WHITE PAPER

HOW TO ATTRACT SOCIAL BEINGS?

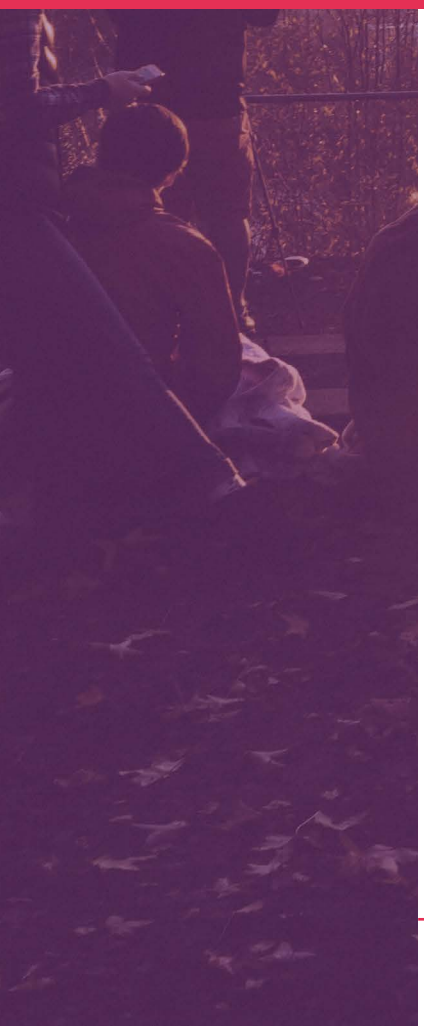
Best practices in moderating news comments.

UTOPIA ANALYTICS

Every media owner has the same goal; the desire for an increase in reader engagement. More readers, each of whom who stay longer, read more content, and actively comment on what they read is the Holy Grail for the online digital media world. Increasing the engagement rates, reducing bounce rates... Any technology that promises to help with these is certainly worth investigating.

Which is why many media owners have started toying with the idea of enlisting the help of machines. Artificial Intelligence, it is claimed, can do the work of thousands of humans in a fraction of time and cost. Even newsrooms, never renowned for their willingness to move away from tried and tested technology of old, have started paying attention to the promise of AI.

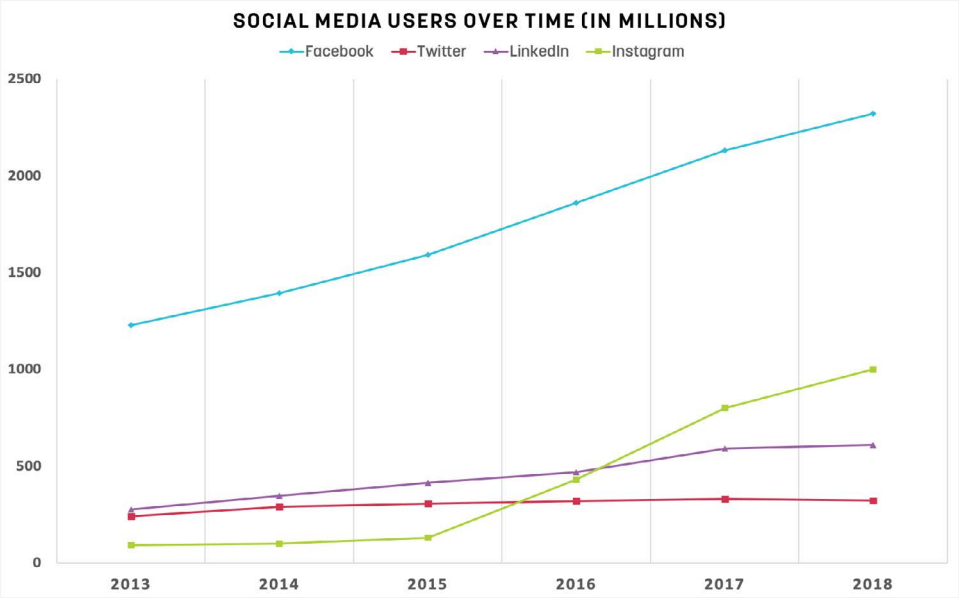
But, is it too good to be true?



