



Artificial Intelligence Semester 4

Project **newsDetector**

***Potential Impact assessment – 10
subjects***

Fontys university of applied science

Mohammed Alharbi

Class: AI41

Student number: 4089553

Impact on society.....	4
Internal - stakeholders:.....	5
External - stakeholders:	5
Privacy:.....	7
Data:.....	7
Transparency:	8
Hateful and criminal actors	9
Future.....	10
Final Conclusion	11
EDA.....	12
Sources.....	12

Potential Impact assessment – 10 subjects

"Assessing the potential impact of your project is just as important as cleaning the data or preparing your models. Without the right impact assessment, the trust in AI algorithms will erode, which will slow the innovative development of AI in our organization and/or society."

Impact on society

Over time the methods which we used to share news have changed from newspapers to television channels and recently on social media. An increase of 37 million users started getting their information through popular social media platforms this as a result allowed the spread of fake news to grasp the user's attention and potentially sell it to other companies by advertising a product for instance. In addition to that the content of fake news is dangerous because humans by nature have individual weaknesses such as **naïve realism** (humans believe that their opinion is correct and other points of view are biased.) and **confirmation** (humans prefer to read news regarding existing opinion they have) that lead to treating fake news as real news and once the fake news is formed it is exceedingly difficult to correct it.

Fake news with topics such as covid-19 is very problematic due to the sensitivity of the topic. A case where the number of deaths increased in Iran in 2020 occurred as result of publish covid-19 tweets regarding alcohol being cure for corona. Therefore, we are dealing with issue of fake news in online articles and social media covid19 fake news which in one of the research projects have been called 'infodemic'

newsDetector is a technology which is intended to reduce the burden of investigating the truth of news reliability. If we looked at the definition of fake news, we would notice that fake news is intentionally written, and it is verifiably false. Therefore, fake news can be checked manually, and its reliability can be investigated while on the other hand humans by nature find it hard to distinguish the reliability of fake news due to individual weaknesses. News readers via social media and online articles will be able to save more time and be sure whether the news they read is 90% or 80% fake or real most of the time which help them overcome the individual weaknesses fake news target and less people are likely to believe biased information about corona which would reduce the number of deaths and hospitalized persons.

Conclusion:

Social media platforms became the dominating approach of obtaining news which allowed the dissemination of fake news. Fake news target individual vulnerabilities in human nature to make them believe biased information and treat it as real news. The technology informs the user about the reliability of the news with confidence percentage to the user so that the investigation problem of the news can be less intensive and help users overcome their individual weaknesses that are being utilized by fake news.

Internal - stakeholders:

The internal stakeholders are my artificial intelligence teachers as for this semester, the following is information about each stakeholder:

Machine learning teacher	Zhao,Qin Q.	q.zhao@fontys.nl
Societal impact teacher	Litsenburg,John J.T.M. van	j.vanlitsenburg@fontys.nl
Data analytics & investigative analysis	Schmitz,Niek N.T.A.	n.schmitz@fontys.nl

Those stakeholders must be updated about the project regularly until the final deadline of the project.

External - stakeholders:

NAME OF STAKEHOLDER	RELATIOSHIP	HOW AFFECTED?	CONSULTED	TAKE INTO ACCOUNT
People that grasp news through online articles.	DIRECT	They will be more confident about articles' reliability beforehand therefore less work required to manually check the reliability.	NO	YES
People who grasp covid-19 news through social media platforms.	DIRECT	The upmost percentage of failure would be 10% for newsDetector and specific evaluation approach has been take to treat covid-19 topic with covid news extremely carefully. Thus, those stakeholders will be informed about covid-19 in a much more reliable way ad the number of deaths due to corona virus should decrease in the country where the technology is used.	NO	YES
Twitter social media platform	INDIRECT	Since most of the data used in the project regards to covid-19 news on twitter social media platform by predicting whether or not the news is fake or real is important to the platform because this might populate the platform as a place for fake articles / news or real articles and news. The platform should be affected positively because the project aims at eliminating fake news which would make grasping news from the platform more reliable.	NO	YES

