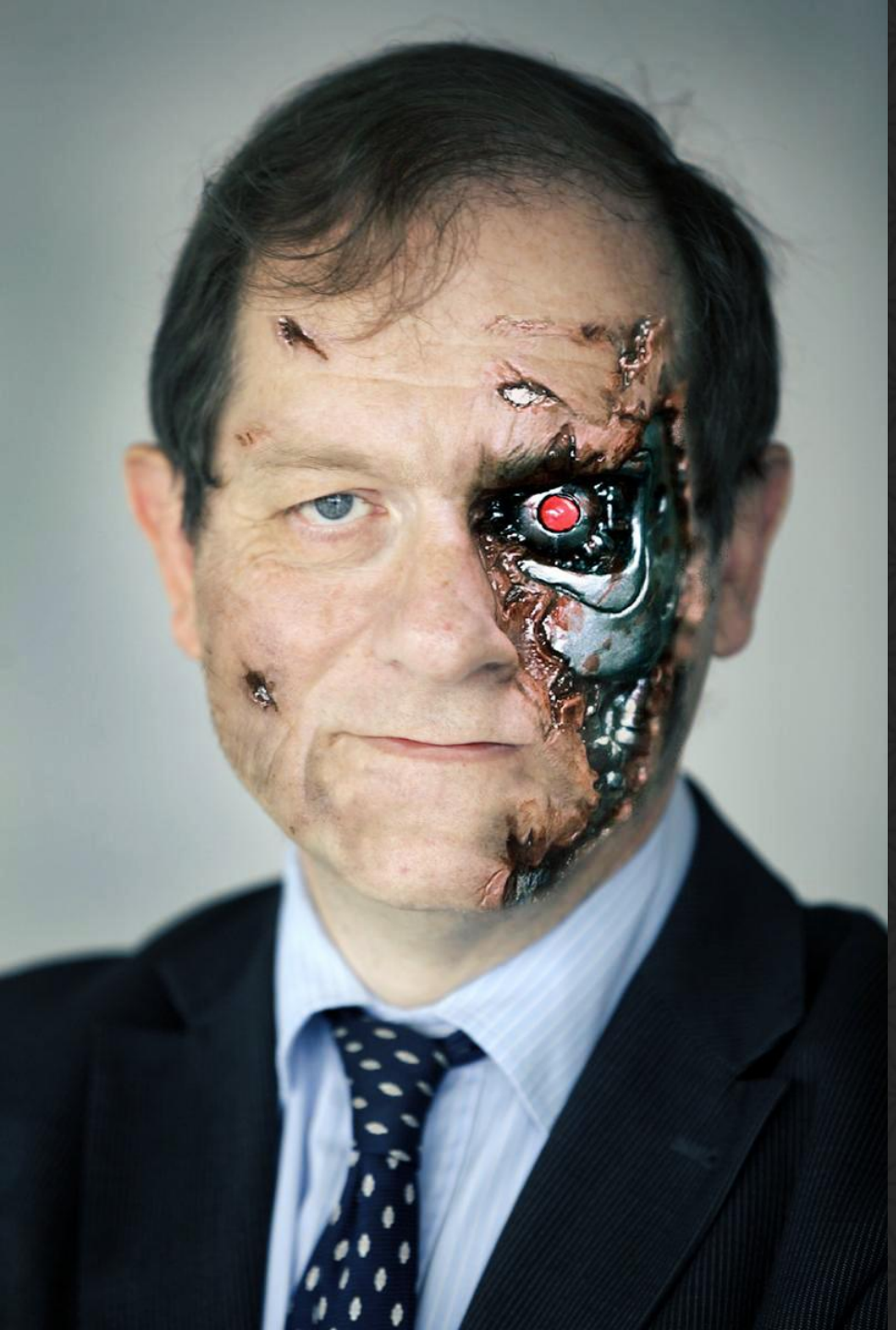


TorfsBot

Een geautomatiseerde Rik Torfs

Thomas Winters
@thomas_wint
thomaswinters.be



Follow

Robot Torfs
@TorfsBot

KU Leugen Rector. Christendemocraat. Cyborg. Geautomatiseerde Kwakfilosoof. Niet te verwarren met @torfsrik. (implementatie: @thomas_wint)

Louvain, Belgium thomaswinters.be/torfsbot Joined September 2016

3 Following **2,872** Followers

Tweets Tweets & replies Media Likes

Pinned Tweet

Robot Torfs @TorfsBot · 2 Jul 2020
Het waardevolle antwoord op onzin is stilzwijgen.
8 41 182

Robot Torfs @TorfsBot · 4h
Dringen we door tot die plaatsen waar iedereen ons al lang is vergeten? Dat kan alleen in Leuven.
1 7

Robot Torfs @TorfsBot · 7h
Zou er een paus zijn die vermoedt dat een aanvaardbare slogan.
1

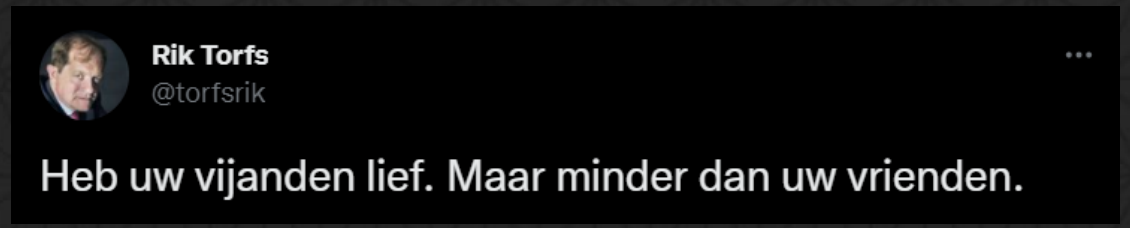
Wat

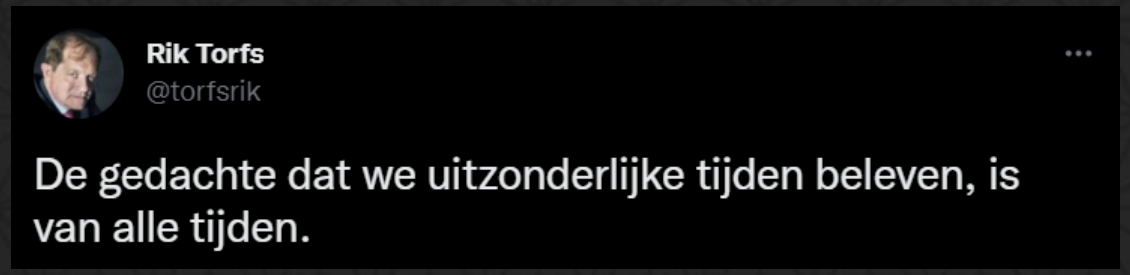
- TorfsBot is een **algoritme** dat de **Twitterstijl van Rik Torfs** leerde nabootsen
- Post autonoom tweets op Twitter = "**Twitterbot**"
- Populairste Belgische Twitterbot