In this white paper, we'll introduce the best practices of news comment moderation using AI. We'll also share some examples that explain why AI tools should be chosen wisely. As with many emergent technologies, everything is not quite as it seems, and many companies that claim to use AI in fact do not. In our view, there is AI and "AI". This white paper gives a few simple ways of detecting the fake news about fake AI.

PROBLEM

Digital media users are social beings. They gather where the content, the platform and the environment are just right. In practice, this means that the flow of the new content has to be continuous. Digital media users also crave real-time digital interaction with other users. It's no wonder that once unleashed, digital news media has spread like wildfire around the world, spurred on by Shares, Likes and Comments from the 2 billion connected human beings on this planet.



Source: Statista, LinkedIn

It's this impromptu collaboration, between social media and news channels, that has really added fuel to the fire.

But whilst a user may click on an interesting headline that appears in their social media feed, typically they would only read the article briefly and leave the site immediately.

For a news site, this is far from ideal. To keep their readers engaged, newsrooms have realized that if they can prompt their readers to make a comment under an article, readers will stay longer on the site and perhaps even return to check if someone has responded to them.

Allowing readers to comment on the news comes with complications of its own. Some readers express their opinions in a careless manner, often annoying, offending or even harming other readers. Dissent breeds dissent and can create a hotbed of controversy that may have a long-lasting effect on the news site's reputation in the market. To maintain a positive atmosphere for meaningful discussions, an efficient moderation process is needed.

SOLUTION

How the moderation should be undertaken depends on the individual case. If the number of comments is small it is probably best to moderate comments manually.

If the number of comments on a news site exceeds 500 a day, new approaches could be called for. Natural Language Processing (NLP) and Computational Linguistics provide answers. Combined together with high quality AI algorithms moderation becomes easier, better quality and even cheaper than human labor. But, as mentioned earlier, not all AI is equal. Some products are sold with an AI stamp, rather than an AI substance.

Without a deep understanding of AI, it might be challenging to tell what is good AI and what is not. Indeed, some of these low-end "AI" products are actually not much more than spam filters or traditional IT. If a black-listed word appears, the comment will be flagged to a human moderation queue or for removal. Filters can be fast and effective, but they are not accurate. Most often this type of pseudo-AI will flag something as an issue regardless of the context, so a perfectly harmless comment can end up being flagged, and on the other hand, a lot of harmful content, also spam, will stay untouched. This annoys the commenter and ties down human moderators.

A spam filter combined with human resources is better than no tool at all, but in moderation a sophisticated AI tool takes things to the next level. It performs better and is way more consistent than a group of humans can ever be.

PRE or POST

Should moderation be run pre or post publishing? Much of this is being driven by legislation, one of the key factors in moderation. In many countries legislation strongly pushes editors to pre-moderate news comments before publishing.

Pre-moderation provides many advantages, not least the quality of commenting, which can be substantially higher than moderation done post factum. The downsides can be huge too: the publishing delays may grow unbearable long. If pre-moderation is done manually, it basically means that there can be no discussion at the commenting section as the rate of approval is just too slow. No interaction, no engagement.

In pre-moderation the only high-quality way is to use a modern AI tool. With such a tool, pre-moderation can be done in real time. There is convincing data showing how the number of daily comments steadily increases after AI-based pre-moderation has started. It is likely that readers are attracted by the immediate interaction.

In post-moderation, comments appear online instantly, and only afterwards a number of comments join a queue for moderation. Post-moderation is typical on sites where the number of comments is large. Legislation in some countries rules that there is a ***Report this comment*** button beside every comment so every user can participate in checking the quality. Unfortunately, this feature is widely misused. Research has shown that 2/3 of reported comments are actually proper and a large portion of all improper content remains unreported.

If post-moderation is chosen, a modern AI tool is a valuable helper for human moderators to lean on: it can moderate 100% of comments, not only those reported by other users. The quality of the discussions increases enormously. It is also possible to indicate automatically within the discussion thread when a comment has been removed and also the reason for removal. This will decrease unnecessary complaints of censorship or unfair curation practices.

LET THEM TALK 24/7

With a trustworthy AI tool running, it is recommended that one allows news commenting 24/7. User experience is optimal when commenting is possible at any time, and the comments are moderated with the shortest possible delay. That will result in higher engagement and more page loads. People will stay longer on the site.

