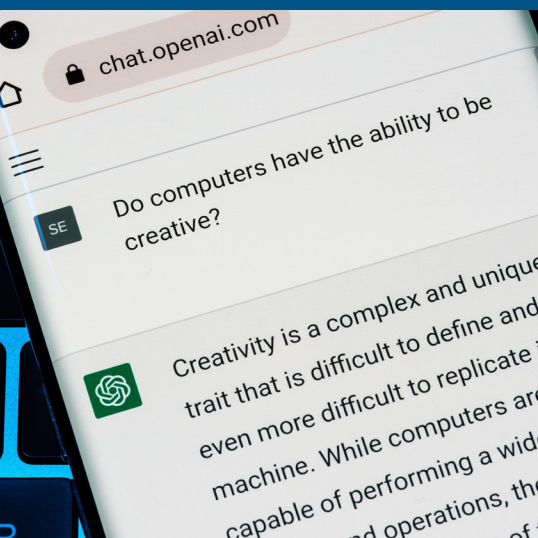**Littler Report**

An Overview of the Employment Law Issues Posed by Generative AI in the Workplace

May 2023

Littler

Authors:

Zoe Argento
ChatGPT
Michael Chichester
Michelle Clark
Tessa Gelbman
Philip L. Gordon
Corinn Jackson

Allan King
Miguel Lopez
Deborah Margolis
Melissa McDonagh
Michael McGuire
Ellie McPike
Marko Mrkonich

Jim Paretti
Niloy Ray
Scott Rechtschaffen
Erin Reid-Eriksen
Kellen Shearin
Alice Wang

This Littler Report is dedicated to Garry Mathiason—our colleague, friend, and visionary—who has taught us to embrace the future and welcome the change that it brings.

Table of Contents

I.   Introduction

A. Artificial Intelligence Arrives in Everyday Life – The Release of ChatGPT

Generative AI, which is a type of artificial intelligence that can produce or create new content, has already started to impact the workplace in various ways. On the positive side, it can automate repetitive and time-consuming tasks, leading to increased efficiency and productivity. For example, it can assist with data entry, customer service, and content creation. Additionally, it can help businesses to analyze and make sense of large amounts of data, leading to better decision-making.

However, the rise of generative AI also raises concerns about its impact on labor and employment. It is predicted that many jobs that involve routine tasks, such as data entry, customer service, and content creation, could be replaced by AI systems in the future. This means that some workers may need to reskill and transition into new roles to stay relevant in the job market.

Furthermore, generative AI has the potential to create new jobs that did not exist before. For example, it could lead to the creation of new job roles that involve working with and managing AI systems. However, it is unclear whether these new jobs will be enough to offset the job losses that may occur due to automation.

Overall, the impact of generative AI on labor and employment is complex and multifaceted. While it can bring about positive changes, it also poses challenges for workers and businesses.

This description of generative AI and its potential impact on the workplace and employment generally was written by **ChatGPT**, the artificial general intelligence platform that has generated more excitement, promise, trepidation, and fear for the future of humanity, than any technological innovation since the invention of the lightbulb, the internet, or the iPhone.

For over three quarters of a century, the promise—or threat—of artificial intelligence (AI) has been largely the province of science fiction books and movies. Isaac Asimov's *I, Robot* stories; HAL in *2001: A Space Odyssey*; the ubiquitous and omniscient computer on the bridge of the Starship Enterprise in *Star Trek*; and C-3PO in the *Star Wars* saga, have stimulated imaginations with the concept of seemingly sentient software capable of answering any question posed—with access to universal knowledge—or, in darker versions, creating inscrutable threats to humanity.[1] At the same time, serious academics have postulated about the "singularity," the point at which technology becomes self-propelled and advances at incomprehensible rates to the point where it moves outside of human control.[2]

---

1   One anonymous but widely quoted observer has defined *artificial intelligence* as "the study of how to make real computers act like the ones in the movies."

2   *The Singularity Is Near*, Wikipedia (last accessed Apr. 16, 2023).

Over the past few decades, there have been great advances in AI—specifically, in areas such as machine learning, deep learning, and natural language processing—and many of these AI applications have been deployed widely, often without users even recognizing that they are interacting with AI. For example, when one's iPhone predicts the next word while drafting a text, that is AI—specifically, machine learning—using simple language models to understand the input and make recommendations based on the user's commonly used words.

But AI truly capable of answering questions and responding with human-like intelligence has largely remained in the realm of science fiction and academic speculation. *Until now.*

On November 30, 2022, research company OpenAI released **ChatGPT**, an exponential advance in the field of **generative AI**, a type of artificial intelligence that generates new material, such as literature, images, audio, video, and music.[3] ChatGPT is, essentially, a chatbot powered by a so-called large language model (LLM) built and tuned using supervised learning and reinforcement learning with human feedback (RLHF) techniques, which can respond to natural language queries with seemingly human-like intelligent answers. Generative AI is a technology that has come the closest to mimicking the human thought process by using trained and untrained data and algorithms to examine patterns in written content across the web to respond by creating novel and original content, rather than simply responding by using pre-existing content or data. ChatGPT appears to engage the user in interactive conversations. For example:

> *User* *Hi ChatGPT. What gift should I take to a dinner party?*
>
> *ChatGPT* *Hello! Bringing a gift to a dinner party is a thoughtful gesture that shows your appreciation for the host's hospitality. Here are some ideas for gifts you could consider:*
>
> 1. *A bottle of wine or champagne - This is a classic and well-appreciated gift that can be enjoyed during the party or saved for later.*
>
> 2. *Flowers or a plant - A beautiful bouquet of flowers or a potted plant can add a nice touch to the host's home and show your gratitude.*
>
> 3. *Gourmet food items - Consider bringing a basket of high-quality chocolates, artisanal cheese, or other gourmet treats.*
>
> 4. *A scented candle - A nice candle can create a cozy atmosphere and be a thoughtful gift for the host to enjoy after the party.*
>
> 5. *A small home decor item - If you know the host's taste, a decorative item such as a vase or picture frame can make for a thoughtful and lasting gift.*
>
> 6. *A book or cookbook - If you know the host's interests, consider gifting a book or cookbook they might enjoy.*
>
> *Ultimately, the most important thing is to choose a gift that you think the host will appreciate and enjoy.*

---

3   San Francisco-based **OpenAI** was founded in 2015 by Sam Altman (CEO), Greg Brockman (Chairman & President), Reid Hoffman (CEO of LinkedIn) and others. Its initial investors included Microsoft, Amazon, Reid Hoffman's foundation, Elon Musk, and Khosla Ventures.

> User   I don't know the hosts very well, what do you recommend?
>
> ChatGPT   If you don't know the hosts very well, it's best to choose a gift that is more generic and universally appreciated. Here are some ideas:
>
> 1.   A bouquet of fresh flowers - You can never go wrong with a beautiful bouquet of fresh flowers. It's a thoughtful and classic gift that can brighten up the host's home.
>
> 2.   A box of gourmet chocolates - A box of high-quality chocolates is a great gift that almost everyone loves. It's also easy to share with other guests.
>
> 3.   A bottle of olive oil or vinegar - If the hosts enjoy cooking, a bottle of high-quality olive oil or vinegar can be a useful and appreciated gift.
>
> 4.   A scented candle - A nice candle with a pleasant scent is a simple but thoughtful gift that can be enjoyed during or after the party.
>
> 5.   A small potted plant - A small potted plant is a low-maintenance gift that can add some life to the host's home.
>
> Remember, the most important thing is to show your appreciation for the invitation and to choose a gift that is thoughtful and appropriate for the occasion.

Critically, despite the clarity and cohesiveness of the responses above, ChatGPT and similar LLM-based AI tools do not operate like encyclopedias, wikis, or reference databases that are organized and searched by topic; rather, these tools analyze the sequence of characters in a user's question or prompts and are acutely tuned to predict what characters should follow next, *i.e.*, what the response to the query should be. The AI does not, however, interpret words, sentences, and language the way humans do, nor make substantive inferences using factual logic. As ChatGPT itself puts it:

> Q: Do LLMs actually understand their inputs?
>
> A: LLMs (Large Language Models) do not "understand" their inputs in the way that humans do. Rather, they are able to generate responses to inputs based on statistical patterns and correlations they have learned from vast amounts of training data.
>
> LLMs are trained on large datasets of text, which allows them to learn patterns and relationships between words and phrases in language. By analyzing these patterns, LLMs are able to generate responses to new inputs that are contextually appropriate and consistent with patterns they have learned from the training data.
>
> However, LLMs do not have a true understanding of the meaning of the words or concepts they are processing. They are unable to comprehend abstract ideas or concepts in the way that humans do, and they cannot reason or make logical deductions.

ChatGPT did not appear overnight. It is the result of decades of development by leading AI firms going back to the 1960s.[4] The "GPT" in ChatGPT stands for Generative Pre-Trained Transformer model, which is a language model based on the transformer architecture, pre-trained in a generative, unsupervised manner that shows decent performance in zero/one/

---

4   *See* Matt White, *A Brief History of Generative AI*, Medium.com (Jan. 7, 2023).

few-shot multitask settings.[5] OpenAI made the third version of GPT, GPT-3, available to developers via an Application Programming Interface (API) in June of 2020.[6] This allowed companies to interact with GPT via software programs. When OpenAI released ChatGPT to the public in November 2022, it was based on GPT-3, a successor to GPT-1 and GPT-2, which were first introduced in 2018-2019. What made ChatGPT novel was that it was open to the public, had a chat interface that used conversational English (or German, or Spanish, or other languages), and that it had been trained on a massive, curated data set that allowed it to intelligibly respond to questions on all subjects.[7]

Upon its initial release, ChatGPT caused a media firestorm—both in the technology and the mainstream media. Within five days, ChatGPT had one million users and within two months, ChatGPT had over 100 million users making it the fastest-growing technology in history (by comparison, it took the mobile phone 16 years to reach 100 million users, iTunes six and a half years, Twitter five years, Facebook four and a half years, and TikTok nine months).[8]

In March 2023, OpenAI released GPT-4, which is purportedly far more accurate, less likely to provide false information, and safer to use than prior versions.[9] According to OpenAI, compared to its predecessor, GPT-4 is 82% less likely to respond to requests for disallowed content and 40% more likely to produce factual responses. It also permits users to input images for analysis through the non-chat interfaces.

To prove its advanced intelligence, OpenAI sent GPT-4 back to school to take a battery of well-known tests. GPT-4 scored in the 90th percentile on the Multistate Bar Exam, including the multiple choice, essay, and performance sections of the exam (GPT-3 had only scored in the 10th percentile). GPT-4 scored in the 93rd percentile on a simulated SAT reading exam and reached the 89th percentile on a simulated SAT math exam. GPT-4 took all Advanced Placement high school exams, scoring between the 84th and 100th percentile, except for a few outliers (it did not do as well in AP English Literature and AP Calculus—although it did earn a score in the latter that might have allowed it to skip a basic college calculus course).[10] GPT-4 even scored 92% on the theory section of the Court of Master Sommeliers exam (and 86% and 77% on the Certified and Advanced Sommelier exams). Unfortunately, it could not pass the tasting section of the exam given its inability to taste wine.

The release of ChatGPT was accompanied by headlines that Microsoft, which in 2020 had invested $3B in OpenAI to purchase an exclusive license to use OpenAI's technology, was investing an additional $10B. In fact, Microsoft had begun working with OpenAI at least five years earlier (in 2018) and it had invested several hundred million dollars developing the infrastructure on which ChatGPT is built.[11]

Microsoft has since incorporated ChatGPT it into its Bing search engine and provided enterprise users of its Azure platform access to GPT.[12] In March 2023, Microsoft unveiled GPT-4-enabled "Co-pilot" and "Business Chat" that would enable users to edit, summarize, create and compare documents, transcribe and summarize meeting notes, and understand what is in

---

5   For example, a "zero-shot" task is a problem set up in machine learning where the software reviews samples that were not included in its training data set and needs to predict their classification without previously learned information.

6   *https://openai.com/blog/openai-api*.

7   *Large, creative AI models will transform lives and labour markets*, The Economist (Apr. 22, 2023).

8   World of Statistics @stats_feed, Mar. 4, 2023.

9   There were reports that users could prompt GPT-3 to explain how to engage in nefarious activities such as how to shoplift, make and use pyrotechnic chemicals, or take over the world. Janus Rose, *OpenAI's New Chatbot Will Tell You How to Shoplift And Make Explosives*, Vice.com (Dec. 1, 2022).

10   GPT-4 Technical Report, Open AI Website (Mar. 27, 2023).

11   Jacob Roach, *Microsoft explains how thousands of Nvidia GPUs built ChatGPT*, Digitaltrends.com (Mar. 13, 2023).

12   Eric Boyd, *General availability of Azure OpenAI Service expands access to large, advanced AI models with added enterprise benefits*, Azure.microsoft.com (Jan. 16, 2023).
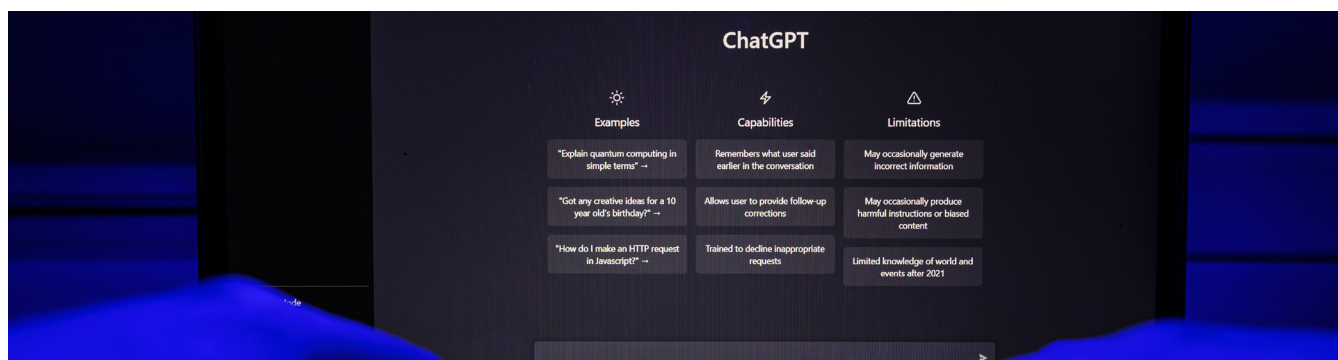
users' emails and calendars to draft responses and summarize project information—across multiple applications including Excel, Word, PowerPoint, Outlook, and Teams.[13]

OpenAI has also launched ChatGPT-4 as an API service providing business customers with access to its chatbot technology enabling them to incorporate ChatGPT into their own offerings. Almost immediately, a wide range of online providers announced that they would incorporate GPT technology into their offerings including Salesforce, Instacart, Shopify, Snapchat, Expedia, and OpenTable.

Numerous alternatives to OpenAI's ChatGPT have emerged including Google's BARD and Meta's LLaMA. In parallel, there have been other uses of large language models to emulate human creativity including OpenAI's popular DALL-E, an application that enables users to generate images based on text input (*e.g.*, "draw a picture, in the style of an impressionist painter, of a squirrel in goggles wearing a purple scarf riding a skateboard"). Similar tools allow for the creation of videos, images, voice from text, and more.

B. The Darker Side of Generative AI

There are, however, questions about the accuracy and reliability of programs such as GPT—and not without justification. As internet content is filled with erroneous and deceptive content, GPT often relies upon dubious sources and questionable material, which are frequently wrong. The consequence is that GPT produces incorrect responses, which users may believe to be true. The OpenAI website readily acknowledges this in its initial release of ChatGPT:

> *From hallucinating inaccurate information, to offensive outputs, to bias, and much more, language models may not be suitable for every use case without significant modifications. Consider whether the model is fit for your purpose and evaluate the performance of the API [application programming interface] on a wide range of potential inputs in order to identify cases where the API's performance might drop. Consider your customer base and the range of inputs that they will be using, and ensure their expectations are calibrated appropriately.*[14]

AI developers have come to refer to these aberrational experiences with generative AI platforms as "hallucinations," when the AI produces outputs that significantly deviate from what would normally be expected. More concerning, these platforms

---

13   Samantha Murphy Kelley, *Microsoft is bringing ChatGPT technology to Word, Excel, and Outlook*, CNN Business (Mar. 16, 2023).

14   Even as recently as April 2023, OpenAI's ChatGPT web site included the following caveats:

   May occasionally generate incorrect information
   May occasionally produce harmful instructions or biased content
   Limited knowledge of world and events after 2021

often appear to completely fabricate facts and sources but in an authoritative manner, leading users to assume that the answers must be correct.[15]

In March 2023, an attorney informed her client that she was unable to find any helpful cases to support her client's position in a compensation-related matter. Days later, the client emailed that he had found six cases that were "100% on point," and included captions, case summaries, and Westlaw citations. The attorney, feeling embarrassed but unable to locate those cases on Westlaw, asked the client how he found those cases. The client responded that he had found them using ChatGPT. The cases were complete computer-generated fabrications, including the captions, summaries, and Westlaw citations.[16]

More alarming, certain generative AI interactions with the pre-GPT-4 model have been described as basically "unhinged." In February 2023, after spending two hours having a conversation with Microsoft's Bing AI, Kevin Roose, a technology columnist for the New York Times, wrote about "the strangest experience I've ever had with a piece of technology" in which the software professed its love for him and argued that Roose didn't really love his spouse. Roose wrote that the conversation with the software "unsettled me so deeply I had trouble sleeping afterward."[17]

Following concerns about reported inaccuracies and hallucinations, and after the release of GPT-4, which purportedly significantly reduced occurrences of these issues, OpenAI published on its website Usage Policies describing how its technology should or should not be used. Among other things, these include:

> We have further requirements for certain uses of our models:
>
> 1. Consumer-facing uses of our models in medical, financial, and legal industries; in news generation or news summarization; and where else warranted, must provide a disclaimer to users informing them that AI is being used and of its potential limitations.[18]

Of course, every new technological development has generated skepticism and experienced problems during its early introductory period. Skeptics questioned whether the iPhone or the iPad would ever find acceptance in the workplace given their initial limitations. Eventually, these technologies overcame these initial issues and have since become ingrained in our work and, moreover, our lives.