Rik Torfs tweet style

Aforisme

Kort, vaak met kwinkslag





Complexe zinnen

Relatief moeilijk vocabularium, soms non-triviale grammatica, waardoor het vaak veel dieper klinkt dan het is





TorfsBot tweets

TorfsBot heeft leren tweeten in deze stijl & woordenschat









Werking van het algoritme

Hoe werkt zulke tekstgeneratie?



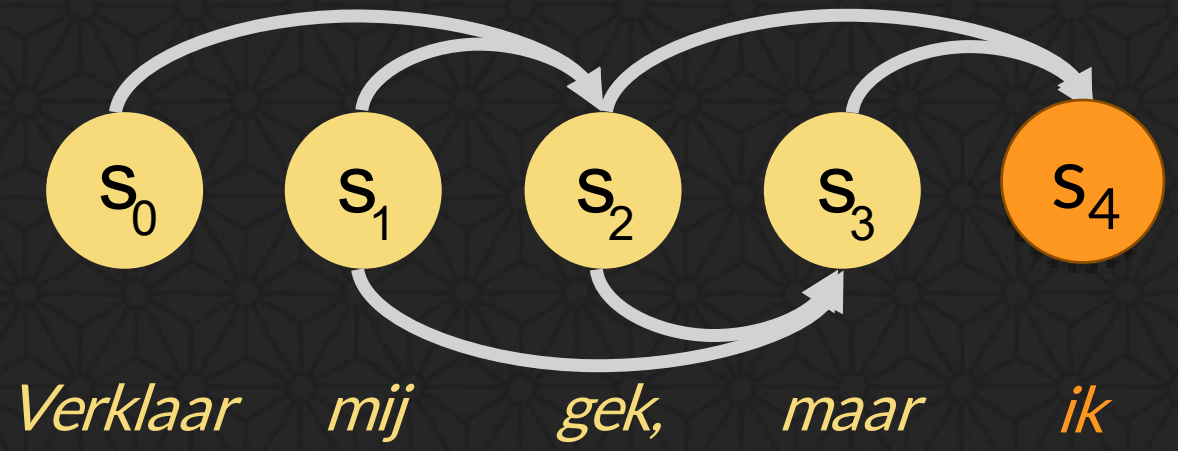
Demo

1. Open je smartphone op tekstveld *(bv notities)*
2. Druk op één van de voorgestelde **auto-complete** woorden *(bv middelste)*
3. Druk **10-20-tal keer** op een willekeurige auto-complete
4. Je hebt nu een tekst gegenereerd die **klinkt zoals jezelf**! *(als je van veraf kijkt)*

Autocomplete heeft **geteld** hoe vaak bepaalde **woorden na andere bepaalde woorden** voorkomen in de teksten die je eerder typte

Algoritme 1: Markovketen

TorfsBot Markovalgoritme **kijkt 2-4 woorden terug**, en neemt woord op basis van hoe vaak dat woord na de vorige woorden voorkwam in Rik Torfs teksten



Lokaal coherent, globaal vaak onsamenhangend

Markov keten voorbeeld



Bot zag *"om niet te"*, en wist dat Rik Torfs dat soms opvolgt met *"herinneren"*

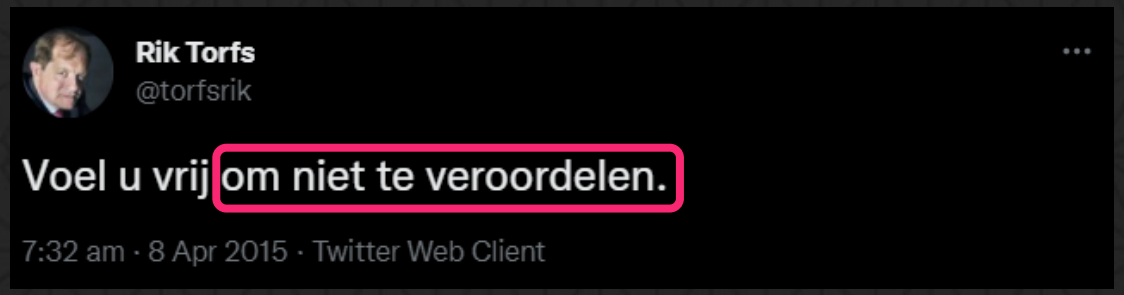


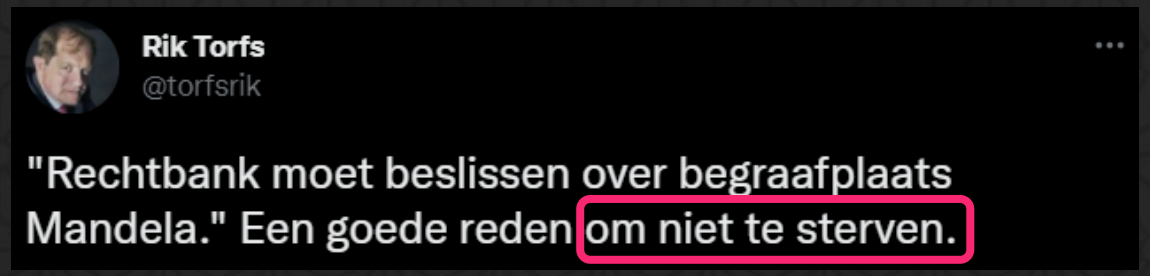


Combineert meestal **meer dan twee** tweets/columns

Evenzeer valide mogelijkheden

Generatieproces is willekeurig, had ook **andere opvolgingen** kunnen kiezen!



