News Article Classification

Tool classifying news articles into true and fake based on features extracted from the title and text of the article.

Created by: Ami
Created on: November 28, 2022 2:33 PM
Changed on: November 28, 2022 11:55 PM

Context of use: Education
Level of education: Bachelor

Impact on society
What impact is expected from your technology?

This category is only partial filled.

What is exactly the problem? Is it really a problem? Are you sure?
The spread of false/misleading information often leads to extreme opinions and contributes to the radicalisation of society. Especially forms like slander and defamation against persons or marginalised people, give rise to violence. Platforms have an obligation to remove or flag such information or limit its spread. It is virtually impossible to manually keep track of and remove problematic content on larger platforms. Therefore, automatised tools have to be used to filter out such content.

Are you sure that this technology is solving the RIGHT problem?
There are deeper levels to this problem, which cannot be solved as easily. The technology would, however, prevent reinforcing the problem that lies on a deeper level and limit its effects.

How is this technology going to solve the problem?
Tools that do not use AI would need to rely on blacklists of words or expressions to filter upon. This approach has a limited effect since users can easily find out which words and expressions get removed and find ways around them.
An AI tool can categorise articles into true and fake with relatively high accuracy. Using this classification, fake articles can be prevented from being published, shared, or at least flagged as such.

What negative effects do you expect from this technology?
There will be false positives and also false negatives. On the one hand, false positives would lead to possibly important information being prevented from getting shared. On the other hand, false negatives might be less likely to be identified as fake by the user since they are not used to questioning the reliability of published/unflagged articles as much anymore.

In what way is this technology contributing to a world you want to live in?
If the technology works as it should, it will reduce the impact of articles promoting misinformation, hatred, defamation, discrimination, racism, etc. and by lowering their impact also decrease their ramifications.

Technology Impact Cycle Tool

News Article Classification

Now that you have thought hard about the impact of this technology on society (by filling out the questions above), what improvements would you like to make to the technology? List them below.
This question has not been answered yet.

Hateful and criminal actors

What can bad actors do with your technology?

This category is only partial filled.

In which way can the technology be used to break the law or avoid the consequences of breaking the law?

If trained accordingly, such a tool could be facilitated to establish censoring. Oppressive regimes for example could use such technologies to suppress opposition and other critical voices or even prevent the publishing of reporting on their own crimes.

Can fakers, thieves or scammers abuse the technology?

This question has not been answered yet.

Can the technology be used against certain (ethnic) groups or (social) classes?

If trained accordingly, such a tool could be used to discriminate against certain groups and keep them from voicing their opinions.

In which way can bad actors use this technology to pit certain groups against each other? These groups can be, but are not constrained to, ethnic, social, political or religious groups.

This question has not been answered yet.

How could bad actors use this technology to subvert or attack the truth?

This question has not been answered yet.

Now that you have thought hard about how bad actors can impact this technology, what improvements would you like to make? List them below.

This question has not been answered yet.

Privacy

Are you considering the privacy & personal data of the users of your technology?

This category is not applicable for this technology.

Human values

How does the technology affect your human values?

This category is not applicable for this technology.

Stakeholders

Have you considered all stakeholders?

This category is not applicable for this technology.

Technology Impact Cycle Tool

News Article Classification

Data

Is data in your technology properly used?

This category is only partial filled.

Are you familiar with the fundamental shortcomings and pitfalls of data and do you take this sufficiently into account in the technology?

Language, writing styles and topics change over time. It may well be that the current algorithm works well within a certain time frame but as soon as topics are not related to any the model has been trained on anymore, or as soon as there is a source which writes in a different language or with a variant approach of writing style the accuracy may drastically drop.

How does the technology organize continuous improvement when it comes to the use of data?

The algorithm is not prepared for such changes, it would need to be continuously retrained with newer data to keep track of current differences in writing styles between fake and true articles and the topics covered.

How will the technology keep the insights that it identifies with data sustainable over time?

As long as there is enough freely accessible news there will be enough data. If the technology would actually be used, the data it is trained on should be safely stored to be able to look into it if necessary.

In what way do you consider the fact that data is collected from the users?

This question has not been answered yet.

Now that you have thought hard about the impact of data on this technology, what improvements would you like to make? List them below.

This question has not been answered yet.

Inclusivity

Is your technology fair for everyone?

This category is not applicable for this technology.

Transparency

Are you transparent about how your technology works?

This category is not applicable for this technology.

Sustainability

Is your technology environmentally sustainable?

This category is not applicable for this technology.

Technology Impact Cycle Tool

News Article Classification

Future

Did you consider future impact?

This category is only partial filled.

What could possibly happen with this technology in the future?

People could get used to the fact that they only get presented with verified information, so if some misinformation gets misclassified it may be harder to convince people that it is indeed false.

Sketch a or some future scenario (s) (20-50 years up front) regarding the technology with the help of storytelling. Start with at least one utopian scenario.

This question has not been answered yet.

Sketch a or some future scenario (s) (20-50 years up front) regarding the technology with the help of storytelling. Start with at least one dystopian scenario.

This question has not been answered yet.

Would you like to live in one of this scenario's? Why? Why not?

This question has not been answered yet.

What happens if the technology (which you have thought of as ethically well-considered) is bought or taken over by another party?

This question has not been answered yet.

Impact Improvement: Now that you have thought hard about the future impact of the technology, what improvements would you like to make? List them below.

This question has not been answered yet.