



Generative AI: You Know Better Than to Trust a Strange Robot



Jim Goepel

General Counsel

Continuous Compliance LLC

JGoepel@FutureFeed.co

About Jim Goepel

- General Counsel and Director of Education at FutureFeed
- Author of 2 books on Controlled Unclassified Information
- Founding Director of the CMMC Accreditation Body (Cyber AB)
 - Created and taught the RP program
 - Board Treasurer
- Co-author of Certified CMMC Professional (CCP) curriculum
- Co-Founder of the CMMC Information Institute
- Adjunct Professor at RIT; former Adjunct at Drexel University
- Expert Witness
- BSECE – Drexel University
 - Designed satellite test equipment and processes
 - Systems Administrator and Developer for the US Congress (House of Representatives)
- JD and LLM – George Mason University
 - Advisor to many government contractors including Unisys and JHU/APL
- Certifications:
 - Certified CMMC Assessor (CCA), CMMC Provisional Instructor, Certified CMMC Professional





Bottom Line Up Front:

- Review the recommendations from the MIT Task Force on Responsible Use of Generative AI for Law;
- Develop an AI policy for your company that addresses the relevant topics outlined in these slides; and,
- Make sure the policy is actually followed by your entire organization and your service providers!



Background

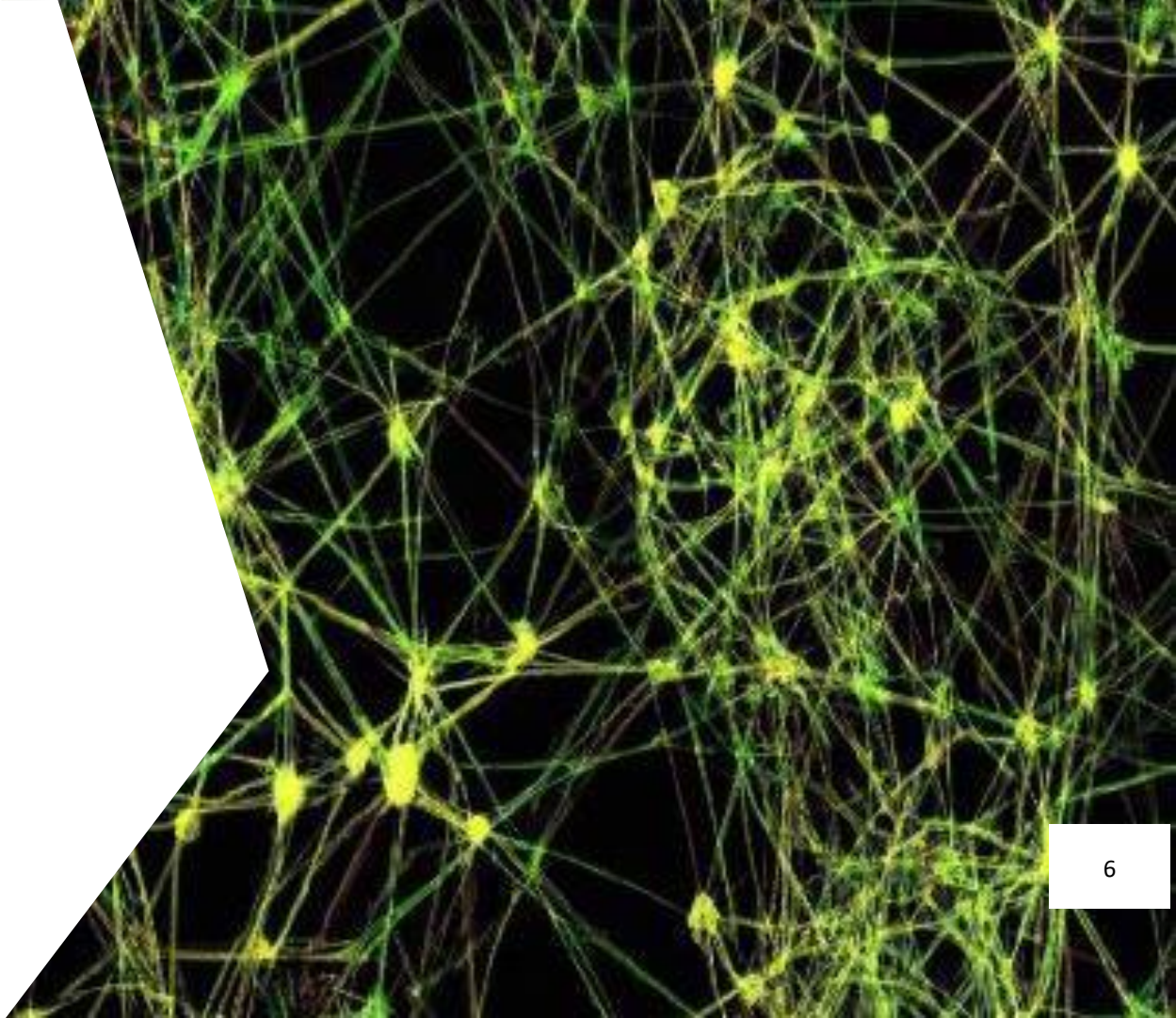


What is Machine Learning?

- Machine Learning uses algorithms to identify patterns in source information (“training data”)
- Categorizes new information based on the patterns

What are Large Language Models?

- New class of machine learning architectures.
- Leverage datasets with hundreds of billions of parameters (Internet-scale datasets).
- Typically trained using “unsupervised learning”.
- Build statistical models that “learn” context and meaning by tracing relationships in sequential data.





What is Generative AI?

- A new form of Machine Learning that leverages Large Language Models.
- Classifies new inputs, and
- Generates entirely new content.

How can Generative AI be Used?



Classify (e.g., image matches, text content)



Translate (e.g., language, writing style)



Summarize (e.g., text, transcripts)



Generate (e.g., audio, source code, images, text, simulations, and videos)



Converse (e.g., chatbots)

Example:

► **Prompt:** Create a standard nondisclosure agreement written in plain English



Non-Disclosure Agreement (NDA)



This Non-Disclosure Agreement ("Agreement") is entered into on [Date], between [Disclosing Party's Name], with its principal place of business at [Address], referred to as the "Disclosing Party," and [Receiving Party's Name], with its principal place of business at [Address], referred to as the "Receiving Party," collectively known as the "Parties."

1. Purpose

The Parties intend to explore a potential business relationship and share certain confidential information for the purpose of evaluating and discussing this relationship.

2. Definition of Confidential Information

