

Answers

For Claims

Argument	Source	Evidence
Using public information to train LLMs is consider ethical when used to support research and educational methodologies	This is an educational facility that is arguing for a viewpoint that would be beneficial for an educational purpose. This perspective is not surprisingly considering the context	2
LLMs using public information is fair use, but there are further concerns regarding copyright and the creator's consent	This perspective is relatively balanced. But an example they use is an artist how has their work used to train a model that is then used to generate artwork in her style. I believe that this is a case where she is concerned about AI's ability to make it very difficult for her to make money, and that heavily influences her outlook on this topic. As if this AI was allowed, they could be used to generate new art faster than the artist could draw it herself, and put her out of business.	3

Against Claims

Argument	Source	Evidence
Author claims that AI uses their works without their permission	<p>This lawsuit refers to generative AI, producing a prequel to the author's most well selling book, 'Game of Thrones' that was titled A Dawn of Direwolves and used "the same characters from Martin's existing books in the series A Song of Ice and Fire."</p> <p>This one is also bias because the author will now be competing against an impossible opponent, which is able to generate stories that are good enough to threaten his ability as a writer, forcing his hand in a lawsuit.</p>	1
LLMs can spit out text and information that is copyrighted without gaining copyright permissions (and without knowing it is copyrighted)	<p>This source is probably the most balanced read out of all of them. It discusses some of the technical aspects of LLMs that make it a challenge to maintain an ethical viewpoint from the developer's perspective.</p> <p>Despite that the publisher is a privacy charity group that promotes privacy across the world, and this viewpoint is also, not surprising as it is telling people that AI is bad for privacy protection purposes. It does a good job of balancing both sides, especially by explaining it through the technological aspects of LLMs.</p>	5
LLMs can use information in a way that violates privacy laws and regulations.	<p>This source argues that LLMs use a lot of scraped information that is gotten by "ill-gains" such as web-scraping bots. I do agree that their take on restrictions regarding scraped personal information should be in tact, but I believe it to be a weak argument because if the web scraped information was interpreted by thousand of people a day as their job, the same outcome would occur. only differences is the magnitude and scale of interpretation that is not possible due to LLMs.</p>	4

Conclusion

Personally I lean towards it being ethically fine for an LLM to utilise public information that is widely available on the web. Mainly because many of the arguments against are profit driven. You could argue that if I was affected personally by AI stealing my job, or stealing my profits/money, I would also have the same outlook as many of these who are directly affected by this. And honestly many individual (probably including myself) would. As a developer I could argue that AI might steal my coding job, but personally I have the mindset that if that's the case, I should learn to utilise it to benefit my ability, to keep me competitive in this industry. Since I don't think AI will be getting halted or stopped anytime soon considering how slow the legal scene is for regulating LLMs.