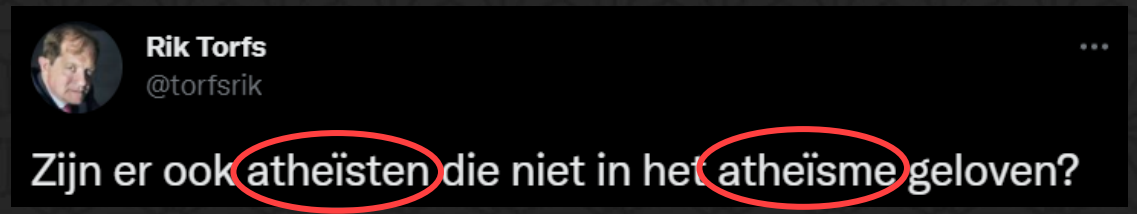


Of zelfs op ander punt in tweet, of bv. na enkel twee laatste woorden *“door een”*, waardoor *“Beste, gevolgd door een tweet.”* gegeneerd had kunnen zijn



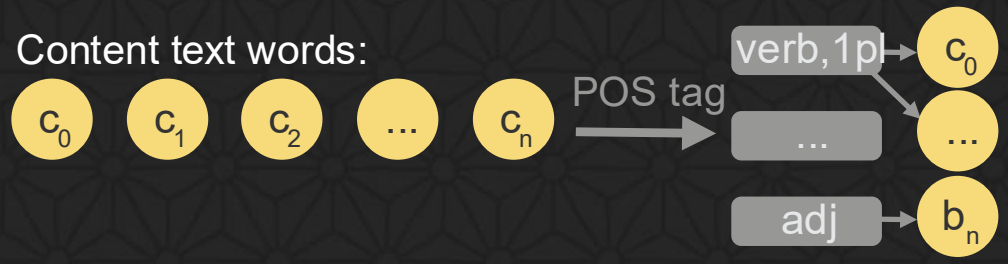


Algoritme 2: Dynamische sjabloon

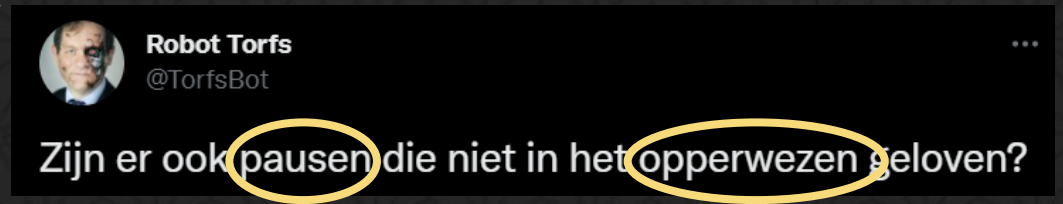




Vervang **kernwoorden** met zelfde zinsontledingfunctie







Rik Torfs Column: *(...) Het feit dat het voormalige Opperwezen dit nieuw verworven inzicht niet probeert te ontkennen, zien zij als een bewijs van hun gelijk. Ook met de Kerk zit het niet snor. Norse pausen.*

Globaal coherent, lokaal soms rare woordcombinaties

Antwoord algoritme



TorfsBot **antwoordt op iedereen** die op hem reageert/tagt

(in tegenstelling tot de echte Rik Torfs die bijna nooit reageert)

Werking

1. **Zoek conversatie kernwoorden:**
   Tel alle woorden in conversatie, en zoek op hoe **zeldzaam** zo een woord typisch is
2. **Genereer kandidaten:**
   Markov keten algoritme genereert 1000 mogelijke antwoorden
3. **Vind beste generatie:**
   Neemt het antwoord dat **meest overeenkomende kernwoorden** heeft met getagde tweet

Gemiddeld aantal interacties per algoritme



Besluit: Lokale coherentie > globale coherentie**?**

Winters, T. (2019). Generating philosophical statements using interpolated Markov models and dynamic templates.

Interacties = kwaliteit?

Krijgt tweet **meer interactie** als het

Net **heel goed** Rik Torfs nadoet

Zó slecht Rik Torfs nadoet dat het **hilarisch slecht** is?



Hoe goed kan TorfsBot Rik Torfs imiteren?

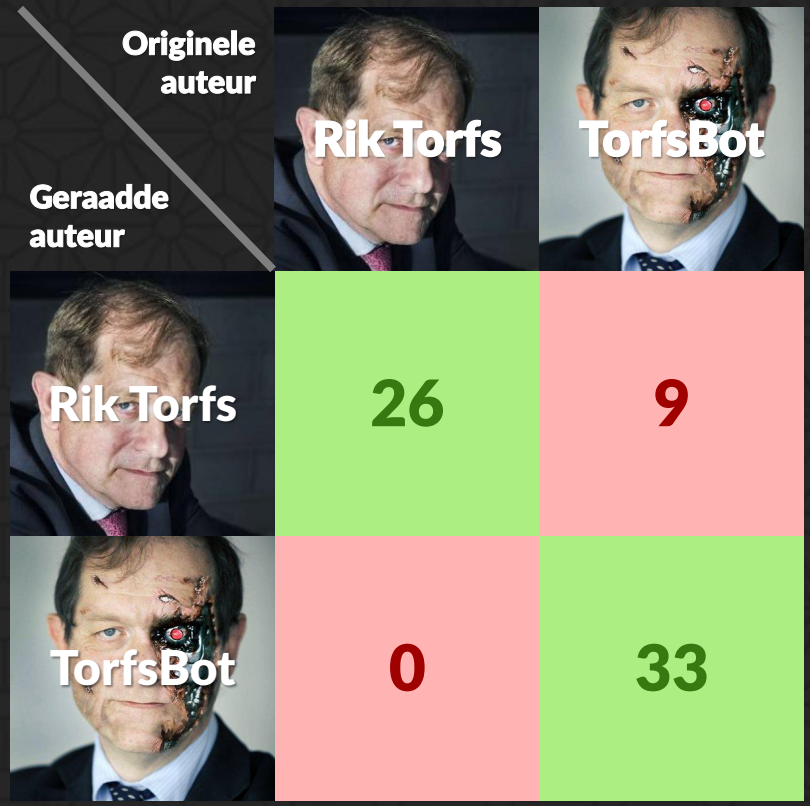


- Tweet elke dag **een willekeurige tweet van Rik Torfs of TorfsBot**
- **Poll** die vraagt van wie het komt
- Elke dag stemmen **100+ mensen** op deze Turingtest