Online article authors	INDIRECT	They will reconsider the content they shared since it would be possible to distinguish reliability of online articles by the reader without much efforts. If the content was fake for the sake of personal or political gain it is highly possible that those kind of articles should start disappearing and the meta would be real articles for business models.	NO	NO
Journalists that report Covid-19 news on social media	INDIRECT	If the result of the technology presented to the end user was fake, it might be possible that those journalists would be reported for publishing fake covid-19 news on social which is pretty dangerous. However, what if the model was wrong? What if they were really spread covid-19 fake news for profit? The technology is not perfect and covid-19 topic will be treated differently in newsDetector technology. In those cases, if the technology was not clear in terms of transparency it is possible that the person will be sentenced to jail or will be investigated with thus the number of journalists will reduced in the world.	NO	NO
Social media platforms	INDIRECT-INDIRECT	Just like what happened to one of the popular social media platforms Facebook. They kept advertising articles which were fake and they lost faith in their end users when it comes to the information presented to them, yet since facebook users are many, there is a large number of people who do not know that. Time is the only playing factor here slowly but surely platforms would be classified and popularized according to the reliability of the news presented to the end user.	NO	NO
Companies that advertise for a product or service	INDIRECT-INDIRECT	They will realize that utilizing fake online articles or fake news on social media would no longer grasp people attention which they some social media platforms like Facebook do and online articles. The end user will likely use the technology to read 'real' news suggested by model. When the majority of readers can trust a technology that is 80% accurate most of the time when it comes to online article news the meta will be shifted to advertising in real news websites and articles which will be either expensive or rejected. Thus, advertising products or services will be harder than what it is now.	NO	NO

Conclusion:

The stakeholders have been considered directly, indirectly, and indirectly-indirectly with the technology people that grasp news through online articles, that grasp covid-19 news through social media have been considered as direct users that get affected by the technology. It is expected that they will get affected in a positive way by reducing the burden of self-investigation of the reliability of the news and number of deaths due to corona virus should decrease. Indirect stakeholders such as journalists who report news on covid-19 are expected to be treated as heroes or as bad people by society if the technology was not transparent enough and the number of journalists should decrease. Indirect-indirect stakeholders such as companies that advertise services or products will notice that people will be gathered more at real news websites, and this will shift the current meta from advertising in appealing fake news to honest real news and this might make the advertising process harder.

Privacy:

newsDetector does not require personal data. The technology processes given text and predicted whether the content of online article is fake or not or the content of covid-19 on social media is fake or not. The technology does not register personal data and the data used in the technology does not have any personal data. What is important to mention is that journalists, media organizations or individuals that report news especially regarding covid-19 topic might be judged by the society and their privacy might be invaded but this only would occur if the direct stakeholders used the technology and it displayed that the news was fake, and they spread this information to the public. However, privacy in this case is invaded by the direct stakeholder and not the technology itself thus it is important to use the technology in the correct manner to keep the privacy which is the most fundamental to us.

Data:

Fake news in social media typically contain the text or the body of the news then such technologies take the take process it and with acceptable accuracy percentage it would be useful. However, the body of the news is not only factor which can be utilized to identify fake news but research have also pointed out that auxiliary information can be applied to further improve and optimize the technology even better such as social engagements. For example, it is possible after using this kind of information to understand if a person's behavior has changed drastically after some engagement which we could consider as bribe. The problem with such point is that the quality of fake news data with social engagement produce extremely big, biased, unstructured data which is considered a challenge itself and it is open for further research. Another interesting point is that post features and exploit network

interactions can be used to further optimize the technology, but they are still undeveloped areas for more research.

In this technology only the body text of the news articles and covid-19 tweets is used. It is important to mention that the data used was manually checked for its reliability because fake news in its definition indicate that it is verifiably false which can be checked manually for its reliability if investigated properly. This as a result makes the input the technology will be training on bias.

Conclusion:

The data currently only uses the body of the news text, but it can be improved by adding social engagement factor and features such as image extraction, post features and exploit network interaction to further optimize the technology. Such factors are hard to include in the data because they require more research and are considered challenging. The input used in the data is biased since each fake news was checked manually for its reliability on the other hand fake news on social media in its definition it indicates that the fake news can be manually identified.