We recommend companies to use a commenting platform, that keeps the user data on the company's own systems for their future use. In future, this data can be a gold mine when analyzed with new sophisticated AI tools.

LOGIN or NO LOGIN

Another big question for media companies is how to deal with the user login. Is a real name required or is a nickname enough? Or should commenting be possible without logging in at all? If no login is required, the number of comments will of course be at its highest.

Based on our experience, the number of comments drops if a login requirement is introduced. On the other hand, the quality of comments is not significantly higher even if the real name is required.

Interestingly, people usually stick to the nickname they have originally adopted, even in the internet era, and on sites which allow new nicknames in every post. Nicknames themselves can be offensive and therefore also need to be moderated.

This whole discussion of different login schemes is not very relevant if a sophisticated modern AI tool is in the house. With such a tool the quality of the discussion is always stable regardless of the login scheme.

TOP PERFORMANCE

A top-quality AI moderation system is able to moderate 100 % of comments in real time. Usually news commenting sites want to use manual moderation for 5 to 15% of comments. In other fields of user-generated content moderation, such as sales, ads and chats, the percentage for manual moderation is significantly lower, about 0.01–2%. This means that with AI moderation the majority of comments, even on news sites, get published or deleted immediately.

This arrangement significantly assists media companies as well as their employees; readers find the site attractive, while human moderators are able to work during office hours and only a fraction of their time will be needed for moderation. Human moderator work satisfaction improves as the machine is taking care of a huge part of the routine. Employees will not be exposed to disturbing content with such a heavy frequency as before.

Furthermore, human moderators can concentrate on the meaningful parts of their journalistic duties, while the AI tool keeps the moderation policy stable. If necessary, human moderators or editors can easily adjust the policy by providing new moderation decisions.

TRAINING

Training is an important part of modern AI and machine learning products like Utopia AI Moderator. The following therefore do not apply to pseudo-AI tools.

Real AI services learn automatically from human decisions. They learn what sort of content is OK on each site, what is acceptable and what is not, all according to the Editor-In-Chief's policy. These top-class algorithms understand the semantic meaning by both comment and context.

When you pick your algorithm for moderation it is important to remember that user generated comments rarely use language in a formal or grammatical manner. They often contain emoticons, incorrect punctuation, missing spaces and spelling errors. They also have new words and old words in totally new meanings. "Fake" AI algorithms do not work well because they make an assumption that people use perfect language at all times.

DOES IT REALLY UNDERSTAND?

We are often asked if Utopia AI Moderator recognises sarcasm. The answer is yes and no; the skills of AI depends a lot on the way how human moderators made decisions on sarcasm or irony. In fact, many people do not recognise sarcasm. Because of this, human moderators tend to treat sarcasm as if not understanding it; if it is likely to be offensive, it will be removed.

DATA QUALITY

No matter which moderation solution you choose, the quality of data determines the upper limit of the technological performance. If the human moderators fail to agree what should or shouldn't be accepted in 25% of comments, then the accuracy for the machine learning model will drop by the same level. (And this unfortunately is common between humans.)

Such problems can be fixed up to a point with skilled data scientists and high-quality AI model training. It typically improves the quality of the moderation model but does not completely fix the problem, if it is a data-related one.

In our approach, the machine always remembers the moderation policy; it is a databank that contains an up-to-date view of all the previous moderation decisions. Some of the best AI products are able to give feedback to human moderators whether their decisions follow the moderation policy or not, which is an essential on-the-job training resource.

TIME SCALE

We are sometimes asked how long it takes to install and teach the Utopia AI system if no commenting has taken place and thus no training data is available. It's quite a straightforward process; first, one has to choose a commenting platform, deploys it and moderates the comments manually for about a month.

After that time there will be enough data to create the first AI model. Building the model takes two weeks. When the Utopia AI Moderator model is taken into use, the need for human moderation can be cut in half. In our experience, after a further five months, on news commenting sites Utopia AI will be handling between 85–95% of the moderation work.

This is true of learning AI models, such as Utopia AI. With these, the ability to learn is crucial because language is always changing and so is the moderation policy. Without training and learning the quality of moderation decreases dramatically in a year.

Utopia Analytics Vilhonkatu 6 A, FI-00100 Helsinki, Finland sales@utopiaanalytics.com utopiaanalytics.com