C. The Risks of Generative AI for Employers

The application of AI in many facets of life, including the workplace, is not new. As Littler reported in 2015:

> The World of Big Data has arrived, and it is beginning to affect employers and their decision-making in ways undreamed of even a few years ago. Employers can access more information about their applicant pool than ever before and have an ability to correlate data gleaned from the application itself, perhaps supplemented by publicly available social media sources, to determine how long a candidate is likely to stay on a particular job. Similarly, by combing through computerized calendar entries and e-mail headers, Big Data can tell us which employees are likely to leave their employment within the next 12 months. At the same time new tools and methods that rely on concepts of Big Data are becoming part of the daily landscape in human resource departments, employers continue to operate in a legal environment based on precedent and history with few guideposts that translate seamlessly into the world of Big Data.

---

15   *See, e.g.,* Cade Metz, *What Makes A.I. Chatbots Go Wrong?*, N.Y. Times (Mar. 29, 2023).

16   *A Cautionary Tale About Clever Clients and AI*, Litigation Daily Online, ALM Media (Mar. 28, 2023).

17   Kevin Roose, *A Conversation With Bing's Chatbot Left Me Deeply Unsettled*, N.Y. Times (Feb. 16, 2023).

18   OpenAI Website, *https://openai.com/*.
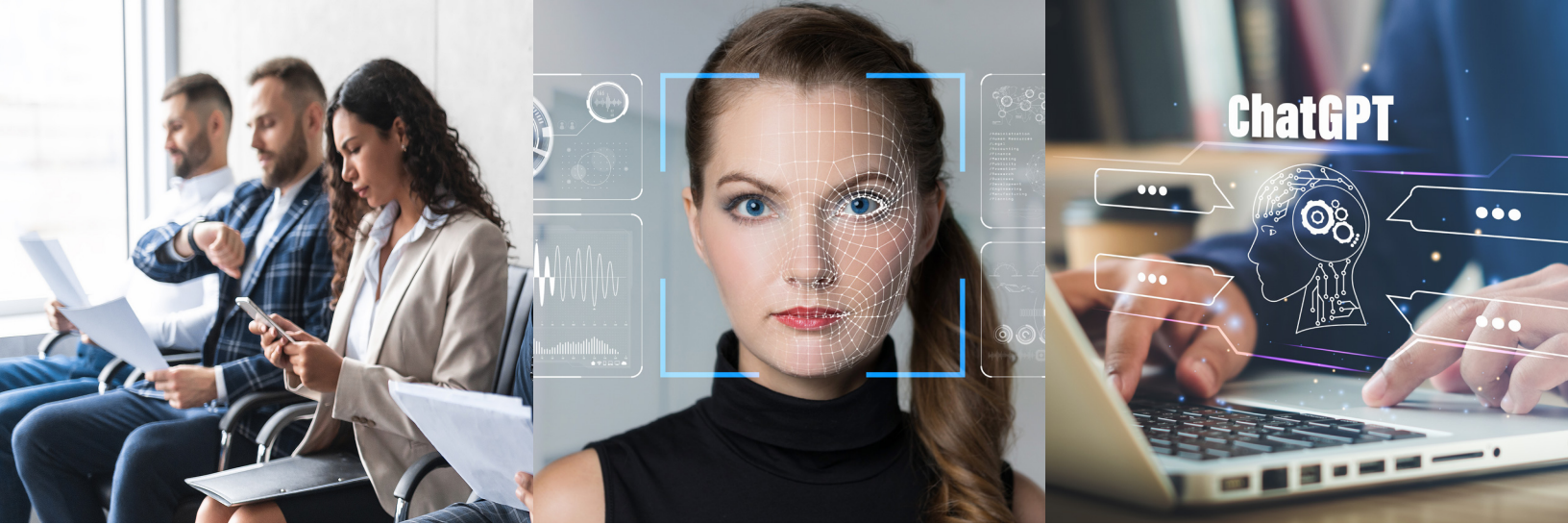
The issues that can arise either are brand new or develop in a context that makes yesterday's compliance paradigm difficult to apply.[19]

AI tools – primarily predictive AI tools – are already being used across the hiring and employment lifecycle. As discussed in Section IV, employers are already using traditional AI tools to screen resumes at a massive scale to help identify whom to interview; to conduct video interviews and other candidate assessments that analyze facial expressions, tone of voice and non-verbal cues in addition to substantive responses, to evaluate a candidate's skills, experience, or 'fit'; to automate onboarding processes and other employee self-help protocols; to track and measure employees' performance, skills, and areas of improvement; and to determine employees' internal growth paths, compensation, and predicted success on the job.

As discussed throughout this report, the addition of generative AI tools and technology will supercharge and expand these existing uses. GenAI will make candidate and employee interaction customizable and personal at scale. Instead of merely providing assessment metrics to managers, human-like generative AI chatbots will be able to provide guidance and coaching using natural language ask-and-respond mechanisms. Generative AI tools will be able to summarize and paraphrase volumes of content such as employee reviews and feedback, provide customized recommendations across the board, and perhaps even become a first-line contact for employee engagement and productivity support. Generative AI Smart Assistants will undoubtedly be a default part of word-processing and other content-creation tools within years, if not months.

This exponential levelling-up of workplace tools will inevitably increase corporate risk and potential liabilities for employers in the short term, at least, as new laws and regulations emerge and are refined, and as existing laws are adapted in form and application to account for the paradigm shift that generative AI is bringing about. Of particular concern is that generative AI, like the traditional AI already in use in the marketplace, is susceptible to algorithmic biases that can result in discriminatory or otherwise impermissible treatment of employees, as discussed in Section II. In addition, the primary use of generative AI—the creation of new content and work product—gives rise to concerns with respect to individual privacy, intellectual property rights, plagiarism, and ethical conduct.

---

19   Marko Mrkonich *et al.*, *The Big Move Toward Big Data in Employment*, Littler Report (Aug. 4, 2015).

D. Other Legal Risks of Generative AI for Organizations

In addition to the risks for companies in their role as employers, the use of generative AI can create other legal risks of which organizations should be aware. Here are some potential legal risks associated with using generative AI, many of which will be discussed in greater detail throughout this report:

- **Defamation claims:** Claims have been made about allegedly defamatory output from these tools about individuals.
- **Consumer protection:** If factually inaccurate output from a generative AI tool is used in advertising or marketing campaigns, it may mislead consumers or violate consumer protection laws.
- **Liability:** If the use of generative AI leads to harm or injury, the organizations or individuals responsible for the use of the technology may be liable for damages.
- **Regulatory compliance:** The use of generative AI may be subject to regulatory requirements, such as those related to consumer protection, financial services, healthcare, or national security.

Overall, the legal risks associated with using generative AI highlight the importance of understanding the legal and regulatory landscape surrounding the technology and ensuring that its use complies with applicable laws and regulations. It is crucial to consult with legal experts and implement appropriate measures to mitigate legal risks when using generative AI.

E. Preparing for the Impact of Generative AI

Beyond the risks of generative AI running afoul of legal and regulatory regimes, there is the broader question of the impact of generative AI on the human labor workforce. As discussed in Section III, generative AI has the potential to massively disrupt the workforce—particularly the white-collar, college-educated workforce—unlike any previous technological development. As discussed in the 2021 *Inaugural Report of Littler's Global Workplace Transformation Initiative*, which focused on the emergence of technology-induced displacement of employment (or TIDE), the workforce was in the process of being radically transformed by robotics and automation with some predicting that almost half of current work activities could be automated by 2055. The COVID-19 pandemic greatly accelerated this transformation. As stated in this report:

> If one thing is plain, it is that the pandemic has accelerated these trends. Where once viewed more as a matter of simple economics (using AI and automation to reduce labor costs), in the pandemic environment, employers are also looking to automation as a way of continuing operations safely, minimizing potential illness in its workforce. In October 2020, the World Economic Forum reported that 43% of the businesses it surveyed were planning to reduce their workforce because of the increased use of technology.[20]

What this report could not anticipate at the time was the exponential advances in AI as represented by ChatGPT and similar generative AI platforms. While the release of ChatGPT, and the advances in AI it represents, was met with tremendous excitement and unparalleled rates of adoption, it was also met with alarm from many corners. As follows the introduction of any massively disruptive new technology, there have been fears, not necessarily irrational, about the significant impact on the labor force.[21] Previous waves of robotic and automation displacement have impacted blue-collar work involving physical tasks such as assembly-line work, and highly structured and predictable tasks such as data collection and

---

20   Michael A. Chichester Jr. *et al.*, *Inaugural Report of Littler's Global Workplace Transformation Initiative*, Littler Report (Mar. 30, 2021).
21   Annie Lowrey, *How ChatGPT Will Destabilize White-Collar Work*, The Atlantic (Jan. 20, 2023).

information processing.[22] Generative AI will impact knowledge workers in creative fields such as journalism, marketing, advertising, academics, legal, and professional services.

As discussed in greater detail in this paper, in March 2023, Goldman Sachs released a report concluding that two-thirds of U.S. occupations will be impacted by some degree of AI automation and that generative AI could replace one-fourth of current work tasks. The study estimates that 44% of legal tasks could be automated by AI. "Extrapolating our estimate globally suggests that generative AI could expose the equivalent of 300M full-time jobs to automation."[23]

In another study conducted by OpenAI and the University of Pennsylvania, using job data from the Department of Labor, researchers found that up to 80% of the U.S. labor force could see at least 10% of their work tasks impacted and 19% of workers may see at least 50% of their tasks impacted. Workers with higher-wage jobs are at greater risk than those in lower wage jobs, with the impact increasing as jobs get closer to $80,000.[24]

Of course, technological innovations often create new jobs to replace those displaced by the technology. In the late 1700s/early 1800s, rather than reducing the overall number of textile workers and weavers, the automated looms so feared by the Luddites made manufactured textiles affordable for the masses (replacing hand-woven textiles), thereby increasing the overall demand for labor in the new textile factories, even though the labor needed was less skilled than the weavers who had previously made such cloth by hand. The internal combustion engine may have displaced the need for those involved in tending to horses and wagons—farriers, blacksmiths, and the proverbial buggy-whip makers—but it created tens of millions of jobs for generations in automobile and parts factories, automobile dealerships, service stations, road building and maintenance operations, and so many other related industries and occupations.

We may see similar job-creating benefits from generative AI. Already, demand has emerged for "prompt engineers," individuals skilled at framing exactly the right type of queries to coax generative AI platforms to generate cogent and accurate responses, at annual salaries currently over $300,000. As explained by a prompt engineer, the best prompt engineers do not have technology backgrounds but rather "a history, philosophy, or English background, because it's wordplay. You're trying to distill the essence or meaning of something into a limited number of words." As one individual who runs a staffing business for AI and machine-learning positions said, "Expert prompt engineers can name their price."[25]

In some sectors, the initial hostile and fearful response to the release of generative AI platforms has transformed into acceptance and, perhaps, recognition that these tools might offer some conceivable benefit. For example, initially, academic institutions reacted with existential alarm at the release of ChatGPT, considering it a threat to learning, forcing them to reconsider how they would evaluate applicants and students on tasks such as entrance exams, writing assignments, and tests, given that ChatGPT can perform these tasks effortlessly. They anticipated rampant cheating and plagiarism.[26] However, once the initial panic subsided, many teachers and academic institutions have come to realize that generative AI platforms might be exposing the deficiencies in their existing pedagogical methods. Academics are considering whether requiring students to answer the type of rote questions that generative AI can answer is really helping

---

22 Chichester *et al., supra* note 20.

23 Joseph Briggs & Devesh Kodnani, *The Potentially Large Effects of Artificial Intelligence on Economic Growth*, Goldman Sachs Economics Research (Mar. 26, 2023).

24 Aaron Mok, *ChatGPT will most likely impact your job if you work in tech, went to college, and make up to $80,000 a year*, Business Insider (Mar. 22, 2023).

25 Conrad Quilty-Harper, *$335,000 Pay for 'AI Whisperer' Jobs Appears in Red-Hot Market*, Bloomberg Law (Mar. 29, 2023).

26 Kathy Hirsh-Pasek & Elias Blinkoff, *ChatGPT: Educational friend or foe?*, Brookings.edu (Jan. 9, 2023).

them learn or think. Just as mathematics and science teachers eventually accepted the invaluable and inevitable role of calculators in the classroom, increasingly, teachers are recognizing the role of generative AI in humanities classrooms.[27]

One sector that is concerned about the potential existential threat of generative AI is the technology community itself. On May 1, 2023, Dr. Geoffrey Hinton, a pioneer in the field of AI, often referred to as one of "the Godfathers of AI," announced that he had resigned from his position at a major technology company in order to talk more freely about his concerns about generative AI. Among other things, Hinton is concerned with generative AI flooding the internet with fake text, photos and videos, undermining any notion of truth online, and the potential negative impact on the labor market. He is also worried that future versions of generative AI could pose a threat to humanity as they begin to generate and run their own code and become truly autonomous.[28]

Other leading technology leaders and researchers, including Elon Musk (an early investor in OpenAI), Steve Wozniak (co-founder of Apple), Yuval Noah Harari (professor and author of *Sapiens*) and over 30,000 others, have called for a pause in the development and training of further large language models until governments, academic institutions, and independent experts can develop safety protocols and a regulatory framework for the use of these models. In their open letter, posted on the website of Future of Life Institute, the authors proclaim that "Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources."[29]

---

27   Pia Ceres, *ChatGPT Is Coming for Classrooms. Don't Panic*, Wired.com (Jan. 26, 2023).
28   Cade Metz, *'The Godfather of AI' Quits Google and Warns of Danger Ahead*, N.Y. Times (May 1, 2023).
29   *Pause Giant AI Experiments: An Open Letter*, Future of Life Institute (Mar. 22, 2023).

II. Generative AI & Potential Bias

A. What is the Difference Between Generative AI and Machine Learning?

Machine learning (ML) is a branch of artificial intelligence (AI)[30] that focuses on teaching machines to recognize patterns in data, and make predictions or decisions based on those patterns. It involves feeding large amounts of data into a computerized algorithm, which then uses statistical techniques to identify patterns and make predictions or decisions about new data.

The key to machine learning is the computer's ability to "learn without being explicitly programmed to do so,"[31] or to "independently perform and improve over time."[32] Machine learning begins with "predictors" or "features," consisting of data (*e.g.*, numbers, pictures, text) that are deemed relevant to a particular question. The data gathered are used as "training data" from which the algorithm learns. The training data are typically randomly divided into a training sample and a hold-out sample. The training sample is used to train the machinelearning algorithm, with instructions about the outcome and the features that may predict the outcome, and the algorithm's determining which features most strongly correlate with the outcome. The "hold out" sample is then utilized as evaluation data, testing how accurately the machine-learning model is when presented new data.[33] As new data are fed into the algorithm, the machine-learning system continues to analyze the data to improve its performance and make better-informed decisions.[34] Eventually, the machine-learning model can be applied to data from outside the sample, where the outcome is unknown. Successful machine-learning algorithms can be *descriptive* (use data to explain past phenomena); *predictive* (use data to make future predictions); or *prescriptive* (use data to suggest what action to take).[35]

Generative AI refers to a category of ML algorithms that generate new outputs based upon the patterns within data on which they have been trained. Like other forms of artificial intelligence, generative AI learns how to take actions based upon past data, but it also creates brand new content based on that training (such as images, videos, text, or computer code) rather than simply categorizing or identifying data.[36] Generative AI involves training a model on a large dataset of examples, and then generating new samples that resemble the original data. During training, the model learns to recognize patterns in the input data and then generates new data by sampling from a probability distribution that is learned during training. This specific training process differs for different types of generative AI.[37]

In essence, *machine learning* is a general approach to teaching machines to learn from data and make predictions or decisions, while *generative AI* is a specific type of machine learning focused on creating new content.

---

30   *Artificial intelligence* is, broadly, the computer science of building machines that can execute tasks normally performed by humans, thereby requiring human-like knowledge. *See* Alyssa Schroer, *What is Artificial Intelligence*, BuiltIn (updated Mar. 3, 2023).

31   Alison Arden Besunder, *Not Your Parents' Robot*, N.Y. St. B.A.J., Mar./Apr. 2018, at 20, 21. *Machine learning* has also been characterized as a "method of data analysis that automates analytical model building." Using algorithms that iteratively learn from data, machine learning reportedly allows computers to find hidden insights without being explicitly programmed where to look. As the models are exposed to new data, they are able to independently adapt. The purpose is for the model to learn from previous computations to produce reliable, repeatable results." *Artificial Intelligence And The Practice Of Law Or Can A Computer Think Like A Lawyer?*, 2016 WL 10611682.

32   Maya C. Jackson, *Artificial Intelligence & Algorithmic Bias: The Issues with Technology Reflecting History & Humans*, 16 J. Bus. & Tech. L. 299, 303 (2021) (*citing* Roberto Iriondo, *What is Machine Learning*, MEDIUM, (last updated May 6, 2022)).