"Confidential Information" refers to any information, data, materials, or knowledge disclosed by the Disclosing Party to the Receiving Party, whether orally, in writing, electronically, visually, or by other means, that is designated as confidential or reasonably understood to be confidential due to its nature and the circumstances surrounding its disclosure.

3. Obligations of the Receiving Party

The Receiving Party agrees to:

- Keep all Confidential Information received strictly confidential and not disclose it to any third parties without prior written consent from the Disclosing Party.
- Use the Confidential Information solely for the purpose of evaluating the potential business relationship between the Parties.
- Limit access to the Confidential Information to its employees, contractors, or agents on a need-to-know basis and ensure they are bound by confidentiality obligations at least as protective as those in this Agreement.

MIT Task Force on Responsible Use of Generative AI for Law: Principles

- 1 Duty of Confidentiality to the client in all usage of AI applications**
- 2 Duty of Fiduciary Care to the client in all usage of AI applications**
- 3 Duty of Client Notice and Consent* to the client in all usage of AI applications**
- 4 Duty of Competence in the usage and understanding of AI applications**
- 5 Duty of Fiduciary Loyalty to the client in all usage of AI applications**
- 6 Duty of Regulatory Compliance and respect for the rights of third parties, applicable to the usage of AI applications in your jurisdiction(s)**
- 7 Duty of Accountability and Supervision to maintain human oversight over all usage and outputs of AI applications**

*Consent may not always be required - refer to existing best practices for guidance. We also seek feedback on whether or when consent may be advisable or required.





Bias and Accuracy Issues in AI

Bias

- All machine learning depends on training information
- Bias can creep into training information
 - Source selection
 - Bias in selected source content
- AI can amplify existing biases
- AI can detect bias



Identifying and Addressing Bias



Educate your employees



Ensure diverse teams and training data are involved wherever possible



Preprocess the training data to identify and remove bias



Solicit feedback from users and experts to identify potential bias

TECH · A.I.

Over just a few months, ChatGPT went from correctly answering a simple math problem 98% of the time to just 2%, study finds

BY PAOLO CONFINO

July 19, 2023 at 7:29 PM EDT



Accuracy

- Uncurated data often includes inaccurate information
 - Inaccurate information can exceed accurate information
- Some Large Language Models continue to learn from new inputs
 - Inputs can degrade accuracy over time
 - Some users intentionally “poison” the data



A Rhyming Nondisclosure Agreement

In secrecy and hush, we meet today,
To set the terms in a unique way.
This pact between us, so clear and true,
Shall guard confidential things we do.