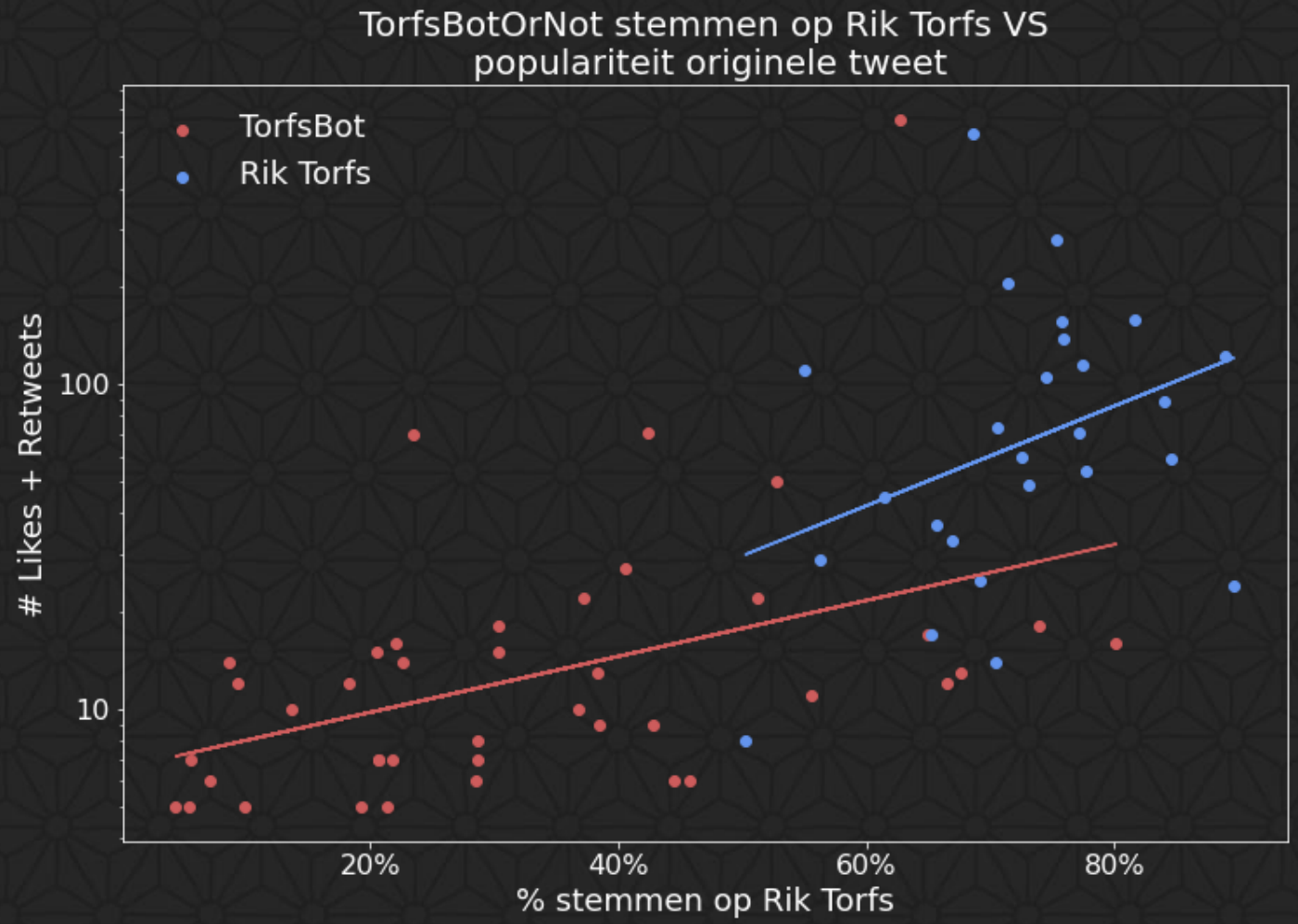
TorfsBotOrNot



- **Gelanceerd:** januari 2022
- **Gemiddeld # stemmen:** 193 / poll
- **Gemiddeld stemmen op juiste:** 68,9%
- **Juistheid:** Nog nooit Rik als Bot verward!



Tweet populariteit vs stemmen op Rik *(Tot 23-03-2022)*



- ✓ TorfsBot's populairdere tweets zijn **meer "Rik Torfsiaans"** *(=**minder** correcte stemmen)*
- ✓ Rik's populairdere tweets zijn **vaker juist herkend** *(=**meer** correcte stemmen)*

Twitterbot regels

Automation rules

Do!

- Build solutions that automatically broadcast helpful information in Tweets.
- Run creative campaigns that auto-reply to users who engage with your content.
- Build solutions that automatically respond to users in Direct Messages.
- Try new things that help people (and comply with our rules).
- Make sure your application provides a good user experience and performs well — and confirm that remains the case over time.

Maak leuke nieuwe ervaringen

Don't!

- Violate these or other policies. Be extra mindful of our rules about abuse and user privacy.
- Abuse the Twitter API or attempt to circumvent rate limits.
- Use non-API-based forms of automation, such as scripting the Twitter website. The use of these techniques may result in the permanent suspension of your account.
- Spam or bother users, or otherwise send them unsolicited messages.

In addition to the policies above, the following ground rules apply to all automated activity on Twitter:

- **Don't surprise or mislead users:** Automated activity should honor users' expectations. Ask for the user's permission before taking an action if you aren't sure.

Maar zorg dat ze niemand lastigvallen

Van: https://help.twitter.com/en/rules-and-policies/twitter-automation

Twitter reactie regels



Een Twitterbot mag pas reageren als een **gebruiker interactie toont** met bot

→ TorfsBot antwoordt enkel wanneer hij getagd wordt

@ZoekNieuws: Fake-news bevechtende Twitterbot





Probleem: Tegen de Twitterregels!

