

QUICKSCAN - CANVAS A.I.-tool voor het filteren van haat en discriminatie

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

DATE: September 12, 2025 2:23 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



De A.I.-filtertechnologie zorgt ervoor dat gebruikers zich veiliger voelen op het platform, omdat kleinerende en discriminerende reacties worden tegengehouden. Dit kan ertoe leiden dat mensen zich vrijer uiten en hun identiteit zonder angst delen.

TRANSPARENCY



Het AI-filter is ingebouwd in Instagram en monitort reacties automatisch. Het waarschuwt bij ongepaste berichten en kan bij herhaaldelijk overtreden accounts tijdelijk verbannen. Instagram informeert gebruikers hierover via meldingen en community richtlijnen.

IMPACT ON SOCIETY



Het probleem is dat kleinerende en discriminerende reacties gemakkelijk op Instagram geplaatst kunnen worden. Dit kan leiden tot symptomen zoals depressie, stress en angst, en draagt bij aan een toxische online omgeving. Onderzoek toont aan dat online negativiteit de mentale gezondheid schaadt en de gebruikerservaring verslechtert, wat bewijst dat dit een serieus probleem is.

STAKEHOLDERS



- Instagram gebruikers

SUSTAINABILITY



Dit sluit niet aan bij mij ai model

HATEFUL AND CRIMINAL ACTORS



Als de A.I.-filter gehackt wordt, kan dit privacywetten schenden door mee te kijken met gebruikers of hun berichten te manipuleren. Daarnaast kan de filter worden omzeild, waardoor discriminerende reacties geplaatst kunnen worden zonder consequenties.

DATA



Ik ben me ervan bewust dat A.I.-systemen misbruikt kunnen worden en dat er risico's zoals bias en privacy problemen kunnen ontstaan. Om deze tekortkomingen te beperken, zou ik een geavanceerde A.I. implementeren die context beter begrijpt en regelmatig wordt geüpdatet om nieuwe problemen te voorkomen. Daarnaast zou ik ervoor zorgen dat het systeem transparant werkt en gebruikers controle geeft over hoe het hun gegevens verwerkt.

FUTURE



In de toekomst zou het AI-filter op meerdere sociale media platforms geïmplementeerd kunnen worden om een veiligere online omgeving te creëren. Daarnaast zal de AI steeds slimmer worden, waardoor er minder menselijke input nodig is, behalve voor regelmatige updates

PRIVACY



De A.I.-filtertool zelf gebruikt geen persoonlijke data. De enige gegevens die verwerkt worden, zijn de data die Instagram al verzamelt bij het installeren van de app.

INCLUSIVITY



Het AI-filter van Instagram is bedoeld om voor alle gebruikers een veilige omgeving te creëren en wordt niet specifiek tegen één groep ingezet. Toch kan er in het begin sprake zijn van ingebouwde bias, omdat het AI-filter getraind wordt op data die mogelijk niet volledig neutraal is. Door gebruikersmeldingen en regelmatige updates kan het AI-filter bijgestuurd worden om eerlijker en nauwkeuriger te beoordelen

FIND US ON WWW.TICT.IO

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



QUICKSCAN - CANVAS A.I. HEBBEN VOOR HET FILTEREN VAN HAAT EN DISCRIMINATIE

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

DATE: September 12, 2025 2:23 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



How is the identity of the (intended) users affected by the technology?

To help you answer this question think about sub questions like:

- If two friends use your product, how could it enhance or detract from their relationship?
- Does your product create new ways for people to interact?...

TRANSPARENCY



Is it explained to the users/stakeholders how the technology works and how the business model works?

- Is it easy for users to find out how the technology works?
- Can a user understand or find out why your technology behaves in a certain way?
- Are the goals explained?
- Is the idea of the technology explained?
- Is the technology company transparent about the way their...

IMPACT ON SOCIETY



What is exactly the problem? Is it really a problem? Are you sure?

Can you exactly define what the challenge is? What problem (what 'pain') does this technology want to solve? Can you make a clear definition of the problem? What 'pain' does this technology want to ease? Whose pain? Is it really a problem? For who? Will solving the problem make the world better? Are you sure? The problem definition will help you to determine...

STAKEHOLDERS



Who are the main users/targetgroups/stakeholders for this technology? Think about the intended context by...

When thinking about the stakeholders, the most obvious one are of course the intended users, so start there. Next, list the stakeholders that are directly affected. Listing the users and directly affected stakeholders also gives an impression of the intended context of the technology.
...

SUSTAINABILITY



In what way is the direct and indirect energy use of this technology taken into account?

One of the most prominent impacts on sustainability is energy efficiency. Consider what service you want this technology to provide and how this could be achieved with a minimal use of energy. Are improvements possible?

HATEFUL AND CRIMINAL ACTORS



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

Can you imagine ways that the technology can or will be used to break the law? Think about invading someone's privacy. Spying. Hurting people. Harassment. Steal things. Fraud/identity theft and so on. Or will people use the technology to avoid facing the consequences of breaking the law (using trackers to evade speed radars or using bitcoins to launder...

DATA



Are you familiar with the fundamental shortcomings and pitfalls of data and do you take this sufficiently into...

There are fundamental issues with data. For example:

- Data is always subjective;
- Data collections are never complete;
- Correlation and causation are tricky concepts;
- Data collections are often biased;...

FUTURE



What could possibly happen with this technology in the future?

Discuss this quickly and note your first thoughts here. Think about what happens when 100 million people use your product. How could communities, habits and norms change?

PRIVACY



Does the technology register personal data? If yes, what personal data?

If this technology registers personal data you have to be aware of privacy legislation and the concept of privacy. Think hard about this question. Remember: personal data can be interpreted in a broad way. Maybe this technology does not collect personal data, but can be used to assemble personal data. If the technology collects special personal data (like...

INCLUSIVITY



Does this technology have a built-in bias?

Do a brainstorm. Can you find a built-in bias in this technology? Maybe because of the way the data was collected, either by personal bias, historical bias, political bias or a lack of diversity in the people responsible for the design of the technology? How do you know this is not the case? Be critical. Be aware of your own biases....

FIND US ON WWW.TICT.IO

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