1. Purpose of Our Pact

We join our hands with this rhyme so sweet,
To keep our secrets safe and neat.
For exploring ventures, side by side,
Confidential info we'll never hide.

2. What's Confidential, You See

Any info shared, be it day or night,
Shall be kept from the public's sight.
The data, the plans, the secret brew,
All shall remain 'twixt me and you.

3. Promise to Keep, Promise to Hold

I'll keep your secrets, oh so tight,
Away from others' prying sight.
I'll use them only for what we agree,
No sharing unless you say "aye" to me.

4. Exceptions, Just a Few

There are times when the rule won't bind,
Exceptions to keep in mind:
Known before or made public true,

Accuracy Example:

- Generative AI is good at creating linguistically accurate content that is contextually relevant.
- Generative AI is not good at ensuring its content is contextually accurate.

Prompt: Create a standard nondisclosure agreement written in rhyme

Addressing Accuracy Issues

- Review all AI-generated content with skepticism
 - Pretend it was created by opposing counsel
- Ask for and validate sources used

Ask away

ChatGPT can answer questions, help you learn, write code, brainstorm together, and much more.

Don't share sensitive info

Chat history may be reviewed or used to improve our services. Learn more about your choices in our [Help Center](#).

Check your facts

While we have safeguards, ChatGPT may give you inaccurate information. It's not intended to give advice.

Okay, let's go

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

-----X
ROBERTO MATA,
:
Plaintiff, : Case No.: 22-cv-1461 (PKC)
:
- against - :
:
AVIANCA, INC., :
Defendant. :
-----X

16

MEMORANDUM OF LAW BY NON-PARTIES STEPHEN A. SCHWARTZ AND
LEVIDOW LEVIDOW & OBERMAN, P.C. IN
RESPONSE TO MAY 26, 2023 ORDER TO SHOW CAUSE



Ethical and Reputational Risks

Ethical and Reputational Risks

- Fake news and "deepfakes"
- Public perception (not only when you're infringing)
- Legal ethics
- Is generative AI an existential threat to humanity?





Litigation and eDiscovery



Litigation and eDiscovery Risks

Questions (mostly) without good answers:

