Scroll Savely: een AI-filter voor schadelijke content op Instagram

AI-tool die gewelddadige of schadelijke posts op Instagram detecteert en verwijdert om gebruikers (vooral jongeren) te beschermen

Created by: Murphyis
Created on: March 23, 2025 1:22 PM
Changed on: March 23, 2025 3:51 PM

Technology Impact Cycle Tool

Scroll Savely: een AI-filter voor schadelijke content op Instagram

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?

Op Instagram komen soms gewelddadige of schokkende video's voorbij. Dit kan vooral voor jongeren erg schadelijk zijn. Daarom heb ik Scroll Savely bedacht. Met deze AI tool worden de posts meteen verwijderd.

Are you sure that this technology is solving the RIGHT problem?

This question has not been answered yet.

How is this technology going to solve the problem?

This question has not been answered yet.

What negative effects do you expect from this technology?

This question has not been answered yet.

In what way is this technology contributing to a world you want to live in?

This question has not been answered yet.

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.

De AI moet beter begrijpen welke beelden echt schadelijk zijn voor jongeren en welke niet, zodat het de juiste content filtert en geen onschuldige posts worden verwijderd.

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?

Als de AI te streng is, kan het onterecht berichten of groepen blokkeren. Maar als het niet goed werkt, kan schadelijke content blijven staan.

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?

This question has not been answered yet.

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.

Er moet controle zijn om te voorkomen dat mensen de AI misbruiken om onterecht posts te laten verwijderen.

Technology Impact Cycle Tool

Scroll Savely: een AI-filter voor schadelijke content op Instagram

---

Privacy

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

This category is only partial filled.

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

Nee, Scroll Savely scant alleen openbare posts op Instagram en verzamelt in eerste instantie geen persoonlijke gegevens. De focus ligt dus vooral op de inhoud van de post en niet op de gebruiker. Maar als iemand veel schadelijke posts plaatst, kan de AI Instagram waarschuwen over dit account.

Do you think the technology invades the privacy of the stakeholders? If yes, in what way?

This question has not been answered yet.

Is the technology is compliant with prevailing privacy and data protection law? Can you indicate why?

This question has not been answered yet.

Does the technology mitigate privacy and data protection risks/ concerns (privacy by design)? Please indicate how.

This question has not been answered yet.

In which way can you imagine a future impact of the collection of personal data?

This question has not been answered yet.

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

Geen persoonlijke gegevens opslaan. Alleen bij accounts die meerdere keren schadelijke content hebben gepost kan Instagram een melding krijgen.

Human values
How does the technology affect your human values?

This category is only partial filled.

How is the identity of the (intended) users affected by the technology?
Gebruikers kunnen veiliger scrollen zonder onverwacht schokkende beelden te zien. Dit draagt bij aan een positiever gebruik van Instagram en hier zal instagram ook weer profijt van hebben.

How does the technology influence the users' autonomy?
This question has not been answered yet.

What is the effect of the technology on the health and/or well-being of users?
This question has not been answered yet.

Now that you have thought hard about the impact of your technology on human values, what improvements would you like to make to the technology? List them below.
Ik wil dat gebruikers begrijpen waarom bepaalde content schadelijk is en wordt verwijderd. Dit kan door middel van korte meldingen en uitleg, zodat accounts bewuster omgaan met wat ze posten.

Stakeholders
Have you considered all stakeholders?

This category is only partial filled.

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

Did you consider all stakeholders, even the ones that might not be a user or target group, but still might be of interest?
-

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

Technology Impact Cycle Tool

Scroll Savely: een AI-filter voor schadelijke content op Instagram

---

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?
Niet relevant

How does the technology organize continuous improvement when it comes to the use of data?
This question has not been answered yet.

How will the technology keep the insights that it identifies with data sustainable over time?
This question has not been answered yet.

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.
-

Technology Impact Cycle Tool

Scroll Savely: een AI-filter voor schadelijke content op Instagram

Inclusivity

Is your technology fair for everyone?

This category is only partial filled.

Will everyone have access to the technology?
This question has not been answered yet.

Does this technology have a built-in bias?
Niet relevant

Does this technology make automatic decisions and how do you account for them?
This question has not been answered yet.

Is everyone benefitting from the technology or only a a small group? Do you see this as a problem? Why/why not?
This question has not been answered yet.

Does the team that creates the technology represent the diversity of our society?
This question has not been answered yet.

Now that you have thought hard about the inclusivity of the technology, what improvements would you like to make? List them below.
-

Transparency
Are you transparent about how your technology works?

This category is only partial filled.

Is it explained to the users/stakeholders how the technology works and how the business model works?
Niet op te lossen

If the technology makes an (algorithmic) decision, is it explained to the users/stakeholders how the decision was reached?
This question has not been answered yet.

Is it possible to file a complaint or ask questions/get answers about this technology?
This question has not been answered yet.

Is the technology (company) clear about possible negative consequences or shortcomings of the technology?
This question has not been answered yet.

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

Sustainability
Is your technology environmentally sustainable?

This category is only partial filled.

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

Do you think alternative materials could have been considered in the technology?
This question has not been answered yet.

Do you think the lifespan of the technology is realistic?
This question has not been answered yet.

What is the hidden impact of the technology in the whole chain?
This question has not been answered yet.

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

Technology Impact Cycle Tool

Scroll Savely: een AI-filter voor schadelijke content op Instagram

Future
Did you consider future impact?

This category is only partial filled.

What could possibly happen with this technology in the future?
Mensen krijgen een positiever wereldbeeld doordat ze minder gewelddadige beelden zien.

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.
-