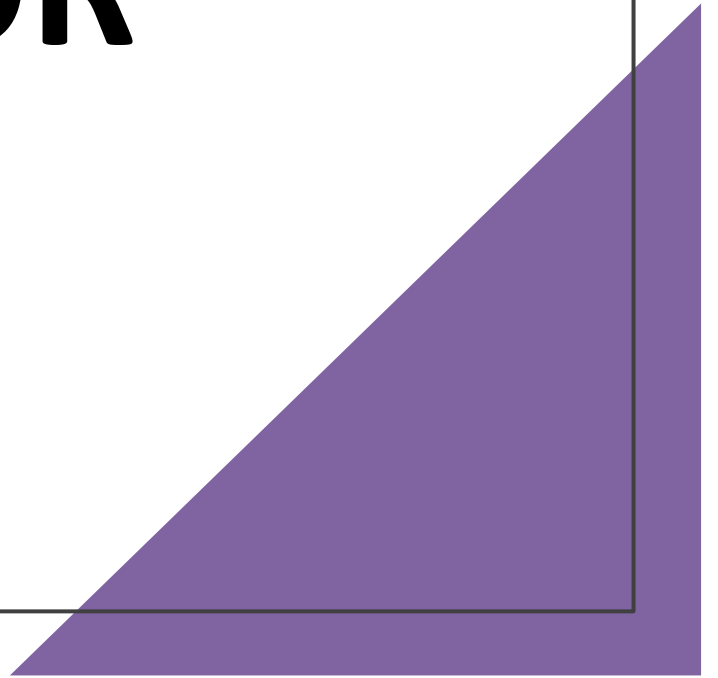


USE OF GEN AI IN CONNECTION WITH VIDEOCONFERENCES

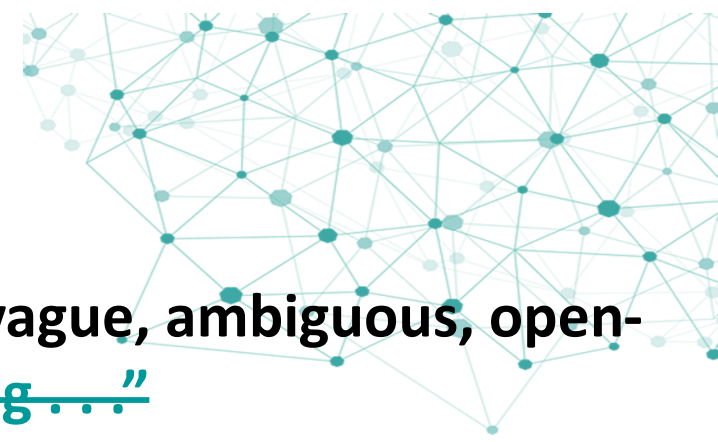


- Many of the videoconferencing systems we all use (*e.g.*, Zoom, Microsoft Teams) now have **embedded automated AI note-taking systems** or **allow recordings to be made of meetings which can be entered into Gen AI systems for transcription or summarization**
- Be mindful that when you use such systems, you are using Gen AI and it may involve information that is obtained from a client, or information that is confidential, privileged, or subject to a protective order; therefore, **you MUST seek and receive the consent of all participants on the videoconference before using such systems**, otherwise you can potentially run afoul of wiretapping and two-party consent laws
- You should always **review and correct the summary or transcript; do not retain records that are inaccurate**

