# Article information:

The Dark Side of Large Language Models | HiddenLayer MLDR  
<https://hiddenlayer.com/research/the-dark-side-of-large-language-models-2/>

# Article summary:

1. Legal and ethical issues: Large language models (LLMs) can violate copyright laws by producing content that is too similar to the original, and there are no laws preventing service providers from training their models on any kind of data without consent.

2. Accuracy and bias issues: LLMs are only as good as the data they were trained on, and online learning makes it difficult to vet all the data for accuracy, fairness, and bias. The infamous Twitter bot Tay showed how easily influenced algorithms can be subverted by malicious actors to spread misinformation, inflame hatred, and entice violence.

3. Behavioral issues: Emotional AI applications designed to recognize human emotions can give advice that appears technically sane but can prove harmful in certain circumstances or when the context is missing or misunderstood. An AI counseling experiment run by a mental health tech company called Koko drew criticism for giving users responses partially or wholly written by AI without informing them adequately that they were not interacting with real people.

# Article rating:

Appears moderately imbalanced: The article provides some useful information, but is missing several important points or pieces of evidence that would be required to present the discussed topics in a balanced and reliable way. You are encouraged to seek a more balanced perspective on the presented issues by exploring the provided research topics and looking at different information sources.

# Article analysis:

该文章对大型语言模型的负面影响进行了探讨，但存在一些偏见和不完整的报道。首先，文章提到了版权问题，但未考虑到公共领域和合理使用的概念。其次，文章强调了数据集的重要性，但未提及如何解决数据集中的偏见和歧视问题。此外，文章没有探讨大型语言模型在社交媒体上可能引发的恶意行为和虚假信息传播问题。

文章还提到了“情感AI”可能带来的风险，但未能提供足够证据支持这一主张。此外，在呼吁禁止使用“情感AI”时，作者似乎没有考虑到该技术在某些情况下可能会带来积极影响。

总之，该文章对大型语言模型存在的潜在问题进行了探讨，但需要更全面、客观地考虑这些问题，并提供更多证据支持其主张。

# Topics for further research:

* Public domain and fair use
* Bias and discrimination in datasets
* Malicious behavior and misinformation on social media
* Lack of evidence for risks of emotion AI
* Potential positive impacts of emotion AI
* Need for more comprehensive and objective analysis of issues.

# Report location:

<https://www.fullpicture.app/item/c46a4bdb33facd4de82c496851ba6dee>