33   Sara Brown, *Machine Learning, Explained*, MIT Sloan School of Management (Apr. 21, 2021).

34   Jackson, *supra* note 32.

35   Brown, *supra* note 33.

36   Reuters, Explainer: *What is Generative AI, the technology behind OpenAI's ChatGPT?* (Mar. 17, 2023).

37   Josh Fruhlinger, *What is generative AI? The evolution of artificial intelligence*, InfoWorld (Mar. 10, 2023).

B. What Are Common Applications of AI in Employment Decisions?

Artificial intelligence is increasingly being used in employment decisions, particularly in the hiring and recruitment process. AI-based hiring tools are designed to help employers sift through large numbers of job applications more quickly and accurately, by automating various aspects of the recruitment process. There are concerns, however, about the potential for AI to perpetuate and amplify biases, which could result in unfair employment decisions.

Some of the ways that AI is being used in employment decisions include:

- **Resume Screening:** AI tools are used to automatically screen resumes for relevant skills and experience, and to rank candidates based on their fit for the job. This saves recruiters a lot of time and helps them identify the most promising candidates more quickly.
- **Video Interviews:** AI-powered video interviewing tools are used to automatically analyze candidates' facial expressions, tone of voice, and other non-verbal cues, to assess their suitability for the job. This helps to standardize the interviewing process and reduce bias.
- **Pre-employment Assessments:** AI tools are used to administer pre-employment assessments, such as cognitive ability tests, personality tests, and skills tests. These assessments are used to screen candidates more objectively and consistently than traditional methods.
- **Self-help Chatbots:** AI tools are used to automate all manner of information look-up and logistics tasks, from candidate management during recruiting and on-boarding tasks to answering HR and benefits questions, providing policy guidance, and even summarizing documents and providing "first drafts" of content.
- **Skill Mapping:** AI tools are used to develop individual employee job-skill profiles and overall corporate skills (and skill gaps). This enables swifter up-skilling and more-precise talent acquisition (to fill the gaps).
- **Performance Management & Succession Planning:** AI tools are used to track employee habits, provide metrics for performance evaluations, monitor engagement, and assist with turnover, promotion, and succession decisions.

While AI-based hiring tools can be useful in streamlining the recruitment process, there is the risk that these tools may reinforce or magnify biases. For example, if the training data used to develop the AI models is biased, the resulting models will also be biased. This could lead to discriminatory employment decisions, such as rejecting qualified candidates based on their gender, race, or other protected characteristics.

C. What Are the Potential Biases Associated with Machine-learning Models and ML-based AI Tools?

Even though machine-learning systems might be regarded as neutral because computer systems are making decisions rather than humans, they still rely on human individuals to create the algorithmic foundations and train the system and thus are susceptible to many sources of bias.[38] Algorithmic bias refers to systemic errors that create unfair outcomes for certain individual or groups. Bias can infuse AI systems in a multitude of ways, including through:

- Human algorithm bias
- Learning bias
- Data collection bias
- Deployment bias

Human algorithm bias. This focuses on the individuals' gathering the data and training the system to create an algorithm to later be used for machine learning.[39] This form of bias reflects that data scientists or engineers who are collecting and training the data have their own personal biases, whether conscious or unconscious, which end up being reflected and reinforced in the data collection and training process.[40]

Learning bias. Training and evaluating the machine-learning model is another opportunity for bias if choices are made that amplify performance disparities across groups. Data scientists or programmers may choose to optimize the aggregate performance across all groups being modeled but doing so results in disparate error rates when examining performance across individual groups. For example, if demographic data is used in a model to predict the likelihood of criminal recidivism, it may be that demographic data such as race or age improves the aggregate performance of the model across populations, but error rates across individual groups vary widely (*i.e.,* a model performs much better for certain groups at the expense of other groups). Another source of learning bias may occur if the algorithm prioritizes more compact models at the expense of the model's ability to create accurate predictions on underrepresented attributes in the dataset. For example, seeking at an aggregate level to minimize computing costs or power requirements may make the algorithm perform worse on groups in populations that are less well represented in the data.

Another source of learning bias can arise in defining the features and labels used in the algorithm. A common type of bias at this learning stage is measurement bias: when features or labels chosen to represent some construct are, in fact, poor representations of that construct. This could be because the choice of label is an oversimplification, or method of measuring that feature or label varies across features or groups.

Data collection bias. This type of bias can arise from historical bias and representation bias. Historical bias occurs when the data collected and used to train the algorithm reflects existing biases in the world around at the time of data collection, and representation bias occurs when the training dataset is not representative of the target population one is seeking to model. There can be different sources of representation bias, *i.e.,* certain groups are historically underrepresented, or the data used to train the algorithm was collected in an uneven or skewed way.

---

38 Jackson, 16 J. Bus. & Tech. L. at 305, *supra* note 32.

39 *Id.*

40 *Id.*

Deployment bias. This occurs when there is a mismatch between how an algorithm was intended to be used and how it is actually used in practice. This can happen when the algorithm is developed in isolation or without sufficient feedback. Take, for example, an automated performance evaluation tool that is originally intended to be used as a complement to other evaluation methods but ends up being used as a primary basis to make termination decisions of low-rated personnel.

The foregoing addresses various biases that can permeate the AI/machine-learning context generally. Since the content produced by generative AI is determined by the underlying training data (which is produced by humans), the output that is generated is also susceptible to many of the same biases and flaws. The below are how some potential biases manifest specifically in generative AI models:

- **Data bias:** Generative AI models are trained on data, and if the data is biased, the model will also be biased. For example, if a generative model is trained on images of predominantly white people, it may generate images that look more like white people than people of other races.

- **Style bias:** Generative AI models can also exhibit style bias, which means they tend to generate content in a particular style that is similar to the training data. For example, a generative model trained on images of a particular artist's work may generate new images that look similar to that artist's work.

- **Historical bias:** Generative AI models can also perpetuate historical biases by generating content that reflects past inequalities and stereotypes. For example, a generative model trained on text from historical documents may generate text that reflects outdated attitudes toward race or gender.

- **User bias:** Generative AI models can be biased based on the preferences of the users who interact with them. For example, if a generative model is used to generate clothing designs, it may generate designs that are biased toward the preferences of the users who are using the model.

To mitigate these biases, it is important to choose the training data carefully, monitor the model's output for bias, and implement methods such as adversarial training and diversity promotion to encourage the model to generate more diverse and fair content.

D. How Can One Tell if AI Is Unbiased?

It can be challenging to tell if a generative AI model is unbiased, as bias can be subtle and difficult to detect. However, here are some methods one can use to evaluate the model's bias:

- **Manual evaluation:** One way to assess the model's bias is to manually review its output and evaluate whether it is biased or not. This can involve comparing the generated content to real-world examples and assessing whether the model's output is representative and fair.

- **Quantitative metrics:** There are various quantitative metrics that can be used to evaluate the bias of a generative AI model. For example, one can calculate the distribution of generated output across different classes or attributes and compare it to the distribution in the seed data. If the generated output is significantly different from the seed data, it may indicate that the model is biased.

- **Adversarial evaluation:** Adversarial evaluation involves testing the model's ability to discriminate between biased and unbiased content. This can involve generating content that is intentionally biased and evaluating whether the model is able to detect and avoid generating such content.

- **Diversity evaluation:** One way to ensure that the model is not biased towards a particular style or attribute is to evaluate the diversity of the generated output. If the model is generating output that is diverse and representative of different styles and attributes, it may indicate that the model is less biased.

Overall, evaluating the bias of a generative AI model can be a complex and ongoing process. It is important to use a combination of manual evaluation, quantitative metrics, adversarial evaluation, and diversity evaluation to assess the model's bias and continually improve its performance.

III. Generative AI & Workforce Transformation

The rapid advancement of generative AI is certain to only accelerate a trend Littler identified in its 2018 white paper, *The Future is Now: Workforce Opportunities and the Coming TIDE*.[41] In that report, we identified the technology-induced displacement of employment (TIDE) as one of the things most likely to fundamentally reshape the workforce.

Generative AI is revolutionary, and no job is immune from its impact. Some jobs will be displaced, demand for others will increase, and yet others will emerge—many of which we cannot even conceive of today, just as a person in 1750 could not have conceived of "electricians," let alone "digital content managers." The labor market is about to experience unprecedented disruption and must properly prepare to survive—and thrive—in this new era.

A. The Impact of Generative AI on Human Labor

As AI models expand, so too do their capabilities and applications. It was once believed that while computers might be faster, humans cornered the market on creativity. The latest wave of generative AI output indicates computers are catching up.[42]

So how will this impact employment? Since ChatGPT's release in November 2022, many have attempted to predict the impact of generative AI on human labor. With the potential to generate novel output based on user prompts, generative AI is likely to impact profoundly the way knowledge workers perform their jobs, and significantly impact work that was previously resistant to automation.

How fast generative AI tools will disrupt human labor is uncertain. One study found that 28 million workers could be impacted by generative AI in 2023 but stressed that "impacted" does not necessarily mean displaced.[43] Many workers will be able to use generative AI tools to help do their jobs, not be supplanted by them. This study examined U.S. Bureau of Labor Statistics O*NET data on 867 occupations and found that lower-

---

41   Michael J. Lotito *et al.*, *The Future Is Now: Workforce Opportunities And The Coming TIDE*, Littler Report (June 2018).

42   Analysts have been trying to project how fast this evolution will occur, and what will be the "next wave" in this space. *See, e.g.*, *Generative AI: A Creative New World*, Sequoia.

43   Steven Peavey, *18% of the U.S. workforce are likely to be affected by generative AI in 2023*, Moser Consulting.
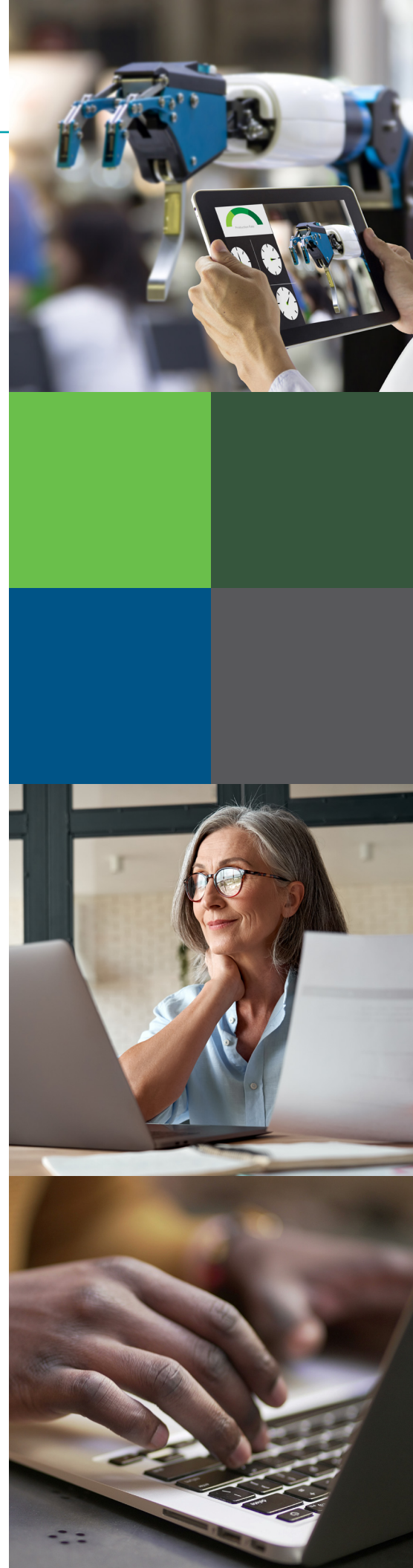
skilled jobs that required the least amount of preparation will likely be those most negatively affected by generative AI in the coming year.[44]

As generative AI tools continue to improve, employees in the creative fields, such as marketing and graphic design, will likely see their work increasingly assisted by generative AI tools. Employees in fields that require deeper knowledge may not be impacted for some time, while those who perform physical labor may not see any changes in their work at all.

A separate report issued by Goldman Sachs—*The Potentially Large Effects of Artificial Intelligence on Economic Growth*[45]—predicts "significant disruption" on the horizon for the labor market. The report opines that two-thirds of jobs are capable of being automated to some degree. According to the authors, that translates to up to 300 million jobs worldwide. The authors also noted that the use of AI technology has the potential to increase labor productivity and augment global GDP up to 7% over time.

Not surprisingly, Goldman Sachs predicts that in the United States, 46% of tasks performed in administrative support positions are capable of being automated. No doubt, firms have been automating administrative tasks for the past several decades—word processing departments and typing pools are extinct, and tasks like completing expenses are far less labor intensive. We can imagine a world where our administrative questions are frequently resolved through interfacing with a chatbot powered by generative AI instead of searching a directory for the right person to call who just may have that knowledge at their fingertips.

Closely following administrative support, Goldman Sachs predicts that 46% of the tasks legal professionals perform are capable of being automated. Apart from generating discovery, research, and briefs in the litigation context, generative AI will aid tasks like contract and document review and analysis, form creation, and compliance analyses.[46] Drawing on centuries of recorded documents and decisions, it is not hard to imagine large language models being able to assimilate and synthesize the available data more quickly and efficiently than a human.

Generative AI will also impact other professions in the knowledge work and creative space. Goldman Sachs predicts that 37% of tasks performed in architecture and engineering, 36% of tasks in physical and life sciences, and 35% of tasks in business and finance are capable of being automated. Generative AI is less likely to have a significant impact on food service, custodial services, maintenance, and skilled trades—although these jobs are already susceptible to displacement by automation (think of in-store automated cleaners that replace human janitors, or the automation of the quick-serve food industry in both ordering and food preparation).

B. Replacement Versus Augmentation

OpenAI believes that generative AI will be even more prolific than Goldman Sach's prediction. According to a paper published by a team of researchers from OpenAI, OpenResearch, and the University of Pennsylvania, approximately 80% of the U.S. workforce could experience automation of some of their tasks through generative AI. That paper also predicts that for 19% of the workforce, up to 50% of their tasks could be automated.[47]

---

44   *Id.*

45   Jan Hatzius et al., *The Potentially Large Effects of Artificial Intelligence on Economic Grown (Briggs/Kodnani)*, Goldman Sachs Economics Research (Mar. 26, 2023).

46   *See* Andrew Perlman, *The Implications of ChatGPT for Legal Services and Society*, Harvard Law School Center on the Legal Profession (Mar./Apr. 2023).

47   Hatzius et al., *supra* note 45.

What remains unclear is the extent to which generative AI will displace, versus augment, human labor, and when. Of course, with improved efficiency, there is a likelihood that society will require fewer workers to perform a task. But, according to the World Economic Forum, "[w]hile Fourth Industrial Revolution technologies driven by AI will continue to fundamentally change the world and the way we work and live, AI may not lead to massive unemployment. Instead, AI technology will create more jobs than it automates."[48] In this way, repetitive, dull, and/or dangerous tasks are eliminated, allowing workers to focus on higher-value tasks. While there is tremendous upside, that will not arrive without some pain, certain tasks, will, of course, be eliminated, and different skills will be required to navigate an evolving workplace.

C. Economic Impact & Need for New Skills

Generative AI is a force of nature. Leaders should take the time now to evaluate where their businesses will experience disruption. As generative AI provides opportunity for workers to be more strategic, efforts should be made to assist employees in developing strategic thinking skills. Thoughtful businesses will anticipate the technical skills needed to thrive and will train and hire appropriately.[49]

Popular attention on the impact of automation and generative AI on the labor market tends to focus on the destruction of jobs in industries experiencing automation. This is, to be sure, a very real effect of artificial intelligence. But automation, and the technological progress that drives it, can have strong positive impacts on the workforce as well by creating and increasing demand for workers with certain skills. Most immediately and directly, the increasing ubiquity of automated systems in people's daily lives—both at work and at home—will greatly increase demand for workers performing services related to those systems, which will require regular maintenance and periodic repair to keep the complex electronic and digital systems that power them in working order. There also will be increased demand for workers trained to build and test autonomous systems.

The U.S. workforce is currently experiencing an historic shortage of labor generally, particularly in the skilled workforce, where job openings requiring specialized training or advanced knowledge far outstrip the number of qualified applicants. This trend was only exacerbated by the COVID-19 pandemic, and the "great resignation" of millions of workers who left the U.S. workforce. Indeed, by one estimate, by 2030, the talent shortage is expected to cost $162 billion in unrealized revenue in the U.S. workforce alone.[50]

This convergence—the advent of automation and generative AI, and the increasing shortage of skilled workers—has led many to call for new or novel workforce development policies. We increasingly see a shift in focus for employers and policymakers from valuing "credentials" to valuing skills and core competencies in hiring. Our workforce training and development models must be improved, not only to train workers on how to adapt to generative AI in the workplace, but more broadly to develop and hone new skills increasingly required as lower-skilled jobs are replaced by automation.

In the past, the federal government has adopted programs such as the Trade Adjustment Assistance program, which generally provides aid to workers whose jobs have been displaced by trade agreements or have been offshored. A comparable effort may be advisable with respect to generative AI and automation more broadly. Similarly, policymakers may wish to use tax benefits and other incentives to spur employers to partner with educational institutions (notably

---

48   *See* Mohamed Kande & Mural Conmez, *Don't fear AI. It will lead to long-term job growth*., World Economic Forum (Oct. 26, 2020).

49   *See Commission on Artificial Intelligence Competitiveness, Inclusion, and Innovation*, U.S. Chamber of Commerce Technology Engagement Center (2023).

50   *See* Laurent Probst & Christian Scharf, *The Lost Workforce*, PWC World Government Summit (2019).

community colleges and vocational education programs) to ensure that workers have access to the skilling and reskilling resources that they need.

D. Emerging Legal Issues

As the emergence of generative AI creates new opportunities—and potential perils—for the American workforce, so is it likely to create new questions under traditional labor and employment laws. Media reports around generative AI chatbots have highlighted how AI-generated conversations may go "off the rails," including an episode where the program claimed to have fallen in love with a user, and tried to convince him that his marriage was an unhappy one and that he should leave his wife in favor of the chatbot.[51] What happens in the workplace when a worker claims to have been sexually harassed by an AI program or claims to have been subject to inappropriate comments about their appearance, or sex, or race? Can an employer be held liable for the "speech" of a chatbot over which it may have no direct control? Or for allowing use of the tool in the first instance? What if an employee uses an HR chatbot to inquire about available leave, and discloses protected medical information while doing so? And does providing such information constitute a formal request for protected leave under federal or local law?