Transparency:

The way newsDetector works is by looking at exists fake news articles which are classified as fake or real and train from them the structure and words that appear in both fake and real news articles. It also applies the same concept but specifically for covid-19 social media tweets. The model would treat any topic such as sport and culture normally but treat covid-19 topic extremely carefully. It is worth mentioning that the technology would at most be wrong 20% of the time in any topic and 10% of the time only for covid-19 topic. This is because covid-19 fake news can have dangerous consequences if believed. The technology only facilitates the process of manually verification of fake news by letting the end user or the direct stakeholder know beforehand that an article for instance is fake with 80% chance then the user can further investigate the matter and with little efforts it is possible to confirm the technology prediction. The technology developer is Mohammed Saleh Said Alharbi, it is possible to contact the developer via email 450632@student.fontys.nl and can check file complaints from time to time, that is in case the project has kept being maintained. There is no underlying business model in technology. The end user would insert the body text of the news and the technology displays the reliability of the new as either real or fake and possible a percentage of confidence prediction. Technology has no dark patterns or skim reading on conclusions of the subjects should summarize the points mentioned without utilizing the fact the most technology users would either agree without reading the term of use or skim quickly through it. The technology is not perfect and has shortcomings, the problem lies if the technology predicts that some covid-19 news is fake while it is in fact real this can cause serious problem especially in such sensitive topic and therefore it is taken into account that false negatives (in reality the news is fake and the model predicts it real) should be minimized as much as possible and the price would be an increment in number of (real news predicted as fake) this is done so

that when 'fake' is the result of the prediction it is at most 20% fake percentage and does need some more effort to manually verify it.

Conclusion:

The technology is AI (Artificial Intelligence) model that trained on a balance amount fake news and real news. The model treats covid-19 topic extremely carefully and when it comes to covid-19 topic the highest error rate would be 10% and any other topic 20% of the time. Thus, the user would investigate more to verify the news if it were fake or not but would put way less effort than they would have without technology. It is possible to contact the developer of the technology Mohammed Alharbi via email 450632@student.fontys.nl no business models applied to technology nor dark patterns exist in the technology. The technology minimizes false negatives with a trade-off false positive (by decreasing amount of fake news being predicted as real news, an increment of real news being predicted as fake news would occur) this is extremely because it allows the user with little effort to verify the news manually.

Hateful and criminal actors

NewsDetector allows us to be aware with high chance if some news published as online article or covid-19 social media news is fake or real. The intention is very clear that reduces the disadvantages of spreading fake news and make our lives easier by investigating less to verify news manually while considering the results of technology. However, bad actors or hackers mainly who have knowledge about AI and machine learning could modify the technology by training it. For example, it could be that a person who have such knowledge changes the technology so that it every article or tweet published by certain individual or media organization as fake news. Therefore, it is important to state that to identify whether an individual or media organization spread fake news that if evidence such as screenshots or videos on the technology predicting in that manner it should not be taken into consideration since the technology can be modified if the hack knows what they were doing. Instead to verify the matter the person who reports an individual or media organization that they do spread fake news should instead report the source links to their published news so that it is always checked with the official version that has not been modified by hackers. In addition, we also need to mention that the technology is not perfect and has error rate which can deviate from 10% up to 20 in case the reported individual or media organization has been demonstrated as spreading fake news with the latest and official version of newsDetector, a further investigation should be carried out but with less effort it would be possible to confirm the matter and take then final decision regarding their actions. This way bad hackers should not be able to abuse technology for their hatred towards a specific person or media organization.

Conclusion:

Even though the intentions of the newsDetector is clearly to eliminate negative impact of fake news. There can be some ways to abuse the technology and one of them being the hacker uses newsDetector to accuse an individual or media organization that they spread fake news. While the accusation might come from a person that genuinely believes they spread fake news the matter should be same the way. By using the latest official version of newsDetector and getting confirmation of the matter then by little effort on self-investigation of the matter if both result point to the fact that the individual or media organization is indeed spread fake news then appropriate actions should take place otherwise the technology can be modified, and no evidence of photos or videos of the technology should be taken into consideration.