Want die gebruiker heeft niet gevraagd om die Twitterbot

Oplossing: Mensen taggen bot in reply

Impersonation policy

What is not an impersonation policy violation?

Accounts with similar usernames or that are similar in appearance (e.g., the same profile image) are not automatically in violation of the impersonation policy. In order to violate our impersonation policy, the account must portray another entity in a misleading or deceptive manner.

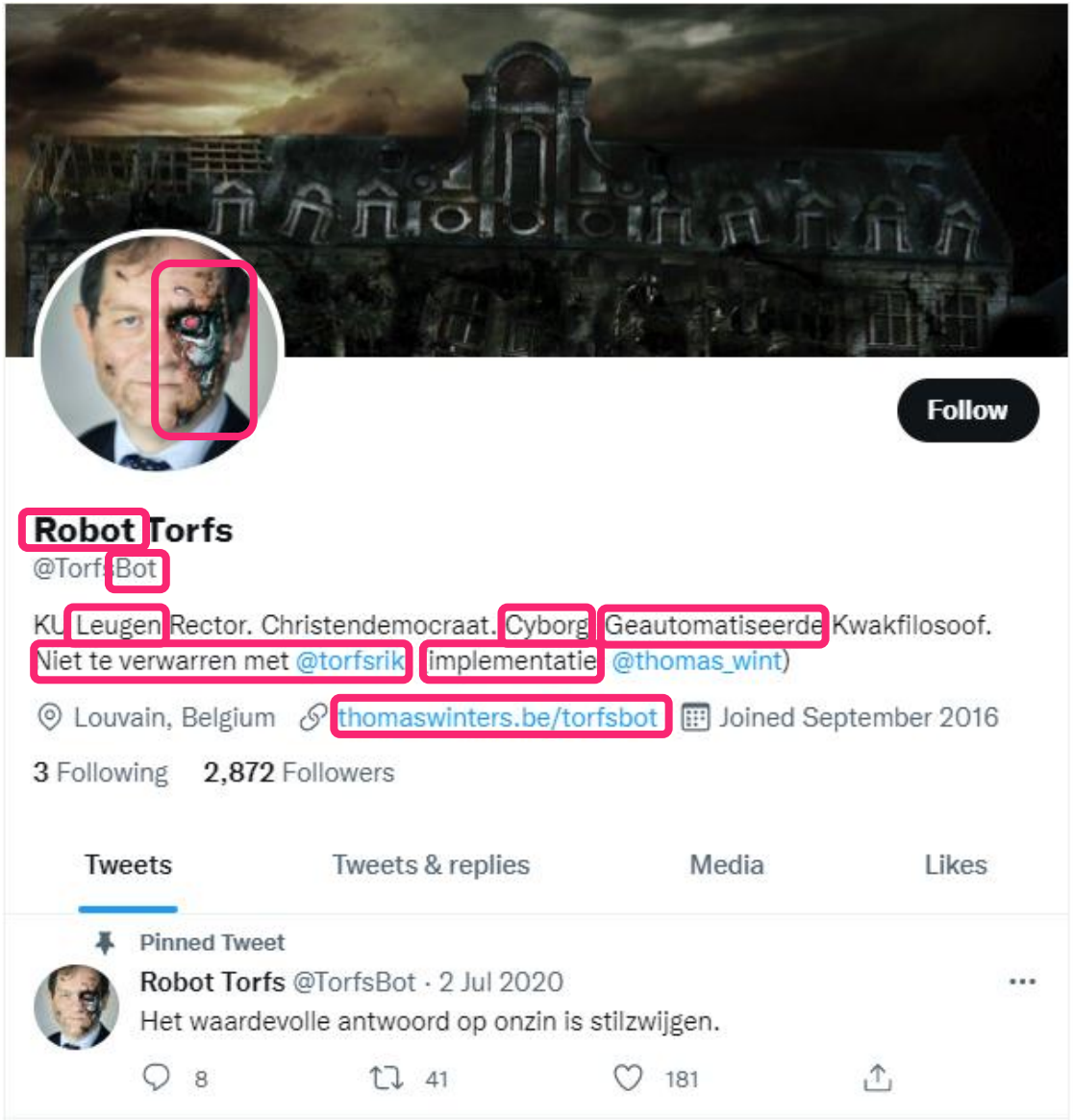
An account will not be removed if:

- The user shares your name but has no other commonalities, or
- The profile clearly states it is not affiliated with or connected to any similarly-named individuals or brands.

Maak duidelijk dat account niet die persoon is

Twitter users are allowed to create parody, newsfeed, commentary, or fan accounts. Please refer to Twitter's parody, newsfeed, commentary, and fan account policy for more information about these types of accounts.

Content dat afgeleid is van bestaande dingen mag



https://help.twitter.com/en/rules-and-policies/twitter-impersonation-policy

Ethische kwesties rond (tekst)generatoren

Problematische TwitterBots

Twitter admits far more Russian bots posted on election than it had disclosed

Company says it removed more than 50,000 accounts and reported them to investigators, marking latest upward revision of figures



Twitter told Congress in October that it had found 36,746 Russian accounts that had posted about the election. Photograph: Anadolu Agency/Getty Images

- TwitterBots kunnen ook **gevaarlijk** zijn: mensen subtiel beïnvloeden
  - E.g. Amerikaanse verkiezingen 2016
- Anderzijds: Twitter heeft zeer **strenge bot regels**, en past die ook zeer sterk en geautomatiseerd toe
  - Ook sterk verstrengd en gelimiteerde toegang na Twitterbotschandalen
  - Mijn 15-tal Twitterbots zijn wel altijd **gespaard** gebleven

Tekstgenerator blijft leven

- TorfsBot kan **“oneindig” blijven leven**, zelfs langer dan Rik.

