

Guard Against Algorithmic Discrimination When Using ChatGPT & Other AI Apps



Beware of using chatbots to make hiring and recruiting decisions.

Like many companies, you may be using ChatGPT, Bard, Bing and other large learning models (LLMs) and generative artificial intelligence (AI) products, aka, “chatbots,” to perform HR functions. While AI technology can simplify and improve personnel management processes, it can also get you into legal trouble, especially if you use it for recruitment and hiring. Here’s a quick look at the discrimination liability dangers and what you can do to manage them.

Pitfall: AI Technology May Contain Hidden Biases

Human rights laws make it illegal to base hiring and other employment decisions on race, religion, age, disability, national origin and other protected grounds. Such discrimination may be overt and deliberate or subtle and unintentional. The latter form of discrimination is the product of the fact that almost human beings harbour some degree of hidden prejudice and bias, even those who earnestly accept and try to practice the principles of equal opportunity and nondiscrimination. That includes the people who generate the data used to train chatbots.

Accordingly, human rights commissioners in Canada, the US and Europe have cautioned that data and algorithms built into chatbots may incorporate the subtle prejudices of the humans who create them. Accordingly, relying on these products to make employment decisions exposes companies to discrimination liability risks.

Example: In 2018, Amazon pulled the plug on an AI-based recruitment program after discovering that the algorithm skewed against women. The model was programmed to vet candidates by observing patterns in resumes submitted to the company over a 10-year period. Most of the candidates in the training set were men. As a result, the AI taught itself that male candidates were preferred over female candidates. In other cases, it’s been reported that LLMs generated code stating that only White and Asian men make for good scientists.

Solution: 4 Ways to Guard Against Discriminatory Use of AI

The simplest way to manage discrimination risks associated with use of AI products is to ban their use for work-related purposes the way Samsung, Apple, Verizon, JP Morgan Chase and other large companies have. But a total ban may be overkill that deprives you of AI’s enormous potential. So, you might want to consider refraining from use of

AI for recruitment, hiring and other HR functions. At the very minimum, you must establish [limitations and guardrails](#) to minimize discrimination risks for HR use of AI.

Action Points:

1. Do a self-audit to rigorously test your algorithms and AI-based selection tools with an eye toward tools that look neutral on their face but have the effect of discriminating against groups protected by the laws, as in the Amazon example above;
2. Caution employees to be sensitive to the hidden risks of bias contained in AI algorithms and data;
3. Ban employees from following the instructions from or using content these platforms generate unless and until HR or another manager with knowledge of discrimination laws vets the material and ensures it's nondiscriminatory; and
4. Add language banning algorithmic discrimination to your company's general anti-discrimination policy.

Model Language about AI Use for Non-Discrimination Policy

Employees must be aware that data and algorithms from ChatGPT, Bing, Bard and other generative artificial intelligence (AI) long learning models (LLMs), or chatbots, may contain hidden prejudices or biases or be based on stereotypes about people of certain races, sexes, age, religions or other protected classes under human rights laws. Accordingly, employees may not use these products for purposes of recruiting, hiring, promoting, retaining or making other employment-related decisions unless and until ABC Company's [*HR director/legal counsel/other*] vets and verifies that those applications and tools relying on AI data are fully compliant with applicable human rights laws and will not have the indirect effect of discriminating against groups or individuals those laws are designed to protect.