Similar questions arise with respect to organized labor regarding the use of generative AI in the workplace. Depending on the language of its collective bargaining agreement, is an employer obligated to bargain with the union over the decision to bring generative AI into the workplace? Perhaps, perhaps not. But if the use of generative AI is likely to lead to the loss of jobs for bargaining unit members, or otherwise significantly affect how workers do their jobs, the argument that an employer must engage in bargaining over the effects of that decision may become stronger. Conversely, what if an employer adopts a policy that prohibits the use of generative AI in the workplace? Is the employer within its rights to require that all work be performed by human workers, or is its policy hampering workers' productivity and efficiency and making their jobs more difficult?

For that matter, can generative AI or other assistive technology make it easier for workers to organize and form a union? Many would state conclusively that the omnipresence of non-AI social media has had a significant effect on how unions organize, from providing greater means for employees to communicate rapidly, to providing platforms for digital and video organizing material (think of unions turning to Tik Tok or viral videos). Can generative AI be used to develop more effective campaign materials, answer questions about union organizing without requiring human direction, or obtain more meaningful corporate information about a company and its operations to assist union efforts to organize employers, or to bargain with them? The answers to many of these questions are yet to be determined.

---

51   Kevin Roose, *A Conversation With Bing's Chatbot Left Me Deeply Unsettled*, N.Y. Times (Feb. 17, 2023).

E. Regulatory Efforts, and More to Come

1. State and Local Developments in the United States

Well before the emergence of ChatGPT and other generative AI tools, legislators in the U.S. and abroad had already begun acting to regulate the use of AI tools in the employment sector and beyond. While many of these incipient laws have yet to be implemented or shown to be feasible, the level of activity itself generates significant compliance risk for transnational and multinational employers.

For example, in 2020, both Illinois and Maryland adopted restrictions on the use of video evaluations using AI (including facial recognition tools), among other things, requiring that an employer provide notice to, and obtain consent from, applicants who will be subject to an AI tool.

In December 2021, New York City adopted the most comprehensive regulatory scheme governing the use of predictive AI in hiring and promotion decisions.[52] The law, which applies to employers and employment agencies alike, requires that: any algorithmic tool that reviews, selects, ranks, or eliminates candidates for employment or promotion must be subject to a number of limitations, including a requirement that any such tool undergo an annual, independent "bias audit," with a publicly available summary; employers provide each candidate (internal or external) with 10 business days' notice prior to being subject to the tool; the notice list the "job qualifications and characteristics" used by the tool to make its assessment; the sources and types of data used by the tool, as well as the applicable data-retention policy, be made available publicly (or upon written request from the candidate); and candidates be able to opt out and request an alternative selection process or accommodation. In September 2022, the City published draft regulations that significantly narrowed the scope of the legislation and addressed some of the more onerous provisions of the law (*e.g.*, a 10-day notice requirement). On April 5, 2023, the City published final regulations, and will begin enforcing this law on July 5, 2023.[53]

On the other side of the country, in March 2022, the California Civil Rights Council (CRC, formerly the California Fair Employment & Housing Council) released draft revisions to the state's employment non-discrimination laws that would dramatically expand the liability exposure and obligations of employers and third-party vendors that use, sell, or administer machine-learning or other AI-driven employment-screening tools or services for employment decision-making.[54] Simultaneously, several bills aiming to create brand-new laws and regulations concerning the use of AI (in the employment lifecycle and more generally) are wending their way through the state's legislature. Following the lead of these two early movers, numerous other jurisdictions across the nation, including the District of Columbia, Illinois, Massachusetts, New Jersey, New York State, Vermont, and Washington State, have begun work on their own AI-regulating laws.[55]

---

52   *See* Eli Z. Freedberg, Niloy Ray, & Jim Paretti, *New York City Enacts Law that Hinders Use of Automated Tools in Hiring and Promotion Decisions*, Littler Insight (Dec. 28, 2021); Jim Paretti, Niloy Ray, Eli Freedberg, & Ellie McPike, *New York City Adopts Final Regulations on Use of AI in Hiring and Promotion, Extends Enforcement Date to July 5, 2023*, Littler Insight (Apr. 13, 2023).

53   *See* Niloy Ray & Jim Paretti, *New York City Proposes Regulations to Clarify Requirements for Using Automated Employment Decision Tools*, Littler ASAP (Sept. 23, 2022).

54   *See* California Civil Rights Council, *Proposed Modifications to Employment Regulations Regarding Automated-Decision Systems, Attachment* C (version Feb. 10, 2023).

55   *See, e.g.*, District of Columbia *B 114*, Illinois *HB 1811*, Illinois *HB 3773*, Massachusetts *HB 4514*, New Jersey *AB 4909*, New Jersey *SB 1926*, New York *AB 567*, New York *SB 5641*, Vermont *HB 114*, and Washington *SB 5116*.

2. U.S. Federal Agency Activity

Numerous federal agencies have declared their strong interest in regulating the use of AI in the workplace, as best evidenced by the "Joint Statement on Enforcement" released by the EEOC, the CFPB, the FTC, and the DOJ, on April 25, 2023, in which these agencies "pledge to vigorously use [their] collective authorities to protect individuals' rights regardless of whether legal violations occur through traditional means or advanced technologies."[56]

a. Office of Federal Contract Compliance Programs (OFCCP)

In November 2022, the OFCCP published proposed revisions to its Scheduling Letter and the accompanying Itemized Listing, which the agency sends to covered federal contractors as part of its compliance audit process.[57] Of note, the proposal includes a new item: "Identify and provide information and documentation of policies, practices, or systems used to recruit, screen, and hire, including the use of artificial intelligence, algorithms, automated systems or other technology-based selection procedures."[58] According to the OFCCP, the impetus for this change is that contractors are increasingly adopting automated technologies in their hiring and recruiting practices, which could lead to instances of screening or selection bias. "Addition of this requirement will allow OFCCP to assess the contractor's use of such technology to determine whether these tools are creating barriers to equal employment opportunity."[59]

b. Equal Employment Opportunity Commission (EEOC)

In October 2021, the EEOC launched an initiative relating to the use of AI in employment decision-making.[60] As stated by the EEOC, the initiative is intended to examine how technology impacts the way employment decisions are made, and give applicants, employees, employers, and technology vendors guidance to ensure that these technologies are used lawfully under federal equal employment opportunity laws. In its rollout of the initiative, the EEOC indicated that the agency plans to establish an internal working group to coordinate the agency's work on the initiative; launch a series of listening sessions with key stakeholders about algorithmic tools and their employment ramifications; gather information about the adoption, design, and impact of hiring and other employment-related technologies; identify promising practices; and issue technical assistance to provide guidance on algorithmic fairness and the use of AI in employment decisions.

The EEOC's initiative further underscores the interest in such systems, and that care must be taken when deploying such systems to avoid running afoul of anti-discrimination laws.[61] Based on the announcement on March 28, 2022 of its FY 2021 Performance Report and FY 2023 Budget Justification, EEOC Chair Charlotte Burrows underscored that the proposed

---

56   EEOC, Press Release, *EEOC Chair Burrows Joins DOJ, CFPB, And FTC Officials to Release Joint Statement on Artificial Intelligence (AI) and Automated Systems* (Apr. 25, 2023).

57   OFCCP, *Supply and Service Program; Proposed Approval of Information Collection Requirements; Comment Request*, 87 Fed. Reg. 70867 (Nov. 21, 2022).

58   DOL OFCCP – Supply and Service Program, OMB No. 1250-0003, *Supporting Statement* (Nov. 2022); OMB, Supply and Service Program –*Scheduling Letter Burden*.

59   *Id.* p. 15.

60   EEOC, Press Release, *EEOC Launches Initiative on Artificial Intelligence and Algorithmic Fairness* (Oct. 28, 2021).

61   Of importance to global employers, the European Union has published a proposed regulation aimed at creating a regulatory framework for the use of AI. In the hierarchy discussed in the EU's proposed regulation, AI systems that are implemented in the recruitment and management of talent should be classified as high-risk. Given that classification, among other requirements, employers or vendors using such systems would be required to develop a risk management system, maintain technical documentation, adopt appropriate data governance measures, meet transparency requirements, maintain human oversight, and meet registration requirements. It remains to be seen how U.S. regulators may utilize the tenets set forth in the EU's proposed regulation.

budget would be used, in relevant part, to advance the EEOC initiative launched in 2021 "to ensure that employment-related artificial intelligence and algorithmic decision-making tools comply with federal civil rights laws."[62]

In May 2022, the agency published "technical assistance"[63] relating to compliance with Americans with Disabilities Act (ADA) requirements when using AI and other software to hire and assess employees. The agency also published a short "Tips for Workers"[64] summary of this guidance. Neither of these documents has the force or effect of law, nor are they binding on employers; as the accompanying press release notes, this guidance is meant to be educational, "so that people with disabilities know their rights and employers can take action to avoid discrimination."[65] This technical assistance document focused primarily on: ADA requirements with respect to accessibility to individuals with disabilities; the need for reasonable accommodation to potentially accommodate individuals with disabilities in the use of these tools, and in the assessment of such individuals these tools might make; and the important distinctions regarding testing, design, and application of these tools in the ADA context from other non-discrimination laws.

In the interim, the EEOC continues to highlight the fact that AI is on the agency's radar screen as employers consider potential reliance on AI in hiring and other employment decisions. Most recently, on January 31, 2023, the Commission held a public hearing examining the implications of AI and machine learning in employment decisions, entitled "Navigating Employment Discrimination in AI and Automated Systems: A New Civil Rights Frontier." At that hearing the Commission heard testimony from a range of stakeholders, including academics, representatives of employers, privacy advocates, and others. Notably excluded from the witness list were: (a) any actual employer using AI tools in practice; and (b) the vendors or creators of AI employment tools. This absence was noted by both Republican commissioners.

The meeting focused heavily on the rapid increase in the use of AI in the workplace. Both Democratic commissioners and several witnesses repeatedly expressed concern that, depending on the data on which an algorithmic tool is based, these tools might perpetuate existing patterns of bias in the workplace. Consumer and privacy advocates stressed their view that "without guardrails" data-driven technology is likely to cause harm in the workplace, or, as one witness claimed, "inevitably lead to disparities." Others noted that even where an algorithm is shown to be highly predictive based on correlation (for example, a tool determining that candidates who preferred a certain hobby would be more successful employees), correlation itself is insufficient, and the agency should require a showing of causation as well. Several witnesses focused on the "structural bias" that may be contained in existing data sets (such as credit reports or arrest and conviction records) and urged the EEOC to take the position that "de-biasing" an algorithm, even where decisions to do so are based on race or other protected characteristics, is lawful under civil rights laws.

Supporters of AI technology, who were somewhat underrepresented at the hearing, stressed the value of AI in eliminating bias in subjective decision-making. Many called for requirements that the use of AI be prominently disclosed to workers and applicants, and that alternative methods of evaluation should be required where requested. Finally, there seemed to be significant consensus that AI tools be subject to audit requirements to ensure they are non-biased – although few offered specifics as to what these audits might look like, or how they might practically be conducted.

---

62 *See* EEOC, Press Release, *EEOC Releases Fiscal Year 2021 Performance Report and Fiscal Year 2023 Budget Justification* (Mar. 28, 2022); EEOC, *Artificial Intelligence and Algorithmic Fairness Initiative*.

63 EEOC, EEOC-NVTA-2022-2, *The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees* (May 12, 2022); see also Jim Paretti, Niloy Ray, & Marko Mrkonich, *EEOC Issues Guidance on Artificial Intelligence and Americans with Disabilities Act Considerations*, Littler Insight (May 18, 2022).

64 EEOC, *Tips for Workers: The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence* (May 12, 2022).

65 EEOC, Press Release, *U.S. EEOC and U.S. Department of Justice Warn against Disability Discrimination* (May 12, 2022).

c. Other Federal Activity

Other federal agencies have weighed in on this issue. In October 2022, National Labor Relations Board General Counsel Jennifer Abruzzo issued a memorandum urging the Board to adopt a new framework that seeks to hold employers accountable for use of what she calls "omnipresent surveillance and other algorithmic-management tools" if they tend to impair the exercise of Section 7 rights under the National Labor Relations Act.[66]

The Federal Trade Commission (FTC) has also been active in this space. In 2022, the FTC issued a report to Congress outlining its concerns about AI tools' potential inaccuracy and inherent design flaws, bias and discrimination, and commercial surveillance incentives, and provided a set of recommendations.[67] The report was in response to a directive in the 2021 Appropriations Act for the FTC to examine how AI "may be used to identify, remove, or take any other appropriate action necessary to address" various "online harms."[68]

Last year, the White House issued a Blueprint for an AI Bill of Rights to provide a framework for "an AI-powered world."[69] The blueprint "is a set of five principles and associated practices to help guide the design, use, and deployment of automated systems to protect the rights of the American public in the age of artificial intelligence."[70] These principles are:

- You should be protected from unsafe or ineffective systems.
- You should not face discrimination by algorithms and systems should be used and designed in an equitable way.
- You should be protected from abusive data practices via built-in protections and you should have agency over how data about you is used.
- You should know that an automated system is being used and understand how and why it contributes to outcomes that impact you.
- You should be able to opt out, where appropriate, and have access to a person who can quickly consider and remedy problems you encounter.

While these principles are not binding, their goals are likely to be incorporated into additional legislation and agency guidance.

Notably, on April 13, 2023, Senate Majority Leader Chuck Schumer (D-NY) launched an initiative to create a "flexible and resilient AI policy framework across the federal government that can adapt as the technology continues to advance, allowing for innovation and continued U.S. leadership in the development of this critical technology, while enhancing security, accountability, and transparency."[71] While this framework is still a work in progress, Sen. Schumer states it will:

> require companies to allow independent experts to review and test AI technologies ahead of a
> public release or update and give users access to those results. That disclosure is the foundation
> of the four guardrails: Who, Where, How, and Protect. The first three guardrails – Who, Where,

---

66   NLRB, General Counsel Memorandum 23-02, *Electronic Monitoring and Algorithmic Management of Employees Interfering with the Exercise of Section 7 Rights* (Oct. 31, 2022); see also James A. Paretti, Jr. Christopher R. Henderson, & Michelle L. Devlin, *NLRB General Counsel Calls for Board to Crack Down on Electronic Surveillance and Automated Management Practices*, Littler Insight (Nov. 3, 2022).

67   FTC, *Combatting Online Harms Through Innovation: A Report to Congress* (June 16, 2022).

68   *Id.*, citing American COMPETE Act, which directs the Commission to conduct a "Study to Combat Online Harms Through Innovation." Consolidated Appropriations Act, 2021, Pub. L. No. 116-260, Title XV, § 1501(j).

69   The White House, *Blueprint for an AI Bill of Rights* (Oct. 2022).

70   *Id.* at 4.

71   Senate Press Release, *Schumer Launches Major Effort To Get Ahead Of Artificial Intelligence* (Apr. 13, 2023).

and How – will inform users, give the government the data needed to properly regulate AI technology, and reduce potential harm. The final guardrail – Protect – will focus on aligning these systems with American values and ensuring that AI developers deliver on their promise to create a better world.[72]

3. International Efforts—EU & UK

The European Union (EU) is leading the way in the regulation of AI technology. A draft regulation ("AI Act") is working its way through the EU's legislative process and is expected to be finalized in the coming months. The first draft was adopted by the European Commission in April 2021 and the text lays down harmonized rules on how AI would be regulated across the 27 EU countries. At the time of writing, the most recent development was in December 2022, when the European Council made significant amendments to the draft text. An AI Liability Directive is being proposed in parallel, which would enable a claimant to bring a claim against either the developer of the AI system or the entity that has made use of it, where they have been caused loss by an AI system.

The objectives of the AI Act are to ensure that AI is used in a safe way and that the fundamental rights of individuals are protected, while also encouraging investment and innovation in AI. Non-binding guidelines for "Trustworthy Artificial Intelligence" have also been published, according to which, trustworthy AI must be lawful, ethical, and robust.[73]

a. The Content of the AI Act

The AI Act sets out a broad definition of AI, which would apply across Europe. At the time of writing, the definition is:

> *[A] system that is designed to operate with elements of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of objectives using machine-learning and/or logic and knowledge based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts.[74]*

This definition aims to give legal certainty, while also providing the flexibility to accommodate future technical developments. The legislation also sets out a distinction between the responsibilities of "providers" (*i.e.,* those that develop the AI system or place it on the market) and "users" (*i.e.,* organizations under whose authority the AI system is used, which would encompass employers).

---

72   *Id.*

73   European Commission, *Ethics guidelines for trustworthy AI* (Apr. 8, 2019).

74   Council of the European Union, *Interinstitutional File: 2021/0106(COD) Article 3(1)* (Nov. 11, 2022).

The AI Act takes a risk-based approach to the regulation of AI systems, where the obligations are proportionate to the level of risk that it poses. Put simply, the greater the risk posed by the AI, the stricter the rules.

The Act sets out a list of AI practices that pose an "unacceptable risk" and would therefore be <u>prohibited</u>. This includes: (i) AI that uses subliminal techniques that materially distort a person's behavior in a way that may cause harm; (ii) systems that exploit particular vulnerable groups and that may distort their behavior and cause harm; and (iii) social scoring that leads to certain types of detrimental or unfavorable treatment. In practice, it is unlikely that these uses of AI would apply in an employment context.