**[BONUS CONTENT:
BEST PRACTICES FOR
PROMPT
ENGINEERING]**

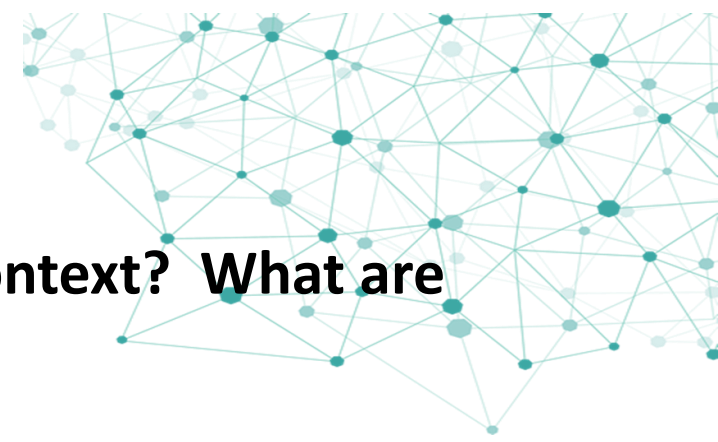


PROMPT ENGINEERING TIPS



- The key is to provide **clear, specific, detailed instructions**, not vague, ambiguous, open-ended questions or statements, *e.g.*, ~~“Write me an email saying . . .”~~
- **Longer instructions are not necessarily better**; what you want is to be **succinct, descriptive, and precise**. Delete unnecessary or superfluous information so as not to confuse or overload the tool.
- **Start by specifying your target audience** so the tool will understand what tone, level, and style to use, *e.g.*: “This is for . . .” or “The tone should be . . .”
 - a novice reader
 - someone who is tech-savvy
 - Serious, persuasive, and professional
 - Casual, conversational, and humorous

PROMPT ENGINEERING TIPS (CONT'D)



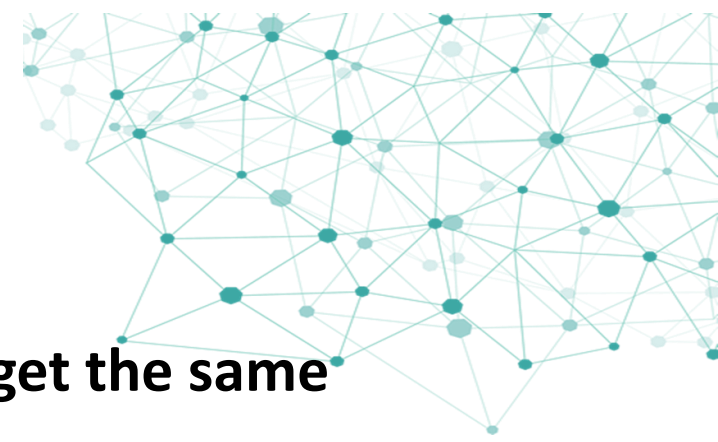
- Next, **provide some background information**. What is the context? What are you trying to accomplish? Specify, *e.g.*,:
 - the subject matter
 - the scope
 - any relevant constraints (“I’m only interested in 2018 to 2022”)
- **Give the tool examples of what you want:**
 - Give actual examples of documents, templates, or cut and paste text or data into your prompt (“show and tell” is very effective)
 - It’s usually better to specify things in positive rather than negative terms, *e.g.*, “I want a concise summary,” rather than “Don’t include too much detail,” but sometimes giving examples of what you do not want can also be helpful

PROMPT ENGINEERING TIPS (CONT'D)



- Specify the format of the output and any limits, *e.g.*, “I want . . .” or “Please include . . .”
 - a bulleted list
 - A report with numbered subheadings
 - Two pages, or 1,000 words or less
- If the task or the instructions are complex and require multiple steps, use chaining or chain-of-thought to break the task or instructions down to one step at a time, *e.g.*, “First do A, then take that answer and do B, and finally add C. Then create a table showing D.”
- If the tool has a “temperature” setting, use 0 for more “truthfulness” and a higher number for more “creativity.”
- If you want to understand how the tool reached its response, ask it to explain its reasoning or process, and/or to provide relevant citations with links.

PROMPT ENGINEERING TIPS (CONT'D)



- **Experiment and iterate, but do not be surprised if you do not get the same response twice in a row.**
- **If you generate a successful prompt for a task you may need to do again, be sure to save it!**

QUESTIONS? THANK YOU!

- **Contact Information**

Maura R. Grossman, J.D., Ph.D.

Research Professor @ David R. Cheriton School of Computer Science & School of Public Health Sciences, University of Waterloo

Adjunct Professor @ Osgoode Hall Law School, York University
Affiliate Faculty Member @ The Vector Institute for Artificial Intelligence

Principal @ Maura Grossman Law, Buffalo, N.Y.

maura.grossman@uwaterloo.ca / maura@mauragrossman.com
/ 212-861-8097



Maura Grossman Law