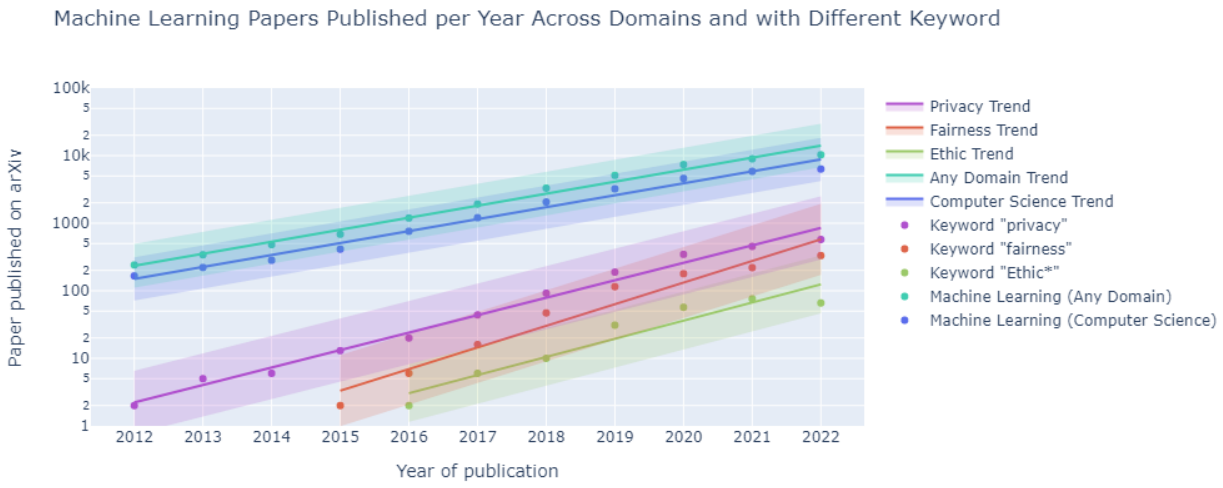
Ethical Machine Learning Algorithms

Here is a brief look at the research around the development of more ethical machine learning systems. As it is now known, many of the machine learning systems currently in use in real-world applications carry over the biases of the data they trained on and the people who build those systems. This can result in injustice when these systems are used to make decisions.
If a select few will build these systems, it seems desirable to ensure the system fairness towards all. It is critical that research explores how to solve this issue. Thus, in this short analysis, we look at the trend of acknowledging and addressing fairness in machine learning research over the years.



Dataset

This dataset was obtained from search queries on arXiv[1] to get the number of research published in a given year with different keywords. For each year, a search was done in the abstract for the keyword “machine learning” alone and with additional keywords (“ethic*”, “fairness”, “privacy”). These keywords are popular words that are used in the discussion around ethical machine learning. Searching in the abstract give us a view of the paper that take into account these concerns without necessarily being overly specific about the subject of the paper. The data collected also include the total number of papers published with the keyword “machine learning” in the abstract, both exclusively within computer science and over all domains.
This data was obtained from the year 2012 up to 2022 as of the 15 of November.
The code to produce the figure can be found on GitHub[2] along with the data used.

[1] https://arxiv.org/search/advanced
[2] https://github.com/Akssi/EthicalMachineLearningResearch

Caveats

This dataset is obtained from a single source for all observations. Any conclusion should take into account that this dataset is not exhaustive and biased towards researchers using arXiv as their main pre publishing/publishing platform.

Only a short list of keywords was used to search through arXiv for papers that would address the fairness issues in machine learning algorithms used in the real-world.

The mention of one of this short list of keywords in the abstract of a paper does not necessarily mean the paper's focus was on solving this issue. It is, however, an indication that the authors acknowledged and potentially took into consideration these issues in their work.

Insights

Over the 10 years analyzed, we can see that privacy has always been a concern, with a trend increasing faster than the machine learning publication trend.

Although addressing the concerns for fairness in machine learning seems to have started later (around 2015) it's growth trend was faster than the trend for privacy, allowing to almost catch up in 2022 with the number of mentions of privacy (same order of magnitude).

Talking about ethical machine learning system is still relatively new. Research in machine learning systems using words such as ethical or ethics seems to have started around 2016. Although only a year after, papers started mentioning fairness, it still lags behind the overall number of papers published by two orders of magnitude and one order of magnitude behind mentions of fairness and privacy.

Another interesting observation that can be made from this data is that the number of paper around machine learning seems to be dropping slightly under trend since last year. This holds for both for the overall machine learning papers numbers and the papers mentioning the specific keywords that were analyzed in this article.

In conclusion, the research around making our machine learning system viable in our diverse world seems healthy from this analysis point of view. Although this dataset strength is weak, it is encouraging to see that research is potentially addressing the problems that were discovered in the usage of machine learning systems. Further research could dive deeper into the specific subdomain of research such as machine learning system that specifically use vision and the ethical concerns that surrounds them.