

IMPROVEMENTSCAN A CANVAS voor het filteren van haat en discriminatie

NAME: A.I.-tool voor het filteren van haat en discriminatie

DATE: September 12, 2025 2:18 AM

DESCRIPTION OF TECHNOLOGY
Een A.I.-filter analyseert reacties op Instagram en waarschuwt gebruikers bij negatieve of discriminerende taal. Herhaalde overtredingen kunnen leiden tot een melding bij Instagram en mogelijke accountblokkade.



HUMAN VALUES

Wij denken dat de AI-filter al bijdraagt aan het verbeteren van online platformen door stress en angst te verminderen. Om dit verder te verbeteren, zouden we gebruikersfeedback verzamelen via korte formulieren op Instagram. Op basis van deze feedback kunnen we de AI-filter verfijnen, zodat deze nog beter aansluit bij de behoeften en waarden van de gebruikers.

TRANSPARENCY

Als we terugkijken naar de eerder genoemde stappen, zien we dat de AI-filter al redelijk transparant is en dat we op dit moment geen verbeteringen kunnen bedenken.

IMPACT ON SOCIETY

In de eerste testfase van het AI-filter kunnen sommige mensen onterecht worden uitgesloten en zal de app niet meteen alle negatieve reacties correct herkennen en filteren. Om dit te verbeteren, zou ik een breder getraind AI-model willen gebruiken met diverse testgroepen, zodat het filter inclusiever en accurater wordt.

STAKEHOLDERS

Wij denken dat deze vraag niet direct op ons van toepassing is, omdat de AI-filter door alle Instagram-gebruikers gebruikt kan worden. Wel zouden we de filter verder kunnen verbeteren door meer maatwerkopties toe te voegen, zodat verschillende groepen gebruikers (zoals contentmakers, bedrijven en gewone gebruikers) hun instellingen beter kunnen afstemmen op hun behoeften.

SUSTAINABILITY

Dit sluit niet aan op onze Ai model.

HATEFUL AND CRIMINAL ACTORS

Ja, wij denken dat een goede oplossing zou zijn om de gebruiker de mogelijkheid te geven de AI-filter zelf in of uit te schakelen. Dit zou echter wel beveiligd moeten zijn met een wachtwoord om een veilige omgeving te waarborgen en te voorkomen dat kwaadwillende er zomaar mee aan de haal kunnen gaan.

DATA

hier hebben we nog niet over na gedacht.

FUTURE

Een verbetering die we zouden toepassen, is dat de AI-filter efficiënter werkt, zodat er minder updates nodig zijn. Dit zou de duurzaamheid en betrouwbaarheid van de technologie vergroten. Daarnaast zouden we een firewall kunnen toevoegen om bestands corruptie te voorkomen, waardoor de technologie in de toekomst veiliger blijft.

PRIVACY

Wij denken dat het een verbetering zou zijn als gebruikers van Instagram regelmatig en op een duidelijke manier op de hoogte worden gehouden van het privacy- en databeleid. Dit kan door middel van korte, begrijpelijke meldingen binnen de app, omdat veel gebruikers anders de voorwaarden niet lezen. Daarnaast zou een compacte en simpele uitleg van de AI-filter en hoe deze omgaat met gegevens helpen om transparantie te vergroten.

INCLUSIVITY

Wij denken dat de AI-filter al breed toegankelijk is voor alle Instagram-gebruikers. Toch zouden we de inclusiviteit kunnen verbeteren door ervoor te zorgen dat de AI beter werkt voor verschillende talen en culturen, en dat mensen met een visuele of cognitieve beperking de instellingen gemakkelijker kunnen gebruiken.