Future

It is expected that technology to get updated in the future and that the performance of the technology should increase along the way. However, what if that is not the case? What if in the future the performance of technology decreases and what will happen in the future? It is possible to simulate or predict such things by making assumptions and scenarios. NewsDetector technology currently accepts 10% error rate with covid-19 topic and 20% error rate on any other topic on online news in a utopian scenario where the model error rate roughly 1% or less news on social media will become the perfect approach of gathering news because currently it is the lowest cost and easy to access approach while it enabled the spread of fake news a technology that almost predicts every news reliability correctly will completely eliminate the disadvantages of the approach as such. As a result, more and more real news will be published, and journalism will start focusing more on reported news specifically on social media. Fewer people will watch television for news and this could lead to news channels closing their business. Advertising products and services through businesses that uses fake news will reduce and the meta will be shifted to advertising in real news websites/articles which would not always be the case due to those sources often very honest and do not sell their audience to third parties for money therefore advertising products and services would be harder. Almost everyone would be able to reliably get news from anywhere, anytime, on any topic which should make our lives easier and save the time we use to investigate the reliability of some news. If we assumed that corona virus would still exist in the future, the news regarding covid-19 and its vacuums would almost always be correct and accurate therefore fewer people would die due lack of information or taking wrong 'vacuum' as was recommended in some fake news. However, if the error rate has increased to about 40% of the time in future, we would be in the dystopian scenario. Social media news would not be as reliable and most of the time people would have to manually investigate the news. The number of fake online articles and fake news would not degrees and their consequences would still exist. Companies would still utilize the fake online articles to grasp people attention and sell It to them then suddenly, a person was reading about some shocking fake news ended up looking at some cheap shoes near where they live.

Conclusion:

We do not know what the future holds but we can utilize assumptions and gathered opinions about fake news matter to try and predict what the future hold for us with the technology. If the future holds for us that technology would work more efficiently than it is now social media news would become almost a perfect method of obtaining news since it would be able to provide users with easy to access news and lost cost almost always reliable news. This would emphasize the need for journalism field possibly new job opportunities. Companies that used to buy your attention from underlying businesses would change their approach and would be hard to let you be informed about something you did not intend to look for. However, if the future holds that technology will be less efficient than it is now then number of fake news on social media would either be the same or increase. Companies would still buy your attention then you would keep facing those ads in appealing articles. Manual investigation on online news would be required more.

Final Conclusion

Social media platforms became the dominating approach of obtaining news which allowed the dissemination of fake news. Fake news target individual vulnerabilities in human nature to make them believe biased information and treat it as real news.

people that grasp news through online articles, that grasp covid-19 news through social media have been considered as direct users that get affected by technology. It is expected that they will get affected in a positive way by reducing the burden of self-investigation of the reliability of the news and number of deaths due to corona virus should decrease.

The input used in the data is biased since each fake news was checked manually for its reliability on the other hand fake news on social media in its definition it indicates that the fake news can be manually identified. It is possible to contact the developer of the technology Mohammed Alharbi via email 450632@student.fontys.nl no business models applied to technology nor dark patterns exist in the technology. The technology minimizes false negatives with a trade-off false positive (by decreasing amount of fake news being predicted as real news, an increment of real news being predicted as fake news would occur) this is extremely beneficial because it allows the user with little effort to verify the news manually. There can be some ways to abuse the technology and one of them being the hacker uses newsDetector to accuse an individual or media organization that they spread fake news. If the technology performance is low, the number of fake online articles and fake news would not decrease and their consequences would still exist and if the performance is high then almost everyone would be able to reliably get news from anywhere, anytime, on any topic which should make our lives easier and save the time we use to investigate the reliability of some news.

EDA

The initial exploratory data analysis has been done and results can be seen in the proposal.

Sources

H. (2021, November 11). *The Dangers of Fake News* / PeoplesBank. Peoples Bank.

<https://www.peoplesbanknet.com/the-dangers-of-fake-news/>

Irwin, L. (2020, November 12). *The GDPR: What exactly is personal data?* IT Governance Blog

En. <https://www.itgovernance.eu/blog/en/the-gdpr-what-exactly-is-personal-data>

Study: On Twitter, false news travels faster than true stories. (2018, March 8). MIT News |

Massachusetts Institute of Technology. <https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308>

Wikipedia contributors. (2021, November 20). *Fake news*. Wikipedia.

https://en.wikipedia.org/wiki/Fake_news#:~:text=The%20tendency%20for%20human%20to,the%20spread%20of%20fake%20news.