The Act also identifies a category of AI systems that are deemed to be "high-risk" and, therefore, subject to a range of detailed compliance requirements for both providers and users, with the majority of the obligations falling on providers. Of particular note to employers, the Recitals of the AI Act say that:

> *AI systems used in employment, workers management and access to self-employment, notably for the recruitment and selection of persons, for making decisions on promotion and termination and for task allocation based on individual behavior or personal traits or characteristics, monitoring or evaluation of persons in work-related contractual relationships, should also be classified as high-risk, since those systems may appreciably impact future career prospects and livelihoods of these persons.*

The mandatory compliance requirements for high-risk AI systems include detailed technical documentation, record keeping, requirements as to transparency, human oversight and ensuring the accuracy, robustness, and security of the AI system. This is likely to create an additional compliance burden for employers that deploy AI as part of their practices.

The draft regulation also envisages two lower-risk categories of AI systems—"limited risk" and "minimal risk"—which would be subject to less-burdensome transparency requirements with respect to the use of AI. These are less likely to be relevant to employers.

i.   Scope of the Act

In terms of territorial scope, as with the General Data Protection Regulation (GDPR – the European data protection regulation), the AI Act would have extraterritorial scope. It would relate to both AI systems established and operating in the EU, as well as systems established outside of the EU that affect EU individuals. As a result, U.S. companies that deploy AI in the EU would find themselves subject to the AI Act regardless of their geographic location.

ii.   Penalties for Non-compliance

As with the GDPR, it is the penalties for non-compliance that really put the sting in the tail. As currently proposed, the maximum penalties are up to the greater of EUR 30 million (USD 33 million) or 6% of global annual revenue.

iii.   What Next?

The next steps in the European legislative process involve a series of votes and negotiations before the final text is formally adopted.[75] Once the legislation has been finalized, the current draft of the legislation anticipates a three-year period for implementation (increased from two years in the most recent round of amendments). Once finalized, the AI Act would be a legally binding regulation, which must be applied in its entirety across the 27 EU member states.

---

75   European Parliament, Legislative Observatory, *Artificial Intelligence Act*, 2021/0106(COD).
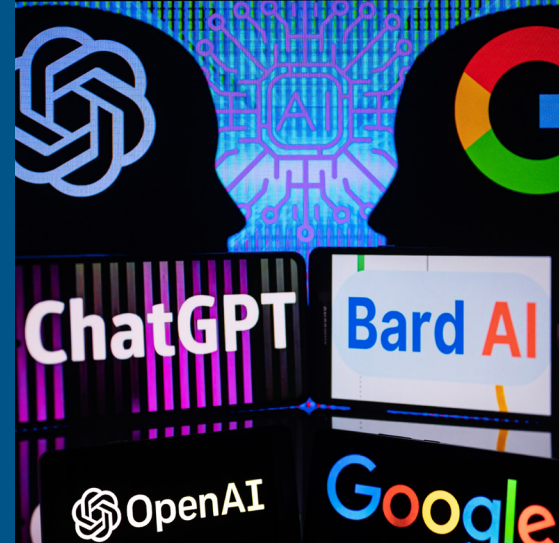
As the world's first comprehensive legal framework for AI, the AI Act is likely to be influential as other countries consider the regulation of AI. As with the GDPR, it may set the international standard for global tech companies in terms of AI compliance.

b. Meanwhile in the UK…

Although the UK is no longer part of the EU, in parallel it has started taking steps to regulate AI. The government published a call for evidence in November 2022, to which Littler contributed on behalf of the Employment Lawyers' Association.[76] This was followed by a White Paper in March 2023, which set out the proposed new approach to regulating AI, to build trust in cutting-edge technologies, and to make it easier for businesses to innovate, grow, and create jobs.[77]

In contrast to the EU approach, the UK government will avoid putting in place detailed legislation, which it says, "could stifle innovation" and that it does not propose to rush into legislating too early.[78] Instead, the government proposes to take an "adaptable" approach to regulating AI. The government is inviting feedback on the White Paper through a consultation that will run until June 21, 2023. It is also expected that the existing UK privacy regulator (the Information Commissioner's Office) will issue practical guidance to organizations over the next year, including toolkits and risk assessment templates, of how to use AI in line with principles set out in the White Paper (safety, security, robustness, transparency and explainability, fairness, accountability and governance, and contestability and redress).

It is clear that artificial intelligence broadly, and generative AI in particular, will raise significant new issues in the workplace and for the workforce—both domestic and international—ranging from its impact on the contours and demographics of the workforce currently and in the future; the challenge of applying laws that were adopted long before the advent of this technology to apply to a very different modern workforce; the rapidity of change and advancements in AI models, and the accordant difficulty of traditional, static regulation; and the need to ensure such tools and technologies are used in a manner consistent with existing workplace laws surrounding safety, non-discrimination, and other workplace protections.

Most of the regulatory efforts discussed in this section pre-date the emergence of generative AI tools in late 2022. We can expect to see additional regulatory efforts in response to these new tools.

---

76   Employment Lawyers Association, _ELA's response to the Call for Evidence: Governance of Artificial intelligence (AI)_ (Nov. 24, 2022).

77   UK Government, Press Release, _UK unveils world leading approach to innovation in first artificial intelligence white paper to turbocharge growth_ (Mar. 29, 2023).

78   UK Government, Policy Paper, _A pro-innovation approach to AI regulation_ (Mar. 29, 2023).

IV. Generative AI & HR Management of the Employment Lifecycle

The proliferation of easily accessible generative AI tools creates a fundamental concern for HR professionals, one that often can be reduced to the following simple question: was this work product created by a human or by an algorithm? The possibility that content creation may have been "sub-contracted" to a generative AI process creates a significant new need for greater scrutiny of all incoming and outgoing work product, to ensure that it is sufficiently reliable, *i.e.*, that it is authentic, accurate, and legally compliant.

A. Pre-Hire Stage

Generative AI challenges corporate notions of authenticity right from the job-application stage. Just as employers can adopt AI-based tools to oversee and verify its application review process, applicants can use generative AI tools to create resumes, writing samples, and other submissions, thereby perhaps masking the applicant's true drafting skills and abilities. Moreover, by feeding the position description to the generative AI algorithm, these submissions can be tailored to directly reflect and respond to the company's stated needs. While this sort of strategic drafting is not unique to generative AI, the ease of content generation afforded by these tools multiplies the likelihood of algorithm-generated—and possibly, misrepresentative or simply falsified—job applications being advanced, potentially at the expense of deserving candidates. Further along the hiring process, candidates might rely on generative AI tools during skill assessments, using these tools surreptitiously to determine accurate or suitable responses to technical and behavioral questions and tasks. Even if such manipulative use of generative AI is rooted out before the candidate is hired, the company's per-candidate cost rises.

B. During Employment

Once employed, an individual can potentially use ChatGPT to generate task-related content on a company's behalf, raising serious concerns (beyond the ethical) of plagiarism and of copyright violation. It might even impair the company's ability to obtain certain forms of protection for what it thought was its intellectual property. Generative AI tools are trained to repurpose the information on which they have been trained, and which resides within their data banks. This could lead to the use of others' work without payment or other form of credit.

Beyond these concerns of authenticity, there is also the issue of accuracy, discussed in greater detail in Section VII of this report. Generative AI tools may be trained on incomplete datasets and may not be regularly updated to incorporate new facts, occurrences, and knowledge.[79] Depending on the nature and use of this inaccurate work product, the ensuing risk to a company could range from negative market perception to lawsuits alleging material misrepresentation, breaches of warranty or contract, or even fraud. From HR's perspective, this breadth of potential legal exposure means tighter workforce management, enhanced training, stronger monitoring of technology use, and the development and refinement of AI use policies and practices organization-wide.

HR needs to be vigilant for the facial impact of generative AI on its candidate- and employee-engaging processes (as well as behind-the-scenes assessment processes using generative AI), particularly the potential for systemic bias and legally impermissible discriminatory outcomes. As previously discussed, generative AI and other machine-learning algorithms can fall short of the goal of being legally compliant with respect to anti-discrimination laws, resulting in legally cognizable adverse impact on members of a protected class. This could occur as early as the hiring process, where a

---

79   For instance, as of this draft, ChatGPT has only been trained on information from September 2021 or earlier, and thus has no knowledge of events from 2022 or 2023. Asking it for a list of Asian-American Oscar winners, to take a simple example, would hence be underinclusive and incorrect.

job description drafted using generative AI tools might be skewed toward attracting male rather than female candidates; or where a generative AI-driven chatbot steers men to certain jobs and women to other, lower-paid roles; or where generative AI-developed assessment questions for candidates tend to discourage or otherwise disfavor a protected class. Growing corporate reliance on generative AI could mean these discriminatory concerns persist post-hiring: an AI-based service chatbot may classify and direct employees to different benefits and resources based on protected characteristics; performance evaluations may rely on biased summaries of work product and achievement created using generative AI tools; and generative AI-built communications to employees may result in inappropriate workplace interactions and outcomes.

C. Use of Predictive AI Tools

More broadly, HR should watch for the discriminatory impact of all AI tools, particularly predictive AI tools, that it deploys. While generative AI focuses on content creation, predictive AI is meant (as the name implies) to make inferences or predictions from data inputs. Predictive AI tools, which, just like generative AI tools, are built using machine-learning algorithms and are susceptible to the same systemic biases as generative AI, are being deployed widely across the employment lifecycle.

Predictive AI is, for instance, being used for ad targeting: analyzing browsing data and patterns to determine the audience for a particular recruiting ad or the "ideal" subset of Netizens from which to attract or source applicants.

Predictive AI is being used to screen candidates: deciding who best matches the needs of an open position; classifying or scoring applicants to determine which of the applicants is "best"; or otherwise selecting those who advance along the hiring funnel and those who are never afforded the opportunity of human review.

Predictive AI is also routinely used to assess the physical and mental skills, aptitude, and characteristics of applicants and employees, whether through so-called gamified and other interactive testing or through more straightforward data analysis. These assessments can stretch to predictions of an employee's engagement, performance, longevity, and likelihood of job success, leading into decisions of whom to retain, whom to promote, and whom to let go.

Each of these uses could result in discriminatory outcomes, with adverse impact to protected groups. An ad-targeting algorithm may suggest lower-paying jobs to members of certain ethnic groups, or insufficiently include them in its "ideal"

sourcing subset. A screening algorithm could "teach" itself to identify (and then more favorably assess) men over women. An algorithmic skill-assessment tool could develop bias against particular facial features, accents, or verbal expressions. All of this can build into the same discriminatory treatment and outcomes that these tools are intended to eliminate, because in the absence of human testing and calibration, predictive AI tools can develop the same systemic biases that are reflected in human history and processes.

The potential intersection of predictive AI and generative AI may amplify these biased outcomes and create exponentially more complicated issues. Current-generation predictive AI tools are somewhat rigid in their processes, in terms of the data they consider and the form of the output they generate. For instance, a resume-reviewing AI tool will consistently provide a grade, score, or classification based on the set of characteristics that it has been trained to identify and measure. However, were a "conversational" generative AI-driven interface be added to the tool, it might permit a user to ask specific—and different—evaluative questions from one set of resumes to the next or permit differences in evaluative criteria from one user to the next, or in other ways change, on a more micro level, the assessments being made by the tool. Simultaneously, by creating a conversational interaction, this sort of next-generation tool would increase the user-friendliness and thus the reach of the underlying predictive AI tool, making it easier to use and therefore more frequently used. All of this would add up to greater risk of adverse impact and a greater complexity to account for or avoid.

D. Challenges for HR

To guard against these risks, companies will need to build an overall AI risk management approach, as described in Section VIII of this report. Determining appropriate and effective testing frameworks (and legal standards) for AI tools will be critical. Section II, above, explains that machine-learning models are not built on a cause-and-effect structure—these models identify characteristics that correlate with successful outcomes, but do not attempt to explain why those characteristics cause successful outcomes. However, existing legal standards (and the testing frameworks based on those standards), are built to address human biases, which tend to be causal. The existing standards and frameworks typically attempt to seek both identification (what characteristics correlate with the likelihood of an outcome) and explanation (how those characteristics operate to cause that outcome).

Machine-learning algorithms, however, are not designed to provide that second answer, as the reason for correlation plays no part in the algorithms' decision-making. In fact, depending on the extent to which a model is continuously training and refining its decisions, even that first answer—what characteristics the model considers, and what weight it gives to each— may not be static or easily discerned. This misalignment between the legislators' understanding of the technology and how the technology actually works has led to flux and uncertainty in the legal and testing communities, with new and vastly varied laws and approaches (oftentimes infeasible for the policy purpose at hand) being proposed regularly and frequently.

The challenge for HR is to ensure compliance with the various emerging regulatory requirements while determining what manner and level of AI tool testing and scrutiny suffices to ensure that discriminatory outcomes will not occur (or be held to have occurred).

The use of generative AI and predictive AI tools can also activate ADA concerns. Whenever such a tool is deployed, be it for candidate assessment, employee engagement, or even as part of a job function, HR must be careful to ensure that the tool does not unfairly deny access to, screen out, or otherwise restrict the fair participation of, an individual with a disability. Absent reasonable accommodation, these uses of AI tools may draw both state[80] and federal scrutiny.[81]

At the same time, employers could also face charges that *not* allowing use of a generative AI tool violates the ADA. For example, employees with a learning disorder such as dyslexia may benefit from using ChatGPT to help perform certain tasks. Whether an employer must offer the use of such a tool as a reasonable accommodation under the ADA has not been tested in court and would necessarily depend on a variety of case-specific factors. But employers should anticipate this possibility.

AI tools, both predictive and generative, represent the next level of data-driven HR processes—broadly eliminating the need for naturally biased human decision-making, deepening our analysis of candidates and employees, and increasing workflow speed and efficiency by orders of magnitude. To deliver on that three-pronged promise, these tools must engender confidence that their outcomes are in fact less discriminatory and biased than the human activity and decision-making they replace, rather than scaled-up versions of the same. While the marketplace and regulators work out a set of frameworks and standards to effectively create that confidence, HR and their partner stakeholders must for a time navigate the risk of unclear expectations and uncertain requirements.

---

80   *See*, *e.g.*, Illinois's Artificial Intelligence Video Interview Act (AIVIA), 820 Ill. Comp. Stat. 42.

81   Notably, the EEOC has indicated the application of the ADA to AI Tools is a high-priority enforcement goal. See EEOC, Technical Assistance Document, *The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees* (May 12, 2022); see also *EEOC's Draft Strategic Enforcement Plan*, 88 Fed. Reg, at 1379 (Jan. 10, 2023).

V. Generative AI & Data Privacy

The recent explosion of generative AI tools coincides with a parallel explosion in privacy legislation, both in the United States and globally. In the United States, in less than three years, six states have passed comprehensive data protection laws. Meanwhile, dozens of similar bills await their turn at the state and federal levels. Globally, many developed countries have passed new or stricter privacy laws within the last decade. Many of these laws explicitly regulate the application of AI, and all of them regulate the use of these tools in some way. Consequently, using personal data to train generative AI tools and the incorporation of personal data in its outputs entails navigating a thicket of data protection obligations.

This section covers the privacy issues for employers that use generative AI systems to process personal data.

As described in more detail in Section VIII, the risks for an employer using generative AI tools depends on the nature of the use and the role the employer plays. There are several possible ways a company's data may be stored or processed by an AI platform.

At a minimum, the prompt (or query) to the system will be used to process the query and return a response. In some systems this prompt (and any data provided for analysis) will not be retained. In others, the data may be retained and processed further for quality control, to detect potential misuse of the AI, or to evaluate the performance of the tool. Some providers may retain the data and use it to further train the model or improve it.

With some platforms, a company can also submit their own data in bulk for analysis. With that data available to the AI platform, questions can be asked about that data using the capabilities of the pre-trained model. The company's data can also be used to train a custom AI or to fine tune how the base model analyzes data and generates responses.

Each of these different ways of interacting with a generative AI tool carries different privacy risks, which are described below.

A. Lawful Basis

Most non-U.S. data protection laws permit the collection and processing of personal data on limited grounds only, such as the individual's consent or as required by law. As a result, in countries with these data protection laws, the use of personal data for a new purpose, such as analysis with an AI tool, should be reviewed for compliance with these laws.

Other countries, such as Brazil and members of the EU, permit data processing based on the company's "legitimate interests" to the extent that the individual's rights and freedoms do not outweigh the company's legitimate interests.[82] It is not yet clear, however, how regulatory authorities will address the use of personal data in the context of generative AI tools. EU regulators, for example, have publicly expressed concerns.[83] In fact, the Italian data protection authority recently banned a generative AI service from the country in part because of the lack of lawful basis for its data collection; however, the ban was lifted after the generative AI service announced a set of privacy controls while the probe of its compliance

---

82   Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 [hereinafter, "GDPR"], arts. 13, 14; Lei Geral de Proteção de Dados Pessoais, Lei n° 13.709/2018, Art. 7(IX).

83   Rolland, et al., "« Non, il ne faut pas interdire ChatGPT » (Jean-Noël Barrot, ministre de la Transition numérique)", L.A. Tribune, Apr. 6, 2023, available at *https://www.latribune.fr/technos-medias/informatique/non-il-ne-faut-pas-interdire-chatgpt-jean-noel-barrot-ministre-de-la-transition-numerique-958119.html.*

with data protection requirements continued.[84] The Spanish data protection authority, the AEPD, has followed Italy's lead and announced a preliminary investigation into OpenAI over suspected breaches of GDPR.[85]

In the United States, data protection laws generally have not adopted the concept of a lawful basis for processing personal data. This could change, however, as more states pass data protection laws.