FIND US ON WWW.TICT.IO

THIS CANVAS IS PART OF THE TECHNOLOGY IMPACT CYCLE TOOL. THIS CANVAS IS THE RESULT OF AN IMPROVEMENTSCAN. YOU CAN FILL OUT THE FULL TICT ON WWW.TICT.IO



IMPROVEMENTSCAN A CANVAS TO HELP IDENTIFY AND REDUCE RISKS OF HATE AND DISCRIMINATION

NAME: A.I.-tool voor het filteren van haat en discriminatie

DATE: September 12, 2025 2:18 AM

DESCRIPTION OF TECHNOLOGY

Een A.I.-filter analyseert reacties op Instagram en waarschuwt gebruikers bij negatieve of discriminerende taal. Herhaalde overtredingen kunnen leiden tot een melding bij Instagram en mogelijke accountblokkade.



HUMAN VALUES



Now that you have thought hard about the impact of your technology on human values, what improvements...

If you think about the impact of this technology on human values and needs. If you think about how this technology affects the identity of the user, the autonomy of the user (can the users make their own decisions?) and the health and well-being of the user. If you think about all that, what improvement would you (want to) make? In the technology?...

TRANSPARENCY



Now that you have thought hard about the transparency of this technology, what improvements would you like t...

If you think about the communication on the way the technology works and the businessmodel. If you think about the explanation on automatic decisions that are made. If you think about complaint procedures and transparency on possible negative effects. If you think about all that, what would you (want to) improve? In the technology? In context?...

IMPACT ON SOCIETY



Now that you have thought hard about the impact of this technology on society (by filling out the questions...

If you think about the real problem this technology is going to solve. If you think about the ability of this technology to solve the real problem. If you think about possible negative effects and whether this technology will contribute to a world you want to live in. If you think about all that, what improvements would you make? In technology? In context? In use?...

STAKEHOLDERS



Now that you have thought hard about all stakeholders, what improvements would you like to make? List them...

If you think about all stakeholders of this technology. If you think about stakeholders that are less obvious. If you think about the way certain stakeholders are affected by this technology and if you want to take them into consideration. If you think about all that, what would you (want to) improve? In the technology? In context? In use?...

SUSTAINABILITY



Now that you have thought hard about the sustainability of this technology, what improvements would you like t...

If you think about the direct and indirect energy use and the materials that are used in the technology. If you think about the lifespan of the technology and the hidden environmental impact of the technology. If you think about all that, what improvements would you (want to) make? In the technology? In context? In use?...

HATEFUL AND CRIMINAL ACTORS



Now that you have thought hard about how bad actors can impact this technology, what improvements would...

If you think about this technology being used to break the law, or avoid the consequences of breaking the law, or to be used against certain groups, or to attack the truth or to pit certain groups against each other. If you think about all of that, what improvements would you (want to) make? In the technology? In context? In use?...

DATA



Now that you have thought hard about the impact of data on this technology, what improvements would you...

If you think about the limitations of data. Things like subjectivity, incomplete datasets and so on. If you think about the way new insights are handled. If you think about the sustainability of the collection of data or the data that is collected from the users. If you think about all that, what would you (want to) improve? In the technology? In context?...

FUTURE



Impact Improvement: Now that you have thought hard about the future impact of the technology, what...

If you think about an utopian and a dystopian scenario. If you think about the way this technology can change the world. If you think about the consequences of a different party buying your technology. If you think about all that, what would you (want to) improve? In the technology? In context? In use? ...

PRIVACY



Now that you have thought hard about privacy and data protection, what improvements would you like to make?...

If you think about this technology invading someone's privacy or collecting personal data and if you think about the way this technology is compliant with prevailing law and mitigates dataprotection risks and concerns. If you think about all that, what improvements would you (want to) make? In the technology? In context? In use?...

INCLUSIVITY



Now that you have thought hard about the inclusivity of the technology, what improvements would you like to...

If you think about accessibility to this technology. If you think about built in biases or automatic decisions that may be biased. If you think about who is benefitting from this technology and the diversity of the team that creates the technology. If you think about all that, what improvements would you (want to) make? In the technology? In context? In use? ...

FIND US ON WWW.TICT.IO

THIS CANVAS IS PART OF THE TECHNOLOGY IMPACT CYCLE TOOL. THIS CANVAS IS THE RESULT OF AN IMPROVEMENTSCAN. YOU CAN FILL OUT THE FULL TICT ON WWW.TICT.IO