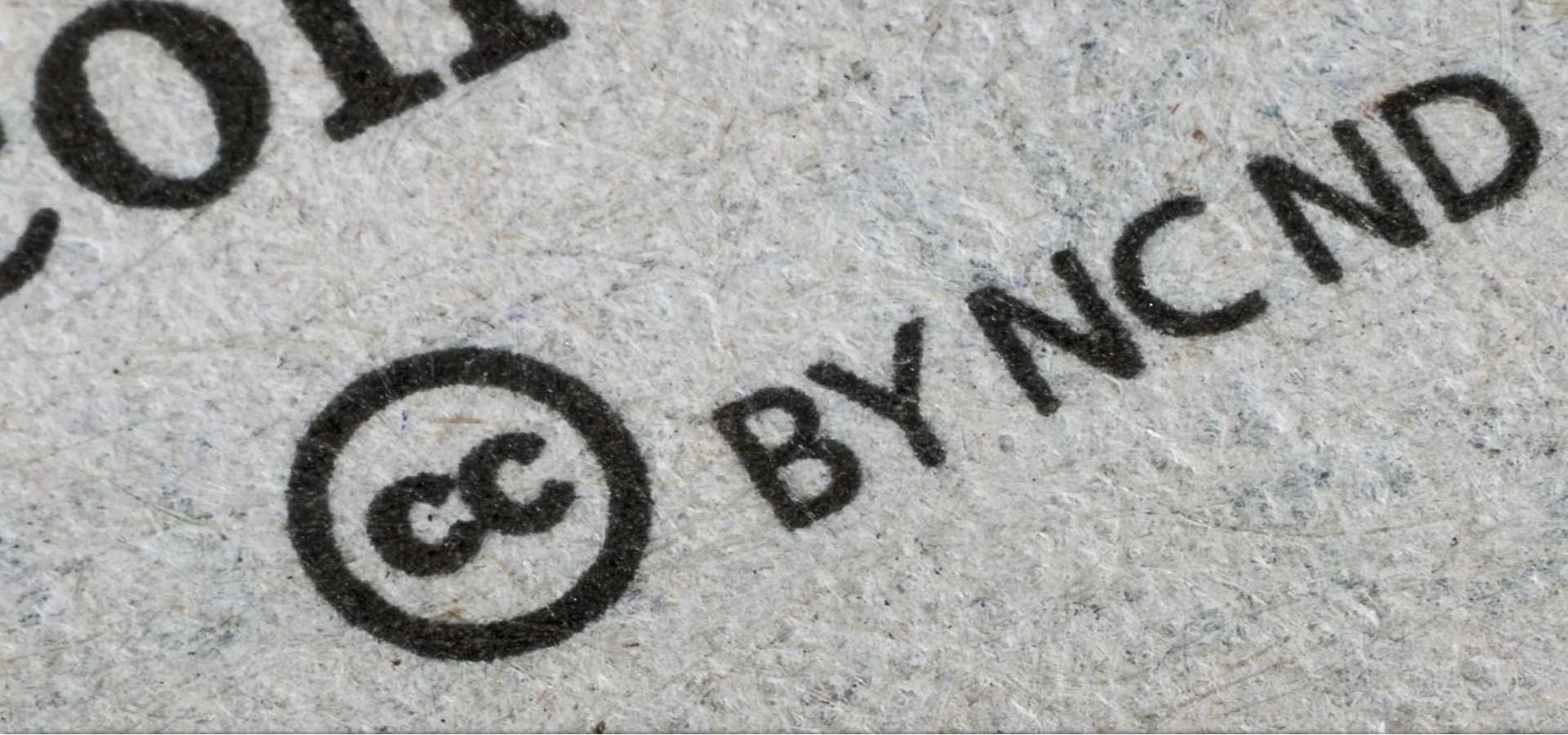
- What do you do when text "hallucinated" by generative AI is responsive?
- Can generative AI draw conclusions from your data that create "knowledge"...even if no one saw the conclusions?
- Can you avoid over-retention of information?
- Where does it get stored, anyway?
- What do we do about deepfakes?



Confidential Information Protection

Confidential Information Protection

- Your inputs may be used to train or fine-tune a model, giving others access to that information
- Your inputs may be deemed public disclosures
 - But see *The Hurry Family Revocable Tr. v. Frankel*, No. 8:18-cv-2869-CEH-CPT, 2023 U.S. Dist. LEXIS 534 (M.D. Fla. Jan. 3, 2023) (quoting a California state court case that "[p]ublication on the Internet does not necessarily destroy the secret if the publication is...limited so that it does not become generally known to the relevant people and where the secret is "not easy to access")
- Your inputs may be hacked, sold, or inadvertently released
- Like other SaaS tools, your employees may retain access after leaving the company
- Bottom line:
 - Tell your employees not to put confidential information into public generative AI platforms!
 - Consider entering into an agreement with a provider that can provide a service with access limited to your company



Intellectual Property

Intellectual Property Protection

- Authorship: Human authorship is the bedrock of copyright law. AI, without human involvement, cannot "author" work. See *Thaler v. Perlmutter*, D.D.C., No. 1:22-cv-01564, 8/18/23). Note that AI-assisted works, with human input, are copyrightable. See, e.g., In re: Zarya of the Dawn, Registration # VAu001480196, 2/21/23 (US Copyright Office's grant of a limited copyright registration for an artificial intelligence-assisted graphic novel).
- Inventorship: AI cannot be the inventor of a patent. The term "inventor" under the U.S. Patent Act construed to require a human inventor. See *Thaler v. Vidal*, 43 F.4th 1207 (Fed. Cir. 2022). Denied cert. on April 24, 2023. Note that AI-assisted invention may be patentable with involvement of human inventor, provided that at least one human rises to level of a coinventor under the law.

Intellectual Property Infringement Risk

- Copyright: Can generative AI use copyrighted content as input? Infringing, but may be permissible under Fair Use doctrine. See *Google Books*, *Warhol*, but also see recent series of cases against generative AI systems (e.g. *Getty Images v. Stability AI*, etc).
- Licensed/Proprietary Data: Can web-scraped or AI bot gathered data from online databases be used as input/training data? Ethical? May be infringing but depends on various factors.
- Right of Publicity and Name, Image, and Likeness (NIL) rights: Use of celebrity NIL likely infringing right of publicity. But note right of publicity is a) governed by state laws, and not uniform, and b) only applies to commercial uses.
- Trademark: Courts likely to apply Section 43(a) of Lanham Act in cases involving unauthorized use of celebrity NIL by generative AI to falsely suggest the celebrity endorses or sponsors a service or product. See, e.g., *White v. Samsung Electronics* and *Wendt v. Host Int'l, Inc.*

Getty Images v. Stability AI Case

- Complaint alleges that Stability AI unlawfully uses Getty's copyrighted images, associated text and metadata as input to train its text-to-image tool Stable Diffusion.