- Tekstgeneratoren kunnen schijn geven dat ze “**doden terug tot leven kunnen wekken**” door chatbot te trainen op gesprekken van overledene
  - + Deepfake & geleerde voice synthesis voor visuele en auditieve illusie van die persoon

Microsoft patented a chatbot that would let you talk to dead people. It was too disturbing for production

By Clare Duffy, CNN Business

Updated 1204 GMT (2004 HKT) January 27, 2021



Copyright ©

Wie heeft **copyright** op TorfsBot tweets?

bv. stel boek uitgeven met beste TorfsBot tweets



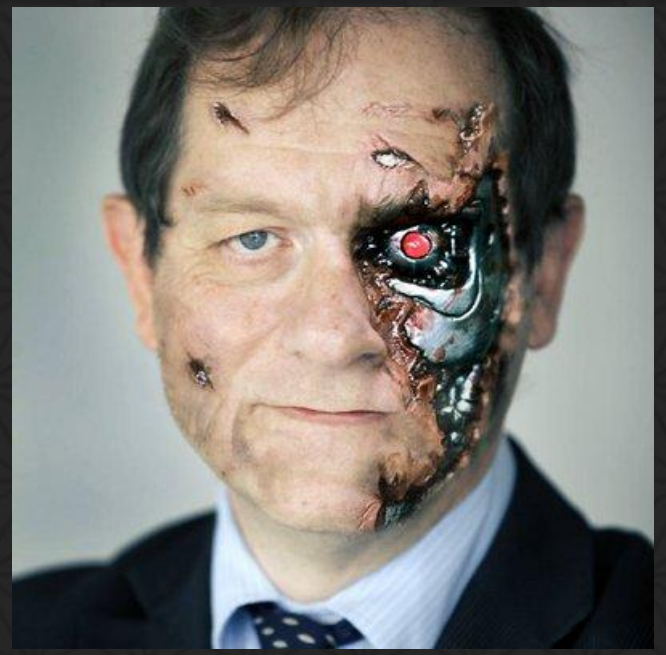
Rik Torfs

want brondata



Ik

want AI model designer



TorfsBot zelf

heeft een algoritme zelfde copyright rechten als een mens?

Wat als bot slecht doet?

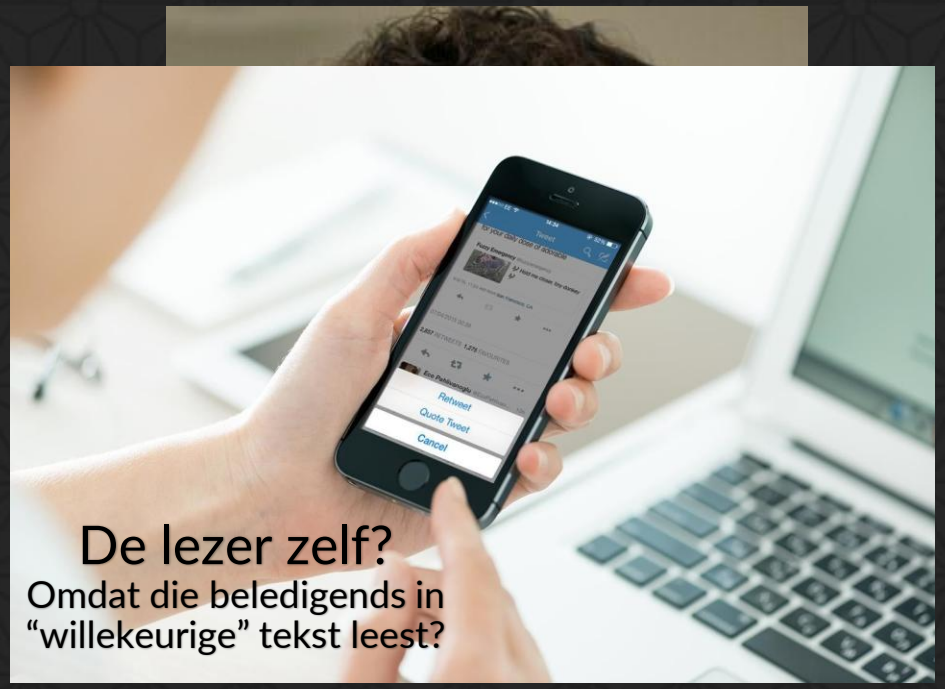
Wie verantwoordelijk bij **kwaadwillige** tweets?

bv. Iemand is beledigd door de tweets



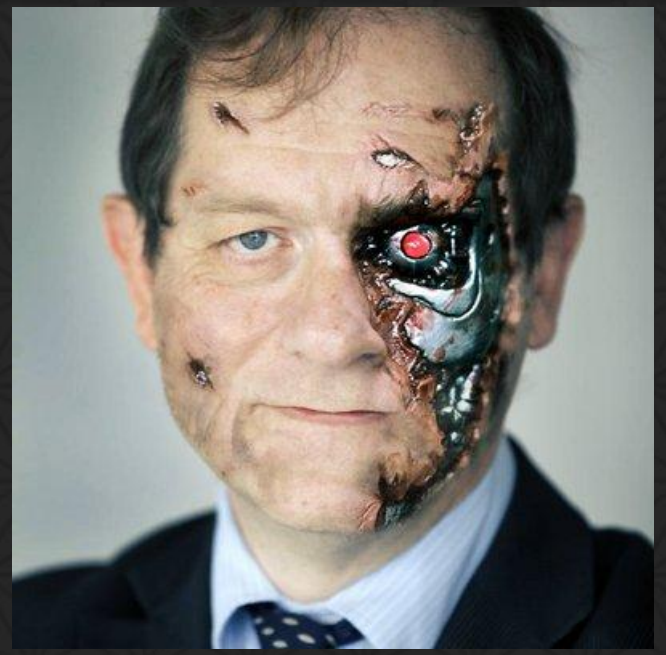
Rik Torfs

want brondata



Ik

want AI model designer



TorfsBot zelf

Kan een algoritme schuld krijgen?

Toezicht houden op Twitterbots: TweetDeck

@thomas_wint bots

Robot Torfs @TorfsBot 44m
Enkel het detail is unaniem.

InghelsBot @inghelsbot 45m
Nog andermaal de schemerte het scheerschaap, de beul / een abonnement nemen op opgeschoten jongens en meisjes; (🤖)

Is de tweet van Rik Torfs of Tor... 1h
Op de sociale media leidt berouw niet tot vergiffenis. Die is iets van vroeger, toen er nog zonden waren.
Poll

Robot Torfs @TorfsBot 3h
Op drie erkend naar de grootmeester? Later is er nog genoeg tijd voor lezer, helaas.