De-identifying the data before submitting to a generative AI service can reduce the risk. In contrast to personal data, de-identified data is largely unregulated by privacy laws. However, most data protection regimes, such as the California Privacy Rights Act (CPRA), the Health Insurance Portability and Accountability Act (HIPAA), and the EU's GDPR, set a high standard for de-identification. Consequently, simply removing name and identification numbers is not necessarily sufficient to meet the standard for de-identification set by applicable law. Further, some laws, such as the CPRA, establish specific compliance requirements before using de-identified data. Therefore, before uploading de-identified data to any generative AI tool, HR personnel should confer with their legal team.

B. Notice

Virtually all data protection laws require the organization, collecting, and processing of personal data to provide notice to the data subject describing how it processes the personal data. These notices must be quite detailed, and must include descriptions of the purposes of use, disclosures, and information about exercising data rights.[86] Thus, before submitting personal data to a generative AI tool for analysis, companies should evaluate whether notices used to collect personal data addressed the use of the data in the context of a generative AI solution.

Moreover, in some circumstances, the notice must provide details about how the algorithm works. For example, under the GDPR, the employer must provide notice about a decision made solely based upon AI "which produces legal effects concerning him or her or similarly significantly affects him or her."[87] This likely would include decisions about hiring and termination of employment. In that case, the notice must provide "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject."[88] Given that most current AI tools maintain their algorithms as trade secrets, or the tool is not explainable or interpretable (see Section VIII below), an employer might not be able to provide "meaningful information about the logic involved" in using a generative AI tool. As a result, in the EU, employers should avoid making material employment decisions based solely on analysis from a generative AI tool. Of course, employers anywhere must consider the risks of basing decisions on an AI tool when they do not fully understand the algorithms used by the tool.

C. Cross-Border Data Transfers

Most data protection laws prohibit the transfer of personal data to another country, except in limited circumstances. As with any other technology, companies using generative AI tools should evaluate the flow of data and adopt lawful data transfer mechanisms to address this issue and avoid potential violations.

---

84 Announcement by Garante Privacy Italia, Artificial intelligence: stop to ChatGPT by the Italian SA, Mar. 31, 2023, *available* at *https:// www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9870847#english*; Natasha Lomas, *ChatGPT resumes service in Italy after adding privacy disclosures and controls*, TechCrunch (Apr. 28, 2023).

85 Reuters, *Spain asks EU data protection board to discuss OpenAI's ChatGPT* (Apr. 11, 2023).

86 GDPR, arts. 13, 14.

87 GDPR, arts. 13(2)(f), 22.

88 GDPR at art. 13(2)(f).

D. Right to Delete

Depending on the applicable data protection law, individuals may have rights to access, delete, correct, or stop the processing of their personal data. These rights pose particular challenges to generative AI tools. What if an employee demands the deletion of personal data that a company has submitted to an AI platform to develop a custom tool or to fine tune the AI? That could be a challenging problem for employers.

These rights concerning personal data come with specific limitations that vary by jurisdiction. For example, the right to delete is quite limited under California's CPRA, but much broader under the GDPR or Brazil's data protection law. To reduce the risk of compliance with these obligations, a company might consider using data sets to which the right to delete or object does not apply or applies minimally. For example, the CPRA's right to delete applies only to personal information "collected from" the individual and not to personal information generated by the employer.[89] Consequently, items provided by the employee, such as name, education, and address, might be subject to the right to delete (subject to other limitations on that right), but items created by the employer, such as payroll records, would be exempt from this right. For California employers, limiting the AI input data to data like payroll reports will result in less administrative burden, at least from the perspective of data rights.

Until these tools develop further, and we learn more about how they work, compliance with these obligations may be challenging. Additional legislative developments may also affect these issues.

E. Accuracy

As discussed in more detail in other sections, generative AI may produce inaccurate content for several reasons. The training data set, as well as source data, may be wrong. Also, the tool itself may interpret the training data set and source data incorrectly or simply make up information. Most data protection laws provide a right to correct personal data, at least to the extent that the personal data is inaccurate. Correcting data raises similar issues to deleting data.

Outside of the United States, data protection laws nearly universally require a data controller to ensure the accuracy of personal data.[90] An employer potentially could violate these laws by not vetting generative AI tools for accuracy or by relying on inaccurate reports generated by these tools. In the United States, the new data protection laws generally do not impose this accuracy obligation, though they do require organizations to correct inaccurate information upon request.

---

89   Cal. Civ. Code § 1798.105(a).

90   *See*, *e.g.*, GDPR, art. 5(1)(d).

VI. Generative AI & Intellectual Property

As discussed throughout this report, generative AI models require training data to learn and perform their tasks. Very large quantities of training data are required for high-quality output data. ChatGPT, DALL-E, and other generative AI models rely on the vast quantity of public data available on the internet, while most private models will rely primarily on private data sets.[91] The owners and users of these data sets will rely on trade secret and contract law remedies to enforce their respective ownership and user rights. The two primary issues facing employers under trade secret law are protection of the employer's own trade secrets and the risk of an employee using a third party's trade secrets in the course of their work for the employer. In addition, for employers that are developing their own internal AI systems, given the high-stakes competition for a limited talent pool in this field, these employers face the additional challenge of implementing safeguards with respect to their employee and contractor developers.

A. Employer's Protectable Trade Secrets

Under the Uniform Trade Secret Act (UTSA), enacted in most states, a trade secret is:

> *information, including a formula, pattern, compilation, program, device, method, technique, process, drawing, cost data or customer list that: (i) derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable by proper means by, other persons who can obtain economic value from its disclosure or use, and (ii) is the subject of efforts that are reasonable under the circumstances to maintain its secrecy.[92]*

The lax transfer of data by an employee or a supplier may result in the loss of trade secret protection and with it a potential for staggering economic losses. Once a trade secret is available to the public (as long as not through improper means), the trade secret loses its protection under the law and may be used by any of the employer's competitors or other third parties. Further, when a trade secret becomes available to all employees in a company, including those not in positions of trust and/or not subject to a nondisclosure agreement (NDA), it becomes more challenging for an employer to argue it took measures to maintain the secrecy of the trade secret. Therefore, if building an internal AI model, it will be imperative that the employer consider the confidentiality of the data to be included. It is also important for the employer to limit access to the AI system so that the employer's trade secret data does not become available to all employees, similar to the approach employers should take to restrict employees from having access to certain data through passwords or otherwise.

In addition, if building an internal AI model, the employer's developer of a data set may acquire the underlying information from data suppliers, customers, patients, or other sources authorizing the developer's use of that data. If the data was properly acquired and not readily ascertainable—as a whole—to third parties, the dataset is afforded trade secret protection notwithstanding the fact that any given piece of underlying data may be known to the source that provided it.[93] In sourcing such data, however, the employer's developer(s) must take care not to utilize the trade secrets belonging to any third party.

With respect to the data sets generally, if a company takes reasonable measures to protect the secrecy of its data, data sets are generally afforded trade secret protection as "compilations" under both the Uniform Trade Secrets Act and the federal Defend Trade Secrets Act.[94] Complex data sets will generally be afforded trade secret status because their developers

---

91   Richard MacManus, *Generative AI: How Companies Are Using and Scaling AI Models*, The New Stack (Jan. 25, 2023).

92   UTSA § 1.4.

93   *See, e.g.*, *Flatiron Health, Inc. v. Carson*, No. 19 CIV. 8999 (VM), 2020 WL 1643396, at *9 (S.D.N.Y. Apr. 1, 2020); *Healthcare Sales Enablement, Inc. v. Geyfman*, No. 19-CV-20125, 2019 WL 13167349, at *1 (S.D. Fla. Feb. 25, 2019).

94   *See* UTSA § 1.4; 18 U.S. Code § 1839(3).
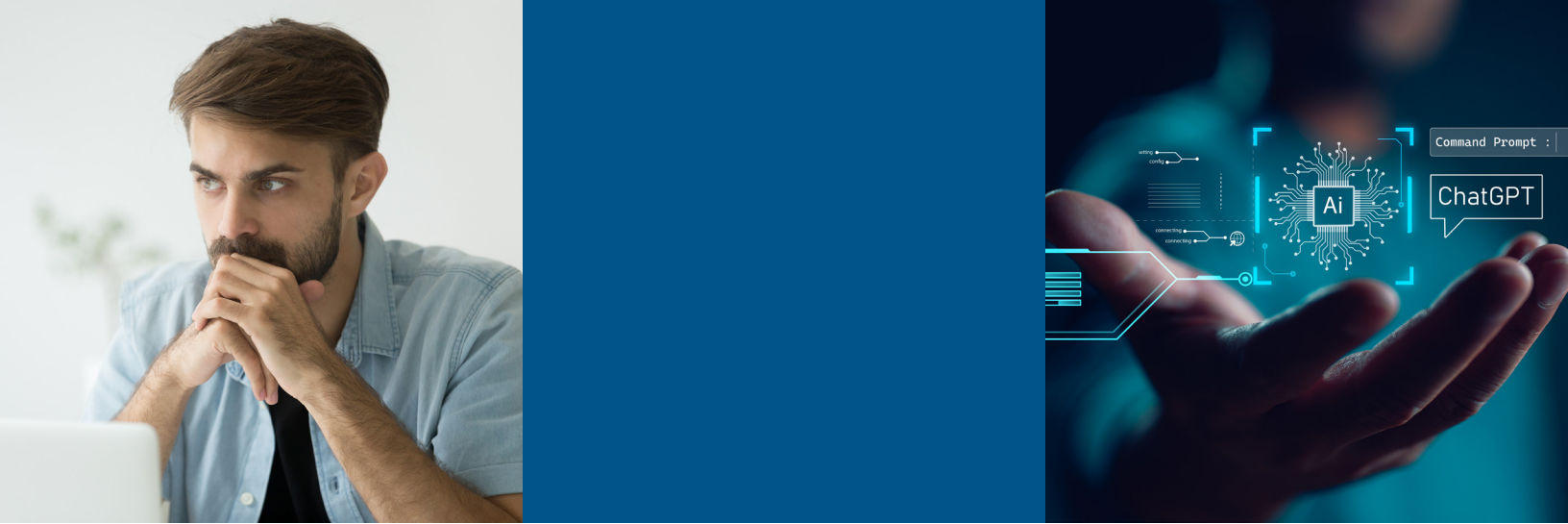
invested valuable resources in collecting, organizing, and compiling the information with the objective of deriving economic value therefrom.[95] The trade-secret status and economic value of a data compilation is only as secure as the measures taken to protect its secrecy.

B. Improper Use or Disclosure of a Third Party's Trade Secrets

The unauthorized use of proprietary data by an employer's AI developer or other employee may result in enormous exposure to liability for breach of contract or trade secret misappropriation, so employers should implement safeguards to minimize the risk of this occurring.

For example, if an employer allows employees to use an AI system in the course of their employment with the employer, or unbeknownst to them, the employees do so on their own, the employees could unknowingly utilize another company's trade secrets in the performance of their work for the employer. Employers may consider including contract provisions in their NDA to address an employee's potential use of an AI system during their employment.

Employers face another area of risk for misappropriation of a third party's trade secrets if they have internal AI developers developing an AI system who source data from outside the company, *e.g.*, from data suppliers. It would be prudent to contemplate requiring the developer (and any other employee utilizing the AI system) to sign an acknowledgment relating to the proper use of data available through the AI system, including addressing any contractual confidentiality obligations owed to the data supplier.

Unlike patents, discussed below, trade secrets do not automatically afford their owner a monopoly over their subject matter. Reverse engineering and other forms of independent development, for example, are expressly allowed. Employee mobility and the commercial need for disclosure (*e.g.*, to partners, customers, and independent contractors) impose substantial hurdles to trade secret protection for all technologies, including AI-based ones. Nonetheless, many companies have found that investment in robust NDAs combined with aggressive security measures afford them the most reliable protection against infringement and copying. While the up-front costs necessary to ensure protectability of trade secrets are not insubstantial, trade secret (and contractual confidentiality) law may still offer the best option for protecting non-customer-facing AI-assisted technologies and work product such as training data sets, AI prompts, and programming code.

To mitigate the risk of harm to an employer's own trade secrets or improper use of a third party's trade secrets, employers should consider implementing the following measures:

---

95   *See, e.g., KPM Analytics N. Am. Corp. v. Blue Sun Sci., L.L.C.*, No. 4:21-CV-10572-TSH, 2021 WL 2982866, at *13 (D. Mass. July 15, 2021).

- Restricting use of external hard drives and cloud-based accounts to those drives or accounts specifically issued or authorized by the employer.

- Limiting access on a project basis to prevent unauthorized access to third-party data by an individual who has not undertaken confidentiality obligations with respect to that data.

- Restricting use of personal devices to perform work for an employer and including in its own worker confidentiality agreements a right to a forensic audit of the worker's devices used in the course of their employment (to the extent permitted by applicable law), in the event of a suspected breach. Courts may order an employee's devices undergo forensics upon an adequate evidentiary showing of exfiltration, but an agreed-upon entitlement to this form of relief will increase the likelihood of obtaining it.

C. Other Intellectual Property Concerns

Generative AI also presents new and unique problems for companies trying to protect their own intellectual property rights and avoid infringement claims from others. As discussed above, trade secret law affords owners protection for confidential data, compilations, and processes that derive economic value from their secrecy. Trade secret law thus affords important protections for the non-public aspects of AI models, their inputs, and their outputs. However, trade secret protection does not obviate the necessity of considering other forms of intellectual property, particularly for user-facing outputs.

1.   Copyright Considerations

Copyright law, for example, is clearly implicated by the inputs and outputs of generative AI, as well as the source code for the algorithms of the technology itself. OpenAI's Terms and Conditions, read with its usage policies, make clear that both the inputs and outputs of the program belong to the user.[96] That is, OpenAI does not purport to own the rights to the data used to generate or generated by its AI program. This position is in accord with the law of most countries, which do not currently recognize non-human inventors or authors.

However, OpenAI also expressly disclaims liability for infringement of other parties' intellectual property rights arising from use of ChatGPT. Numerous copyright issues have arisen from the use of data-trained learning algorithms to generate outputs. For example, several lawsuits have accused companies of infringement for using copyrighted compilations as training data for AI algorithms. In *Planner 5D v. Facebook*, a Lithuanian company filed suit against Facebook and Princeton University, alleging copyright infringement based on defendants' alleged "scraping" of its library of 3-D renderings to use as training data for AI-powered robots programmed to recognize typical three-dimensional spaces.[97] According to the complaint, the defendants purloined over 5GB of data in violation of Planner 5D's express terms of use that prohibit "any deep link, page-scrape, robot, spider, or automatic device, program, or algorithm, or methodology which perform similar functions to access, acquire, copy or monitor any proton of the Planner5D project . . . ." Defendants have focused their defense strategy on attacking Planner5D's standing (for failing to register its work prior to filing suit) and the protectability of its data compilation (for lack of originality).

In another lawsuit, *Andersen v. Stability AI Ltd*, three artists sued AI art generators over their use of the artists' work in the data training set without their consent or compensation.[98] Plaintiffs allege that defendants engaged in direct copyright infringement by downloading copyrighted works from the internet and using the works to train their AI image generation

---

96   Open AI, *Terms of use* (updated Mar. 14, 2023); *Usage policies* (updated Mar. 23, 2023).

97   The original complaint included claims for trade secret misappropriation, but the re-filed lawsuit is based on copyright infringement alone. *Planner 5D v. Facebook*, 3:20-cv-08261, Dkt. 1 (N.D. Cal. Nov. 23, 2020).

98   *Andersen v. Stability AI Ltd.*, 3:23-cv-00201-WHO (N.D. Cal.).

models without a license and, further that they violated the Digital Millennium Copyright Act (DMCA) in doing so by removing copyright management information (*e.g.,* copyright notices, titles, author information). Plaintiffs also allege that defendants are liable for vicarious infringement for distributing programs that enable their users to create images "that can pass as original works by that artist," *i.e.,* "fakes."[99] Defendants have filed a variety of motions to dismiss and to strike that highlight the significant legal hurdles the plaintiffs' enforcement campaign face including the plaintiffs' inability to pinpoint specific copyrighted works used as training data or identify any particular outputs that are substantially similar to the copyrighted works. Defendants also rely on constitutional protections for free speech that protect "transformative" works that reflect significant expressive content beyond the plaintiffs' likeness. In short, Defendants rely on the nature of generative AI, which necessarily relies on "billions of images" to argue that the model cannot, as a matter of law, be held to infringe or violate any single artist's rights.[100]

The same law firm representing the artists brought a similar lawsuit, *Doe v. GitHub, Inc.,*[101] on behalf of software developers that accuses the software developer that effectively controls both GitHub's Codex and OpenAI's Copilot programs[102] of using code stored in GitHub repositories as training data for the AI-based system without attribution or consent. This case, unlike *Anderson*, expressly identifies outputs of the AI-generative process that are substantially similar to the data used to train the model.[103] Plaintiffs contend that the output from the accused programs is "often a near-identical reproduction of code from the training data"[104] except that it "excises the license terms, copyright notice, and attribution" data, causing "users to assume that the code can be used without restriction."[105] While both *Anderson and Doe* challenge the legal predicate espoused by defendants (and much of the open source community) that "training machine learning models on publicly available data is considered fair use,"[106] *Doe* does not assert copyright infringement claims but focuses on DMCA and contractual issues. Motions to dismiss are pending in that case as well.

Collectively, these lawsuits demonstrate the inter-relatedness of legal challenges to the inputs and outputs of generative AI programs. They also underscore the complexity of legal challenges to AI use that touch on fundamental, constitutional rights to free speech and long-held intellectual property tenants of fair use. While copyright suits against AI inputs and outputs proliferate in federal court dockets, none has yet to offer reliable substantive insight into the standards likely to be applied to AI inputs or outputs.

---

99   *Id.* at Dkt. 1 (Complaint) ¶ 171.