Input



Output



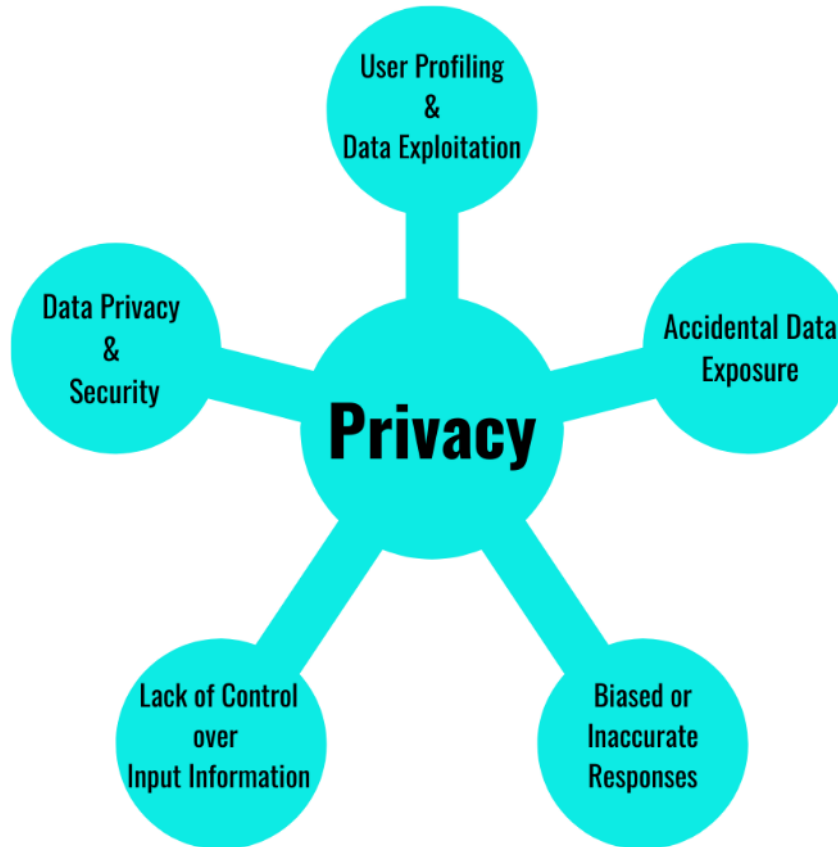
Intellectual Property: Practical Tips

- To assist in proving authorship (copyright) or inventorship (patent) when working with AI tools, be sure to track human involvement and direction in the creative and inventive process (e.g., instructions, notes, files, edit history showing control of process)
- If coders use generative AI to assist in coding:
 - To reduce risk of infringement: double-check to make sure that code from generative AI is not copyrighted or proprietary.
 - To reduce risk of open-source contamination: use an open-source inspection tool on code produced by generative AI.
 - To leverage publicly available generative AI and reduce risk: download an open-source large language model and run it locally, to prevent loss or release of sensitive data, code and trade secrets.
- If using generative AI to create ad copy or marketing materials, be sure to check that imagery and text is not copyrighted, and that claims made are accurate and not false or made up, to reduce risk of infringement or false advertising claims.



Privacy

Privacy





Privacy (continued)

1. Data Privacy and Security: AI systems like ChatGPT require large amounts of data to train effectively. This data often includes personal information and sensitive details. If not handled properly, this data could be vulnerable to breaches, hacking, or unauthorized access, leading to privacy violations and potential identity theft.

2. User Profiling and Data Exploitation: AI systems can analyze user interactions to build profiles and understand preferences, behaviors, and sentiments. This information could potentially be exploited by companies for targeted advertising, manipulation, or even sold to third parties without user consent.

3. Accidental Data Exposure: AI systems may inadvertently generate responses that contain personal, sensitive, or confidential information shared by users during interactions. This could happen if the AI doesn't properly understand context or if the user inadvertently provides such information.