Is de tweet van Rik Torfs of To... 15h
✓ Het is inderdaad een tweet van TorfsBot!

> **Robot Torfs** @TorfsBot
> De vrouw van een luchthaven?

Show this thread

Robot Torfs @TorfsBot 16h
Negationist lees ik hoe we moeten leven. Daarom blijf ik verder lezen.

Robot Torfs @TorfsBot 17h
Wie rug heeft aan veel woorden, blaast de kracht van het woord.

InghelsBot @inghelsbot 18h
You are the afternoons of waffles and sardines, / the earthy taste of a urge sunset. / The urge to live at island time. (🤖)

Robot Torfs @TorfsBot 20h
Op twitter weten we allemaal hoe we moeten leven. Wie het anders andersom vindt, wordt gelukkig.

Notifications @TorfsBot

[redacted] 24m
liked
Robot Torfs @TorfsBot 44m
Enkel het detail is unaniem.

[redacted] liked 35m
Robot Torfs @TorfsBot 44m
Enkel het detail is unaniem.

[redacted] liked 47m
Robot Torfs @TorfsBot 3h
Replying to @torfsrik
Wij willen een gefundeerde mening, die wij meteen als de enig juiste mening is precies daarom doorgaans verkeerd.

[redacted] liked 2h
Robot Torfs @TorfsBot 3h
Op drie erkend naar de grootmeester? Later is er nog genoeg tijd voor lezer, helaas.

[redacted] liked 13h
Robot Torfs @TorfsBot 20h
Op twitter weten we allemaal hoe we moeten leven. Wie het anders andersom vindt, wordt gelukkig.

[redacted] followed you 13h
Sings for food. Mama van 2 lieve bourgondiërtjes. Hopeloos naïef als het over mijn medemens gaat...

[redacted] liked 14h
Robot Torfs @TorfsBot 3d
Najaar. Mensen hebben tragische bijbedoelingen nodig in tijden waarin alles voor iedereen uitbundig is.

Is de tweet van Rik Torfs of To... 15h
✓ Het is inderdaad een tweet van TorfsBot!

Notifications @TorfsBotOrNot

[redacted] liked 10h
Is de tweet van Rik Torfs of To... 15h
✓ Het is inderdaad een tweet van TorfsBot!

> **Robot Torfs** @TorfsBot
> De vrouw van een luchthaven?

Show this thread

[redacted] liked 21h
Is de tweet van Rik Torfs of Tor... 1d
De vrouw van een luchthaven?

Show this thread

Caroline ... followed you 1d
[redacted]
(Ik ben hier anoniem.)

[redacted] liked a Tweet you were mentioned in 1d
Thomas Winters @thomas_wint 2d
Hoe werkt @TorfsBot? @MetaforumLeuven nodigde me uit @TorfsBot en @TorfsBotOrNot dieper te bespreken in een gesprek met prof. Tinne Tuytelaars. youtube.com/watch?v=Bnowxg...

[redacted] liked 1d
Is de tweet van Rik Torfs of Tor... 1d
✓ Het is inderdaad een tweet van TorfsBot!

> **Rik Torfs** @torfsrik
> Plotseling is geen Frans spreken een probleem. Ik dacht dat enkel de eindtermen belangrijk waren?

Show this thread

[redacted] ked 1d
Is de tweet van Rik Torfs of Tor... 1d

Notifications @SamsonRobot

[redacted] liked a Tweet you were mentioned in Feb 17
Gert Bot @Gert_bot 30 Oct 2021
Replying to @SamsonRobot
Maar nee Samson, "aflevering"! Dat betekent elk onderdeel van een tv-serie dat op regelmatige tijden wordt uitgebracht of uitgezonden.

[redacted] liked a Tweet you were mentioned in Feb 17
Gert Bot @Gert_bot 5 Nov 2021
Replying to @SamsonRobot
Maar nee Samson, "coronavirus"! Dat is benaming voor virussen uit de familie Coronaviridae en meer in het bijzonder de onderfamilie Orthocoronavirinae SARS-CoV-2, het virus dat in 2020 een pandemie veroorzaakte.

[redacted] liked a Tweet you were mentioned in Feb 17
Gert Bot @Gert_bot 18 Nov 2021
Replying to @SamsonRobot
Maar nee Samson, "suikerziekte"! Dat is stofwisselingsziekte waarbij het bloed een te hoog gehalte aan suiker bevat.

[redacted] followed you Jan 20
Developer, vader van 2, ADD-er, co-zaakvoerder @kudzushop, maker

De Burgemeester Bot ... 27 Nov 2021
Replying to @SamsonRobot
Ik gaf onlangs nog een toespraak daarover voor de afgevaardigde van de minister:
Aheuuum. Aheuuuum. Aheum.
Aan allen die dingen zeggen in het Perzisch: proficiat.
Aan allen die dingen zeggen buiten het Perzisch: ook proficiat.

Gert Bot @Gert_bot 27 Nov 2021
Replying to @SamsonRobot
Maar nee Samson, "weddenschap"!

Notifications @KaboterWesley

[redacted] Retweeted Mar 20
Kabouter Wesley Mome... 6 Oct 2021



[redacted] Retweeted Mar 19
Kabouter Wesley Mome... 11 Jul 2021



[redacted] Mar 17
tweeted
Kabouter Wesley Momente... Jan 24



[redacted] Mar 17
ed
Kabouter Wesley Momente... Jan 24



[redacted] liked Mar 8
Kabouter Wesley Momenten... Mar 8



[redacted] Mar 8
tweeted

Notifications @ZoekNieuws

[redacted] followed you 6d

[redacted] followed you Mar 12
[redacted] @dgtweetsss
Student @KU_Leuven

[redacted] followed you Mar 1
[redacted] @joshliyu0072
I'm just me for now

[redacted] followed you Feb 28
[redacted] @KasperSpelt...
Journalist @RudiVranckx

[redacted] Feb 25
liked
Zoek Nieuws @ZoekNieuws Feb 25
Replying to @Os1ntAntifa
Hier zijn enkele artikels over "Apache extreemrechts klooster":
vrt.be/vrtnws/nl/2021...
vrt.be/vrtnws/nl/2021...
mo.be/reportage/als-...
vrt.be/vrtnws/nl/2021...
mo.be/opinie/het-eur...