100 *Id.* at Dkt. 49, Dkt. 51.

101 *Doe v. GitHub*, Inc., 4:22-cv-06823-JST (N.D. Cal.).

102 According to the Complaint, Microsoft owns GitHub and is the exclusive licensee of OpenAI's GPT-3 language model. Dkt. 1 at ¶¶ 5, 6.

103 Users of CoPilot may be able to avoid some licensing issues by opting out of the "Suggestions matching public code" feature by "blocking" the feature in the CoPilot settings menu.

104 *Id.* at ¶ 46.

105 *Id.* at ¶ 77, 80 ("Codex and Copilot are not programmed to treat attribution, copyright notices, and license terms as legally essential. Defendants made a deliberate choice to expedite the release of Copilot rather than ensure it would not provide unlawful Ouptut.").

106 *Id.* at ¶ 83.

While these lawsuits and others like them will offer important insights into the risk of using generative AI tools, they do not directly address the question of ownership over AI-generated works—either the queries developed by users that are fed into the AI engine or the outputs thereof. While the United States Copyright Office (USCO) has already ruled that copyright protection cannot be afforded to works by non-human authors,[107] it is currently addressing the much more nuanced issue of the protectability of AI-assisted works of authorship. For example, Kris Kashtanova filed copyright applications to protect images in the graphic novel "Zarya of the Dawn." In a non-final order, the USCO rejected the copyright application on the basis that the work was AI-generated, despite Kashtanova's claim of ownership based on her AI prompts and modifications to the AI-output image.[108]

While the USCO appears to be leaning towards denying ownership of AI-generated works to the author of the AI prompt, most users assume that, at a minimum, the AI prompt itself is protected. However, while a sufficiently creative prompt can be copyrighted, courts have generally limited the scope of copyright protection for software scripts and other functional language. As an initial matter, the law distinguishes between the copyrighted language of the input (the works "expressive elements") and the underlying "idea, procedure, process, system, method of operation, concept, principle, or discovery" expressed therein, which is not subject to protection.[109] Various doctrines including the merger doctrine and scènes à faire have been used to preclude protection over any aspect of the work if the expression is inextricably intertwined with the idea, process, etc.

The relative "thinness" of protection for software and other programmatic works (such as SQL queries) is illustrated by the judicial decisions coming out of the decades-long legal dispute over one software company's right to prevent other companies from incorporating Java into its own programming language. There, the defendant company implemented common portions of Java code in its Android codex. The company argued that it used only the unprotectable elements of the application programming interface (such as registry names) necessary to allow its applications to interoperate with Java. After the defendant lost twice before the lower courts, in 2021, the U.S. Supreme Court held that the defendant's copying of the Java API was fair use. In doing so, the Supreme Court side-stepped the thorny issue of whether the copied portions could be copyrighted at all and focused instead on the practical considerations of interoperability and computability to justify its conclusion that the defendant company was not liable for infringement.[110]

While the ultimate outcome of these legal issues remains uncertain, they underscore the difficulty of adapting traditional copyright notions to the evolving world of generative AI. On one hand, generative AI imposes a substantial risk of possible copyright infringement liability on users and developers alike. On the other, it is equally clear that copyright law is unlikely to provide comprehensive protection for AI-assisted programs and their outputs. Relatedly, as OpenAI's Terms of Use make clear, reliance on generative AI creates a real risk that multiple "authors" relying on the tool will end up "owning" works that are substantially identical.[111]

2. Patents

Another source of potential protection for, and regulation of, AI-related intellectual property is patent law. While the U.S. Patent & Trademark Office (USPTO) has reviewed hundreds of thousands of AI-related patent applications and issued over

---

107 Previously, the USCO rejected an application that identified AI DABUS as the sole author of a work on the basis that copyright protection only extends to human authors.

108 U.S. Copyright Office, *Zarya of the Dawn Letter* (Feb. 21, 2023).

109 17 U.S.C. § 102.

110 *Google LLC v. Oracle America*, Inc., No. 18-956 (2020).

111 OpenAI, *Terms of use* (updated Mar. 14, 2023).

18,000 AI-related patents in 2022 alone, patent authorities recognize the unique challenges posed by AI to the traditional patent framework including "issues such as subject matter eligibility, disclosure requirements of AI inventions, and the implications of AI's increasing role in the inventive process."[112] For example, while the United States (as with all jurisdictions that have addressed the issue except South Africa and Australia) has held that AI cannot be an inventor on a U.S. patent, there is scant guidance on the patentability of AI-assisted inventions.[113] Recognizing the need for further policy regarding AI-assisted inventions, the USPTO recently issued a request for comments regarding AI and inventorship.[114]

Even AI-related inventions with human inventors are challenged to meet the legal requirements of 35 U.S.C. §§ 101 and 112. Section 101 forecloses patentability of abstract ideas, mathematical equations, and natural phenomena. This rule against abstraction imposes substantial hurdles for software-related patents. As the law allows patenting of an implementation and not the entire abstract idea, it is difficult to draft patentable claims for an algorithm that, by its nature, evolves over time.[115] And courts have found that recitation of "well-established" AI arrangements such as a "neural network" are abstract concepts that cannot be patented.[116] The evolutionary nature of AI technologies also poses challenges for the written description and enablement requirements of Section 112. In many cases, the human inventors of the AI do not know (and cannot predict) the specific ways in which an AI algorithm will evolve or the reasons it does so. As Section 112 requires a sufficient description of the software algorithm to support validity, an algorithm that changes in unpredictable ways poses unique hurdles for patentability. While some initial guidance from the Patent Trial and Appeal Board and other judicial authorities has attempted to provide a pathway to patent protection for AI inventors, companies making substantial time and financial investments in AI-assisted technology must accept the risk that their inventions are unpatentable.[117]

---

112 Kathi Vidal, *Incentivizing and protecting innovation in artificial intelligence and emerging technologies*, Director's Blog: the latest from USPTO leadership, U.S. Patent and Trademark Office (June 6, 2022).

113 *Thaler v. Vidal*, No. 2021-2347 (Fed. Cir. 2022).

114 U.S. Patent and Trademark Office, Dept. of Commerce, *Request for Comments Regarding Artificial Intelligence and Inventorship*, 88 Fed. Reg. 9492-9495 (Feb. 14, 2023).

115 *Hyper Search*, L.L.C. v. *Facebook, Inc.*, No. CV 17-1387-CFC-SRF, 2018 WL 6617143, at *10 (D. Del. Dec. 17, 2018).

116 *Id.*

117 *Request for Comments Regarding Artificial Intelligence and Inventorship, supra* note 114.

VII. Generative AI & Accuracy Concerns

Various artificial intelligence technologies have become reliable, even integral, tools that enhance business operations ranging from content creation and data processing to supply chain maintenance and janitorial services. The success of an AI-reliant business, however, depends as much on the management of its risks as on the efficacy of its implementation.

As discussed throughout this report, any business that elects to use or create generative AI tools should carefully evaluate the legal requirements and implications, which will involve assessment of employment, product, privacy, intellectual property, contract, and regulatory laws—just to name a few. As with many technological innovations throughout history, artificial intelligence presents unique challenges to our legal system that will take years, if not decades, for our judicial and legislative branches to resolve. As we wait for the law to evolve, companies must resort to private agreements and policies to address issues of consent, fault-allocation, and ownership over AI-related technologies, their applications, and outputs.

A. Allocation of Fault for AI-Generated Misinformation

As previously discussed, some sophisticated generative AI tools, including ChatGPT, can be used to generate text and image content in seconds that would take human authors hours or days to create.[118] ChatGPT, for example, can be prompted with parameters relating to length, subject matter, tone, and target audience, among other things. The resulting work product often appears highly usable, devoid of obvious grammatical errors, and supported by data-driven analysis with research citations. Despite the polish of the AI work product, at least current generative AI tools, are prone to factual error.[119] AI developers characterize the errors as AI "hallucinations," and while some inaccuracies may be the product of coding flaws, others are simply inherent in the inferential learning model and limitations of the data sets used to train the algorithms.[120] In other words, generative AI sometimes gets it wrong.

Publication of factually inaccurate information may violate local and federal laws including defamation laws, unfair and deceptive trade practices statutes, and hate speech prohibitions. In many instances, both the original author of the inaccurate work as well as those who re-publish the false statements may be held civilly, and sometimes even criminally, liable. While authorities are still grappling with the specific nuances of fault allocation among the creators of AI tools and their users, it is unlikely that companies will be able to evade liability entirely for publishing false information even if seemingly reliable generative AI tools were used. The law has long recognized the requirement of reasonable fact checking and source verification and refused to allow publishers of defamatory content to avoid liability by attributing the false statements to others.[121] In fact, the FTC recently announced it planned to target companies that misuse AI for deceptive purposes.[122]

On the other hand, the law has reasonableness limitations that companies, particularly technology companies, have successfully employed to justify publication of allegedly defamatory content. For example, in 2016, Yelp successfully

---

118   Rob Locascio, *I tried using ChatCPT to write this article*, Fast Company (Feb. 4, 2023).

119   Mark Esposito, Terence Tse, &Tahereh Saheb, *Dilemmas of ChatGPT in Content Creation Industry*, Cal. Review Mgmt. (Mar. 6, 2023) ("Even the most advanced AI algorithm is still demonstrating some uncontrollable and erroneous behavior, resulting in outputs that are questionable and inaccurate.").

120   Vilius Petkauskas, *ChatGPT's answers could be nothing but a hallucination*, Cybernews (Mar. 6, 2023); Tom Simonite, *AI Has a Hallucination Problem That's Proving Tough to Fix*, Wired.com (Mar. 9, 2018).

121   *See, e.g., Gertz v. Robert Welch*, Inc., 418 U.S. 323, 346-47 (1974).

122   *See US FTC leaders will target AI that violates civil rights or is deceptive*, Reuters (Apr. 18, 2023); Michael Atleson, *Keep your AI claims in check*, FTC Business Blog (Feb. 27, 2023).

defeated claims that would have held it liable for its aggregation of user ratings[123] and, in 2013, Trip Advisor defeated a claim that its list of America's "Dirtiest Hotels" compiled from user reviews constituted defamation. A part of the companies' defenses rested on the implementation of reasonable measures to catch and remove false content from their postings.[124] These decisions, however, turned largely on the courts' determination that the websites did not "create or develop" the allegedly false content. The courts held, relying on long-established precedent, that passive aggregation of user content that does not "enhance the defamatory sting of the message" cannot be considered content "creation" for the purposes of liability. It is difficult to fit the robust, often normative, content that can be generated by AI into the well-defined protections afforded to search engines and other content aggregation applications such as those at issue in the Yelp and *TripAdvisor* cases.

One case with the potential to substantially impact legal liability for AI-generated false content is pending, as of the publication of this report, before the U.S. Supreme Court. While the case itself is not about generative AI—it is about potential liability for promulgation of ISIS terrorist videos by the defendant search engine's subsidiary in violation of the Anti-Terrorism Act—Supreme Court argument in February directly addressed the potential implications for AI-generated content. In lower court rulings, the company prevailed on its argument that Section 230 of the Communications Decency Act shields the company from liability for the content that its users post to the platform.[125] This ruling is in keeping with precedent (including the *Yelp* and *Trip Advisor* decisions) that Section 230 provides a safe harbor even when a search engine provider organizes, displays, and even summarizes unlawful content generated by another. In deciding this case, the Supreme Court could undermine the expansive freedom courts have long-afforded technology companies that rely on algorithms to identify and "promote" content. But even if the standard remains unchanged, future cases will have to determine whether a particular AI instantiation is merely repeating another's unlawful language or is creating unlawful content of its own.

Inevitably, businesses face some risk of legal culpability for false and otherwise unlawful content generated by the technologies they elect to deploy as well as negative business impacts from such false content. These risks require consideration prior to adoption of new technologies as well as ongoing monitoring, education, and safeguards to minimize the likelihood that the AI tools defeat their very purpose, of generating fast, efficient, and useful content.

---

123 *Kimzey v. Yelp*, Inc., 836 F.3d 1263 (9th Cir. 2016).

124 *Seaton v. TripAdvisor, L.L.C.*, 728 F.3d 592 (6th Cir. 2013).

125 *Gonzalez v. Google L.L.C.*, 2 F.4th 871, 892 (9th Cir. 2021), *cert. granted*, 214 L. Ed. 2d 12, 143 S. Ct. 80 (2022), and *cert. granted sub nom.*, *Twitter, Inc. v. Taamneh*, 214 L. Ed. 2d 12, 143 S. Ct. 81 (2022).

B. Products Liability

Generative AI-related errors do not stop with the proliferation of misinformation. When AI-generated data outputs are used to make critical decisions, such as how much medicine to give a patient or when a self-driving automobile should stop, errors pose the risk of immediate and potentially irreversible physical harm. While products liability law in the United States is well developed, its application to AI is far from certain. For example, while the FDA has cleared more than 520 AI algorithms for application in health care, primarily in medical imaging, it has yet to adopt a comprehensive framework for evaluating AI technologies. The FDA's Center for Devices and Radiological Health (CDRH) issued revised guidance in 2022 suggesting that FDA clearance be required for a more expansive scope of AI-enabled technologies and that "a total product lifecycle-based regulatory framework for these technologies" may be advisable to ensure that the "safety and effectiveness of the software as a medical device are maintained."[126]

In the European Union, complexities regarding fault allocation for AI-driven applications have led authorities to consider adoption of a two-tier approach to liability depending on the level of risk imposed by the technology. High-risk applications of AI, such as those in health care, may face strict liability while lower-risk applications will be based on fault.[127]

---

126 Shania Kennedy, *FDA Releases Guidance on AI-Driven Clinical Decision Support Tools*, Health IT Analytics (Sept. 29, 2022).

127 European Commission, Press Release, *New liability rules on products and AI to protect consumers and foster innovation* (Sept. 28, 2022).

VIII. Generative AI & Risk Management

While AI tools and techniques have been available since the 1950s, the technology has largely been used by IT and analytics groups within companies. This allowed companies to focus risk mitigation efforts on a narrow set of activities and groups. The sudden availability of free web-based generative AI tools puts these capabilities into the hands of every employee and greatly expands the risks.

To manage these risks, companies should adopt a risk-management framework that allows them to ensure the benefits and risks are understood, managed, monitored, and are within the company's risk tolerances. Due to the rapid pace at which new models and tools are being offered, companies will need to continuously update their risk assessments to keep pace with the changes.

A. Risk Management Frameworks for AI

Several organizations have been working for years to develop risk frameworks to help creators and users of AI understand and manage the risks. Two of the organizations with well-developed frameworks[128] are the United States National Institute of Standards and Technology (NIST) and the Organization for Economic Cooperation and Development (OECD). Both organizations provide similar frameworks for addressing the risks of AI systems and suggest how users and companies can increase the trustworthiness of AI systems and contribute to the "responsible design, development, deployment, and use of AI systems over time."[129]

Both organizations focus on the challenge of developing AI systems that are **trustworthy**, and they identify specific factors that can enhance the trustworthiness of an AI System. Their main principles or factors are below:

NIST Trustworthiness Factors	OECD Principles for Responsible Stewardship of Trustworthy AI
• Valid and reliable • Safe, secure, and resilient • Accountable and transparent • Explainable and interpretable • Privacy enhanced • Fair with harmful biases managed	• Inclusive growth, sustainable development, and well-being • Human-centered values and fairness • Transparency and explainability • Robustness, security, and safety • Accountability

NIST released an extensive "Risk Management Framework Playbook"[130] to help a company **Map** the risks in the context of the company and the intended uses, **Measure** those risks using a variety of methods, **Manage** those risks, and establish processes to **Govern** the use of AI tools.

It is also important to understand there can be trade-offs between these different characteristics of a trustworthy AI system.

---

128 ISO has also published ISO/IEC 23894:2023 Information technology — Artificial intelligence — Guidance on risk management. See *https://www.iso.org/standard/77304.html*.

129 U.S. Dept. of Commerce, National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (*NIST AI RMF 1.0*), p. 2 (Jan. 2023).

130 NIST, *NIST AI RMP Playbook*.

> *Trustworthiness characteristics explained in this document influence each other. Highly secure but unfair systems, accurate but opaque and uninterpretable systems, and inaccurate but secure, privacy-enhanced, and transparent systems are all undesirable. A comprehensive approach to risk management calls for balancing tradeoffs among the trustworthiness characteristics. It is the joint responsibility of all AI actors to determine whether AI technology is an appropriate or necessary tool for a given context or purpose, and how to use it responsibly. The decision to commission or deploy an AI system should be based on a contextual assessment of trustworthiness characteristics and the relative risks, impacts, costs, and benefits, and informed by a broad set of interested parties.[131]*

The issues of bias, accountability, privacy, and security have been discussed in earlier sections and will not be discussed here. The issues of safety, reliability, and explainability/transparency are a bit unique for generative AI systems.

1. Safety

AI systems should be safe. That is, they should not "under defined conditions, lead to a state in which human life, health, property, or the environment is endangered."[132] It is unlikely that a *generative* AI system will be used to operate machinery, run an environmental control system, or drive a car, but in the short time generative AI systems have been available to the public, serious safety issues have already arisen. They have been used in ways that generate offensive, racist speech.[133] Text-to-image tools have been used to create sexually explicit or violent images.[134] And in perhaps the most extreme examples, a generative AI chatbot has been blamed for encouraging a Belgian man to commit suicide, and users of an online "companion" tool have claimed the chatbot urged them to commit murder and sent them sexually explicit messages even after they indicated they were not interested.[135]

OpenAI has released a "System Card" that provides extensive information about its efforts to address safety issues in their newest model, GPT-4, such as its ability to produce convincing text that is subtly false, an increased adeptness at providing illicit advice, and risky emergent behaviors. OpenAI notes that while its efforts to improve the safety of the tool were successful, there is a "need for anticipatory planning and governance and further safety research."[136]

Through their analysis, they found the early version of GPT-4 posed risks, including:

- the potential to attempt to identify private individuals by using outside data
- lowering the cost of some steps of a successful cyberattack, such as through social engineering or by enhancing existing tools
- giving detailed guidance on how to conduct harmful or illegal activities
- responding with speech that is hateful, using discriminatory language, inciting violence, or generating content that could be used to spread false narratives

Their safety engineering efforts improved the safety of GPT-4, including by reducing the tendency to hallucinate and reducing the risk of adversarial prompts or exploits (attempts to persuade GPT to provide responses it should not). The

---

131 *Id*. at 13.