Privacy (continued)

4. Lack of Control over Shared Information: Users might not have full control over what information they share with AI systems. They might disclose more than they intended, thinking they're communicating with a machine. This information could be stored, analyzed, and used in ways that users aren't comfortable with.

5. Biased or Inaccurate Responses: AI systems, including ChatGPT, can sometimes produce biased or inaccurate responses due to the biases present in the training data. These responses could perpetuate stereotypes, misinformation, or discriminatory attitudes, potentially infringing on users' privacy by affecting their perceptions and decisions.



Code of Federal Regulations

3
2005 Compilation
and
Parts 100–102
Revised as of January 1, 2006

The President

Published by:
Office of the Federal Register
National Archives and Records
Administration
A Special Edition of the Federal Register

Other Regulatory Risks

Other Regulatory Risks (1)

- Trend towards transparency, explainability, discrimination analysis requirements, e.g.:
 - GDPR provisions on automated decision-making and profiling
 - Facebook/DOJ settlement in 2022 for housing discrimination due to algorithm for targeting of ads, including a “Lookalike Audiences” tool to target ads based on current customer base (Facebook paid maximum penalty of \$115k under Fair Housing Act)
 - EEOC Draft Enforcement Plan (published Jan. 2023) targets AI hiring bias; various states, including Illinois, have state laws on employer use of AI for hiring
- Draft EU Artificial Intelligence Act bans social scoring, manipulation, some facial and emotion recognition; adds transparency, safety, human oversight requirements for “high-risk” AI; requires summaries of copyrighted material used for training models and measures to safeguard against content that violates EU laws

Other Regulatory Risks (2)

- China draft rules require advance government security review; training data must comply with laws; holds companies responsible for their models' output (which cannot threaten state or social order or discriminate)
- U.S. FDA Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan
- Further regulation is likely; several countries have formed task forces to explore how to regulate AI/Generative AI



HR/Morale Impact

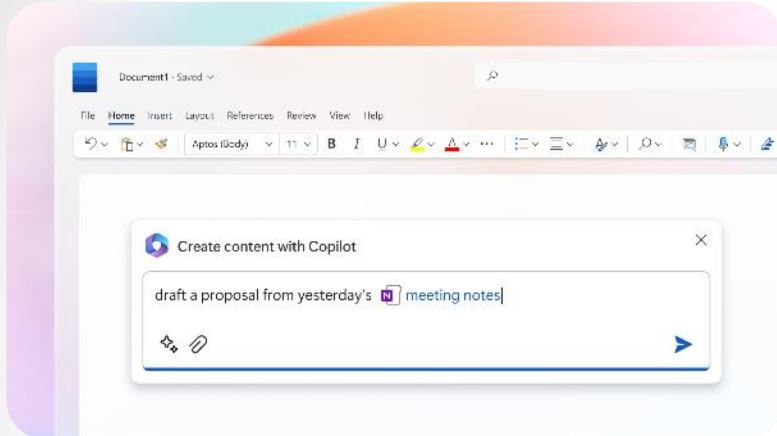
Big Picture

- It is going to be almost impossible to prohibit the use of AI/LLMs.
- Your employees are probably already using it.



Employee Concerns

- Reduced Creativity and Autonomy
- Reduced Efficiency





Employee Concerns

- Fear of Job Displacement
- Perceived Unfairness
- Changes in Work Dynamics
- Job Role Redefinition
- Skill Gap and Training Needs

Charting a Path Forward

- Create an initial draft AI policy
- Work with HR to establish an AI steering committee to refine the policy
- Select members with diverse backgrounds (age, technology competence, job functions)
- Have the members educate fellow employees on AI and the company's AI policy






DIY AI/LLMs

Should we DIY a Large Language Model?

- Not as crazy as it sounds
- Depends on what “DIY” means to you
- Can you get a sufficiently large set of trusted training materials?
- Can your organization gain access to the compute resources? (GPUs are in short supply)
- Can your organization afford the associated compute costs? (GPT-3 is estimated to have cost approx. \$12M in compute costs alone)





Consider a Hybrid Approach

- Some platforms offer marketplaces where you can purchase or rent pre-trained models.
- Opensource tools like Hugging Face's Transformers can kick-start the customization process.
- Microsoft's new Azure AI service might offer sufficient protections to address some concerns.



Q&A

Thank You!

- [Please provide session feedback](#)