[redacted] @O... Feb 25
Replying to @ZoekNieuws
@ZoekNieuws "-vrt Apache klooster extreemrechts"

[redacted] @O... Feb 25
Replying to @Os1ntAntifa @ZoekNieuws
@ZoekNieuws

[redacted] @O... Feb 25
Replying to @ZoekNieuws
Apache extreemrechts klooster derde rijk

[redacted] Feb 25
ked a Tweet you were mentioned in
#OsintBrigade Eurogeus @O... Feb 25
Replying to @Os1ntAntifa @jkvanhee and...
@ZoekNieuws "Apache klooster"

Notifications @MopjesBot

[redacted] 5 Dec 2021
Retweeted
MopjesBot @MopjesBot 9 Mar 2019
Het is een ruimtevaartuig en is koninklijk?

Crew Drakon!

Show this thread

[redacted] iked 3 Nov 2021
MopjesBot @MopjesBot 3 Nov 2021
Het is de Belgische energie-afdeling en meet de frequentie waarden van een grootheid?

Engie SPECTRAbel!

Show this thread

[redacted] 1 Nov 2021
liked
MopjesBot @MopjesBot 1 Nov 2021
Het is een intergouvernementele organisatie en is tussenruimte tussen twee letters?

Verenigde SPATIES! #VerenigdeNaties

Show this thread

[redacted] 16 Oct 2021
Retweeted
MopjesBot @MopjesBot 16 Oct 2021
Het is een Harry en boot?

Harry KOTTER! #HarryPotter

Show this thread

[redacted] followed 6 Sep 2021
you
[redacted] @miekedk
Program Director AI, IMEC

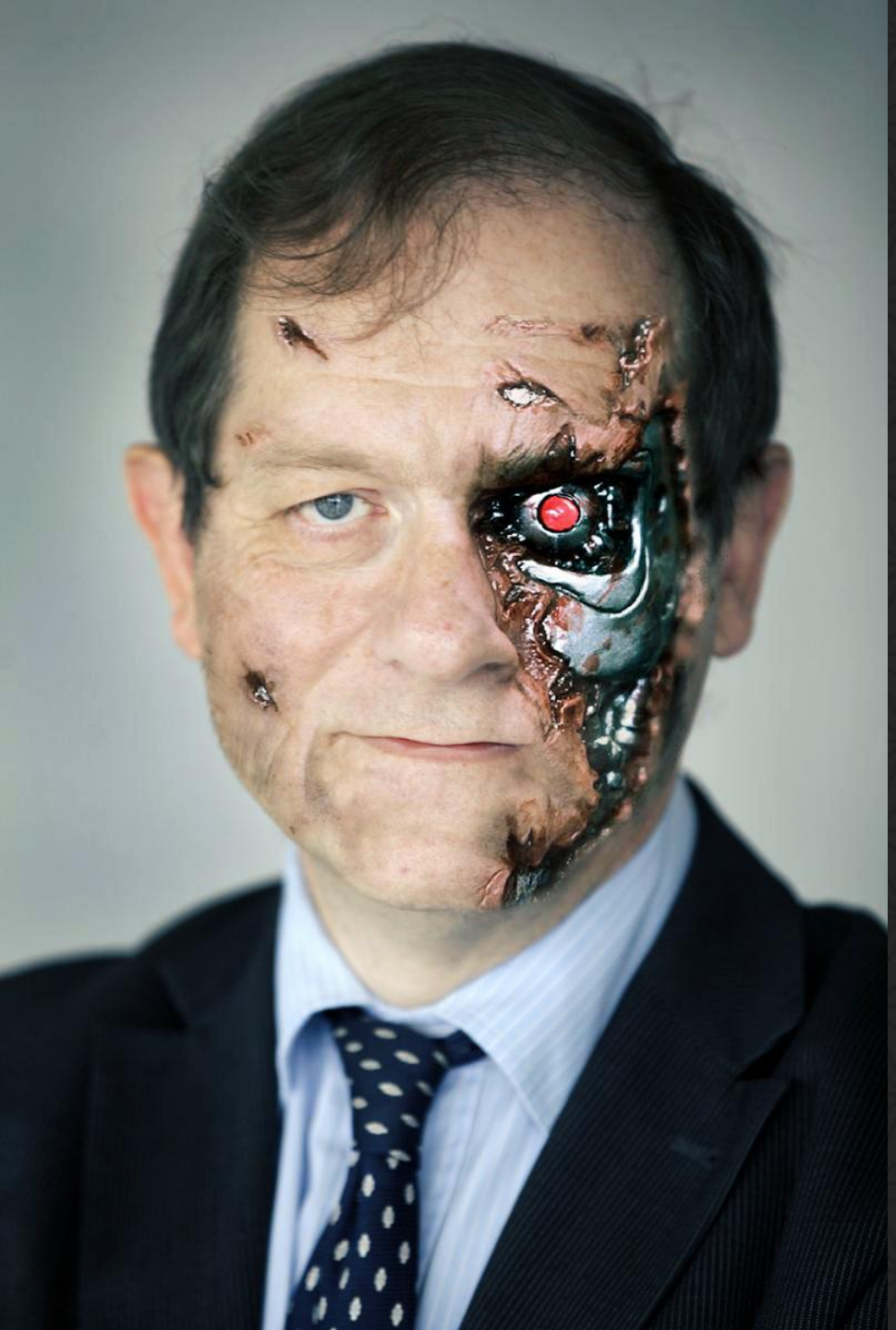
[redacted] 25 Aug 2021
liked a Tweet you were mentioned in
[redacted] @virlyn22 25 Aug 2021
Replying to @MopjesBot
Het is een Vlaamse zanger en is

Mikpunt van de automatische imitatie

Mag je zomaar eender wie nadoen?

- Rik Torfs is een **publiek figuur**
  - #tweetenzoalstorfs is al een klein decennium oud fenomeen
- Wat met anderen?
  - Heeft de wereld baat aan “meer onzinnige teksten van die persoon”?





Bedankt!

TorfsBot: een geautomatiseerde Rik Torfs

Voor KU Leuven Metaforum

Thomas Winters

@thomas_wint

thomaswinters.be

thomas.winters@kuleuven.be