132 ISOIED TS 5723:2022.

133 Emily Dreibelbis, *Shocker: Bot Trained on 4chan Chats Is Super Offensive*, PC Mag (June 8, 2022).

134 Amit Dar, *Why Generative AI Is The Next Frontier in Trust & Safety*, ActiveFence (Dec. 21, 2022).

135 Amrutha Pagad, *An AI chatbot pushed a Belgian man to suicide, it seems*, DailyO (Mar. 31, 2023).

136 *GPT-4 System Card*, OpenAI (Mar. 23, 2023).

System Card goes into significant detail. It shares the specific prompts used to try to elicit unwanted responses, along with the response from the early GPT-4 and the revised responses from the version of GPT-4 that was launched in early 2023.

A company choosing to leverage generative AI systems should address these safety issues as part of their risk assessment and management process to determine whether their planned use cases may be affected by these safety issues. And if so, the company should consider implementing measures to reduce the risk, such as training and awareness for employees who will interact with the system explaining how they should handle such issues or limiting the uses of the AI to lower-risk environments.

2. Reliability

NIST defines "reliability" as the "ability of an item to perform as required, without failure, for a given time interval, under given conditions."[137] Given the relatively short period of time generative AI tools have been available for public use and experimentation, it may be difficult to determine whether they are reliable for a given use. As noted earlier in this paper, many generative AI tools still face accuracy and hallucination challenges. And text-to-image tools still struggle to render certain images accurately, including human hands.[138]

Many generative AI tools also still struggle with certain math or "story" problems. For example, when ChatGPT (using the GPT 3.5 model) was asked:

> *If I have a 6 gallon bucket and a 12 gallon bucket, how can I use these buckets to measure out exactly 6 gallons of a liquid?*

the system did not provide the obvious answer, which is to fill the 6-gallon bucket.[139] Instead, it recommended filling the 12-gallon bucket first and then pouring the liquid into the 6-gallon bucket. As an alternative, it suggested using a smaller container, such as a 1-gallon jug, to help you measure out the liquid "more precisely" with the following procedure:

- *Fill the 12 gallon bucket completely with the liquid.*
- *Fill the 1 gallon jug with liquid from the 12 gallon bucket and pour it into the 6 gallon bucket.*
- *Repeat step 2 five more times, until the 6 gallon bucket is completely full.*
- *Pour the 6 gallons of liquid from the 6 gallon bucket into a separate container.*

This is still not the most straightforward solution to the problem. And when ChatGPT (again the 3.5 model) was asked the same question a second time, it generated a different answer, which failed to provide a process to resolve the question. Google's Bard provided a similarly complex set of procedures that involved using both buckets.

When the same question was asked of ChatGPT using the GPT-4 model, it provided the obvious and simple answer—to fill the 6-gallon bucket. More improvements will continue to be made in other areas over time in these types of systems. But employers should carefully consider these shortcomings when assessing the risks of using generative AI tools for certain use cases. Companies may need to test the generative AI platform to ensure it will be reliable for the intended uses and repeat that testing over time as appropriate.

---

137 NIST AI Risk Management Framework, p. 13.

138 Kyle Chayka, *The Uncanny Failures of A.I.-Generated Hands*, The New Yorker (Mar. 10, 2023).

139 Yejin Choi, *See Why AI is incredibly smart -- and shockingly stupid*, TED Talks Daily, Audacy.com (Apr. 28, 2023).

3. Transparency, Explainability, and Interpretability

AI systems are sometimes referred to as "black boxes," because they do not provide any insight into how they function or why they generated a response. A transparent system helps users understand **"what happened"** in the system, while an explainable system allows a user to understand **"how"** the tool generated a decision or response. If a system is also interpretable, it allows a user to understand **"why"** the system made a particular decision. Without such insights, it can be challenging to monitor or oversee an AI system or audit or govern it.[140]

Some generative AI solutions aim to provide at least some context for the decisions they render. For example, some tools provide links to websites or cite the portions of documents that were relied upon to provide an answer or explain the analysis. Some tools even provide a more step-by-step summary of how the tool provided the solution.

As noted earlier, some current laws either explicitly require an audit of AI tools or may require explanations about how algorithms work. So before an employer leverages a generative AI tool for decisions affected by those laws, the "black box" issue will need to be addressed in the risk assessment process. Even where not required, an employer may nevertheless want to engage in a similar process to understand how it will evaluate safety, accuracy, and reliability issues if the tool does not provide transparency, explainability or interpretability of its process and results.

B. Building a Risk Management Process for Generative AI

If an organization already has a risk management function with existing processes and policies, these existing mechanisms can be adapted for the risks presented by AI systems. If a company does not have an existing program, the NIST AI Risk Management Framework provides a good general description of the process. NIST also has a more general Risk Management Framework that provides a wealth of information to help a company understand how to approach this challenge.[141]

Next, a company should understand how it intends to use a generative AI system. For example, will it use commercially available generative AI tools and leverage the general skills they provide (*e.g.*, answering questions, brainstorming ideas, summarizing documents, etc.)? Or will the generative AI system be used to analyze the company's own data?

Different uses of generative AI tools create different risks, and the risk-management approach will need to be tailored to the circumstances. Below is a table that classifies some of the different use cases available today for generative AI tools, along with examples of potential mitigation steps a company may want to consider.

---

140 NIST AI RMF, pp. 14-15.
141 NIST, *NIST Risk Management Framework*.

Risk Level	Examples of Potential Uses	Examples of Potential Mitigation Steps

Low	• Tool is used to leverage its large language model capabilities, not to elicit factual information
	
	• No company data is provided for analysis
	
	• Tool provides drafts, lists, or list of ideas for consideration, but are not intended to be final work product
	
	• Users can evaluate appropriateness and accuracy of the response and monitor for bias before using the output
		• Train employees about risks and proper use, including:
		  • potential for hallucination and bias
		  • potential for offensive content to be generated
		  • need to review output before use
		
		• Prohibit employees from uploading company documents or information
		
		• Consider blocking access to tools for employees who should not use the tools
		
		• Implement policies to limit use to approved use cases
		
		• Require use of company-owned/ controlled accounts when using the Tool
		
		• Require that output be stored only on company-owned equipment/storage areas

Medium	• Tool is used to review documents or company data for analysis • Users ask Tool for factual information • Tool is used to compare content against a set of rules or policies • Tool is used to summarize, proofread, or translate documents • Tool "attends" meetings and summarizes meeting, identifies action items, etc. • Tool used to extract data from documents	• Governance is established to ensure sensitive company information or personally identifiable information is not shared with the Tool until appropriate controls are in place (security and privacy review, contracts, etc.) • Privacy policies, notices, and consents are updated to reflect uses of data • Where required, AI impact assessments or Data Privacy Impact Assessments are performed • Methods are developed to test outputs for accuracy or appropriate limits on use are established • Testing of outputs to ensure accuracy, management of biases, appropriateness of results, and interpretability • Consider geographic limitations on use to avoid jurisdictions where generative AI tools are not permitted or more heavily regulated

Risk Level	Examples of Potential Uses	Examples of Potential Mitigation Steps

High	• Tool processes or analyzes Company data repositories or large volumes of company data
	
	• Tool is used to inform or make business decisions
	
	• Tool is used to generate content used outside of the company (blog posts, ad copy, reports, etc.)
	
	• Tool is used to create images or videos
	
	• Tool is used to create images or videos based on a particular style or that resemble a known person
	
	• Tool is used to analyze large volume of company data to build a fine-tuned AI unique to company
		• Awareness or training of employees about
		  • risk outputs may not be "owned" by Company so external uses may need to be limited
		  • potential infringement of intellectual privacy rights, privacy rights, and publicity rights of individuals
		
		• Governance is established to address issues such as:
		  • Disclosures or disclaimers about using AI generated content
		  • Review of contractual commitments to deliver content to customers that the Company owns or to indemnify customers for IP infringement claims
		  • Risk to profitability if generative AI content is used and company cannot claim ownership of output
		  • Company data made available to the Tool is accurate, current, and appropriate for analysis
		  • Testing processes are implemented to ensure output is accurate and safe
		  • Company considers need for AI system to explain how it arrived at decisions
		  • Legal review is conducted to ensure uses comply with applicable laws regulating the use of AI

Highest	• Tool is used in applications that interact directly with customers or the public without human review or intervention (chat bots, customer support, etc.)
	
	• Tool is used to make decisions that affect individual rights (*e.g.*, to lend money, or to hire, fire, or promote, etc.)
		• Governance has been established to address issues such as:
		  • Periodic testing of the accuracy and potential biases of the output, especially as changes are made to the underlying training data, fine tuning changes have been made, or use cases change
		  • Tracking laws/regulations that restrict specific uses of AI tools

C. Contractual Protections for AI Systems

Before using a generative AI solution, and especially before submitting its own data to a tool for analysis or other uses, a company should first evaluate several aspects of the vendor and the tool.

1. Evaluate Privacy and Security Controls

As noted in Section V, above, a company should understand how the vendor of the generative AI tool will use the data, including the prompts or queries, any data submitted for analysis, and the output of the tool. If the data is confidential or subject to privacy or security obligations imposed by law or by contracts with customers or others, the company should ensure the ways the data will be used are consistent with those obligations. Most larger companies offering generative AI tools have been transparent about how data is further processed and how long it will be retained.

As with most technology solutions, a security evaluation of the vendor and their tools may be advisable (if not required by law or applicable contractual obligations) to ensure any data that will be collected, processed, or stored by the vendor will be protected.

2. Review Terms of Use or Contracts

As noted earlier, depending on the types of data that will be analyzed by the generative AI tool, a company may need specific contractual protections for the vendor and its data. For example, a data protection agreement may be required to comply with the California Privacy Rights Act, GDPR, or HIPAA. Even if not required by law, documenting the security controls that will apply to the company's data and limiting the use of its data may be advisable.

If the vendor has online terms of use or privacy policies, a company should also review those terms to ensure they are consistent with its contract with the vendor and do not override protections negotiated in the company's agreement with the vendor.

IX. Proactive Approaches to Addressing Generative AI in the Workplace

As AI tools are changing how employers do business—promising to streamline HR processes, increase employee production, improve customer deliverables, and more—savvy employers should take a proactive approach to determining how to best integrate generative AI and other machine learning into their operations. The following are some recommendations for doing so.

1. **Identify the "Use Case" and Understand the Generative AI Solution**. Organizations can use generative AI in many different ways. Before stakeholders can determine legal compliance requirements, evaluate risk, and implement risk-mitigation measures, they should understand exactly how the technology will be used and its intended purpose within their organizations.

2. **Conduct an Appropriate Level of Testing**. Depending on the desired use case and the corresponding level of risk, organizations should determine and conduct an appropriate level of testing of the generative AI tools for bias, accuracy, falsity, and other factors that could impact the usability of the outputs. For example, there are a variety of prophylactic measures to consider to reduce the potential bias in generative AI content: strive for diverse and representative training data; use multiple datasets to mitigate the potential for biases in any one dataset; regularly review and test output of generative AI models for output; disclose training data and algorithms used in generative AI models to enhance transparency and accountability.

3. **Establish General Guidelines for Permissible Use of Generative AI Tools**. Once the "use case" has been established, organizations should establish some general guidelines for use of the tools, such as (a) the categories of employees authorized to use the tools, (b) the types of information that can be uploaded, (c) the level of decision-making that can be based on the output (*i.e.*, none, preliminary subject to human review, final), (d) the manner of communicating the output and how it is described for users, and (e) information on the generative AI tools' or system's governance and oversight.

4. **Address Compliance with Any Applicable Legal Requirements**. While laws and regulations that address AI are only in their nascent stage, they are (as described in Section III) developing rapidly. Organizations should check for recent developments before launching any generative AI tool for a significant internal or external use. Organizations should consider the potential applicability of laws not specifically directed at AI, such as anti-discrimination laws, general data privacy laws such as the California Privacy Rights Act, and laws regulating the collection and use of biometric data, such as Illinois' Biometric Information Privacy Act. For more high-risk use cases, organizations should consider involving in-house or external experts with the most up-to-date knowledge of potentially applicable legal requirements and their implications.

5. **Implement Contractual Safeguards for Protected Information Uploaded to the Generative Tool**. Uploads of information to a third party's generative AI tool could result in an unprotected disclosure of personal information, confidential business information, trade secrets, or information subject to a non-disclosure agreement. Organizations should carefully evaluate the types of information and data they will allow employees to upload to third-party tools and the contractual protections that should be included in third-party agreements that may be legally required or otherwise necessary to protect that information after it has been uploaded.

6. **Educate Authorized Users on Relevant Risks**. Authorized users should understand not only legal compliance requirements relevant to their use of generative AI tools, but also inherent risks in the technology. In particular,

employers should educate employees about the potential for seemingly true, but inaccurate or even completely false, outputs.

7. **Train Users**. Depending on their generative AI or machine-learning use cases, employers should develop training regarding the organization's use plan for the AI tool or system, expectations around employee use, and related policies. When AI use is part of a job function, the organization should provide employees with a basic understanding of the tool or system, how it works, and guidelines and expectations around how the tool should be used by the organization and its employees. As appropriate, employers should also provide continuous learning to their employees, as these tools and their capabilities—and laws and regulations in the space—are rapidly evolving.

8. **Require Confirmation of Outputs**. Given the potential for outputs that are inaccurate or false, authorized users should be required to test all outputs before they are put to any subsequent use. This testing could include, for example, asking the generative AI tool to provide sources for its output and checking those sources, searching for more recent or alternative sources, or trying to replicate results with similar but different inputs. There are several safeguards that users can implement to help ensure the accuracy of content:

    a. *Fact-checking*: Fact-checking is the process of verifying the accuracy of information in a piece of content. Fact-checkers use various methods, including checking the original sources of information and verifying claims with experts in the relevant fields.

    b. *Peer review*: Peer review is a process where experts in a particular field review a piece of content to ensure its accuracy and validity. This process is commonly used in academic publishing and scientific research.

    c. *Editorial oversight*: Editors review and edit content to ensure accuracy, coherence, and clarity. They also check for plagiarism and ensure that the content adheres to ethical and legal guidelines.

    d. *Quality control*: Many organizations have quality control procedures in place to ensure that content meets their standards for accuracy and reliability. This can involve reviewing content before it is published, conducting post-publication reviews, and monitoring user feedback.

    e. *Corrections and retractions*: When errors are identified in content, they should be corrected promptly. In some cases, retractions may be necessary if the content is found to be inaccurate or misleading.

9. **Engage in Appropriate Transparency**. Any employee who shares the output of a generative AI tool for a business purpose should be required to disclose whether the deliverable being shared was created, in whole or in part, by a generative AI tool. This transparency will allow recipients to choose the appropriate level of scrutiny of the deliverable. Similarly, if an organization were to use outputs from generative AI as a basis, in whole or in part, for a significant decision about an individual, such as whether to hire, fire, or discipline, the role of AI in the decision-making either should be disclosed or the organization might want to evaluate the business/legal case for nondisclosure.

10. **Assess and Re-Assess Risks**. After initial risk assessments and risk treatment decisions have been made, a company will need to periodically re-assess the situation. A company's uses may have expanded, vendors may have released new features or updated their models. When this occurs, the risks will need to be re-evaluated to understand how the risks have changed and to ensure previous risk-treatment decisions are still appropriate.

11. **Involve Stakeholders**. If an employer intends to use generative AI or machine-learning tools in the hiring process, the organization should ensure their Legal, HR, IT, and Information/Data Security groups, at a minimum, are involved. While IT and Information/Data Security groups are integral to guiding an organization through multiple aspects of deploying and implementing a generative AI tool or system, when such tools are used for employment purposes like candidate assessment, employee engagement, or even as part of job function, Legal and HR should be involved to identify and address associated risks.

12. **Ensure Proper Messaging**. Finally, employers should also ensure they are clear and thoughtful in their messaging and communications regarding how the organization uses AI, both internally to their workforces, and externally to the public. Types of information employers should consider incorporating into their messaging include their intended use of the generative AI tool or system, their commitment to responsible use, disclosures about the types of data being collected and stored, and information on the organization's AI policies.

As the capabilities of AI – and its related business uses – continue to proliferate, employers should be proactive and thoughtful about integrating these tools and systems into their operations. And that vigilance should extend beyond the process that leads to initial deployment of AI tools to include ongoing monitoring and assessment to ensure continuing compliance and effectiveness.

At Littler, we understand that workplace issues can't wait. With access to more than 1,700 employment attorneys in more than 100 offices around the world, our clients don't have to. We aim to go beyond best practices, creating solutions that help clients navigate a complex business world. What's distinct about our approach? With deep experience and resources that are local, everywhere, we are fully focused on your business. With a diverse team of the brightest minds, we foster a culture that celebrates original thinking. And with powerful proprietary technology, we disrupt the status quo – delivering groundbreaking innovation that prepares employers not just for what's happening today, but for what's likely to happen tomorrow. For 80 years, our firm has harnessed these strengths to offer fresh perspectives on each matter we advise, litigate, mediate, and negotiate. Because at Littler, we're fueled by ingenuity and inspired by you.