7 things I've learned about journalism in 7 years of being a journalist. (2021, January 24).

YouTube. <https://www.youtube.com/watch?v=Rr7povAInwQ&t=342s>

Fake News Detection on Social Media: A Data Mining Perspective. (2022).

<https://arxiv.org/pdf/1708.01967v3.pdf>

Fighting an Infodemic: COVID-19 Fake News Dataset. (2022).

<https://arxiv.org/ftp/arxiv/papers/2011/2011.03327.pdf>

Gottfried, J., & Shearer, E. (2016, May 26). *News Use Across Social Media Platforms 2016*. Pew Research Center's Journalism Project.

<https://www.pewresearch.org/journalism/2016/05/26/news-use-across-social-media-platforms-2016/>

A Heuristic-driven Ensemble Framework for COVID-19 Fake News Detection. (2022).

<https://arxiv.org/pdf/2101.03545v1.pdf>

Newspapers remain most credible source. (2022). Altoonamirror.Com.

<https://www.altoonamirror.com/opinion/editorials/2018/06/newspapers-remain-most-credible-source/>

Optimization and improvement of fake news detection using deep learning approaches for societal benefit. (2021, November 1). ScienceDirect.

<https://www.sciencedirect.com/science/article/pii/S2667096821000446>

Transformer based Automatic COVID-19 Fake News Detection System. (2022).

<https://arxiv.org/pdf/2101.00180v3.pdf>

Wikipedia contributors. (2021, December 29). *Twitter*. Wikipedia.

<https://en.wikipedia.org/wiki/Twitter>

7 things I've learned about journalism in 7 years of being a journalist. (2021, January 24).

YouTube. <https://www.youtube.com/watch?v=Rr7povAInwQ&t=342s>

(2022). *Total population by child and adult populations* | KIDS COUNT Data Center.

<https://datacenter.kidscount.org/data/tables/99-total-population-by-child-and-adult-populations#detailed/1/any/false/870,869/39,40,41/416,417>

Fake News Detection on Social Media: A Data Mining Perspective. (2022).

<https://arxiv.org/pdf/1708.01967v3.pdf>

Fighting an Infodemic: COVID-19 Fake News Dataset. (2022).

<https://arxiv.org/ftp/arxiv/papers/2011/2011.03327.pdf>

Gottfried, J., & Shearer, E. (2016, May 26). *News Use Across Social Media Platforms 2016*. Pew Research Center's Journalism Project.

<https://www.pewresearch.org/journalism/2016/05/26/news-use-across-social-media-platforms-2016/>

A Heuristic-driven Ensemble Framework for COVID-19 Fake News Detection. (2022).

<https://arxiv.org/pdf/2101.03545v1.pdf>

Lynch, M. P. (2016, November 28). *Opinion / Fake News and the Internet Shell Game*. The New York Times. <https://www.nytimes.com/2016/11/28/opinion/fake-news-and-the-internet-shell-game.html?>

Meet the Queen of shitty robots. (2016, March 6). YouTube.

<https://www.youtube.com/watch?v=nOSnW3HdOao>

Newspapers remain most credible source. (2022). Altoonamirror.Com.

<https://www.altoonamirror.com/opinion/editorials/2018/06/newspapers-remain-most-credible-source/>

Optimization and improvement of fake news detection using deep learning approaches for societal benefit. (2021, November 1). ScienceDirect.

<https://www.sciencedirect.com/science/article/pii/S2667096821000446>

Transformer based Automatic COVID-19 Fake News Detection System. (2022).

<https://arxiv.org/pdf/2101.00180v3.pdf>

Utopia is Dystopia. (2017, April 16). YouTube.

https://www.youtube.com/watch?v=cDeMzg31T2I&ab_channel=KnowledgeHub

Wakefield, B. J. (2016, June 15). *Social media “outstrips TV” as news source for young people*.

BBC News. <https://www.bbc.com/news/uk-36528256>

Wikipedia contributors. (2021, December 29). *Twitter*. Wikipedia.

<https://en.wikipedia.org/wiki/Twitter>