

Hoe controleer je AI-tools?

Designing
Journalism



> FOR SOCIETY



Jessy de Cooker

Fontys Journalistiek

VVOJ Regiomiddag, 8 oktober 2024

Hoe controleer je AI-tools?

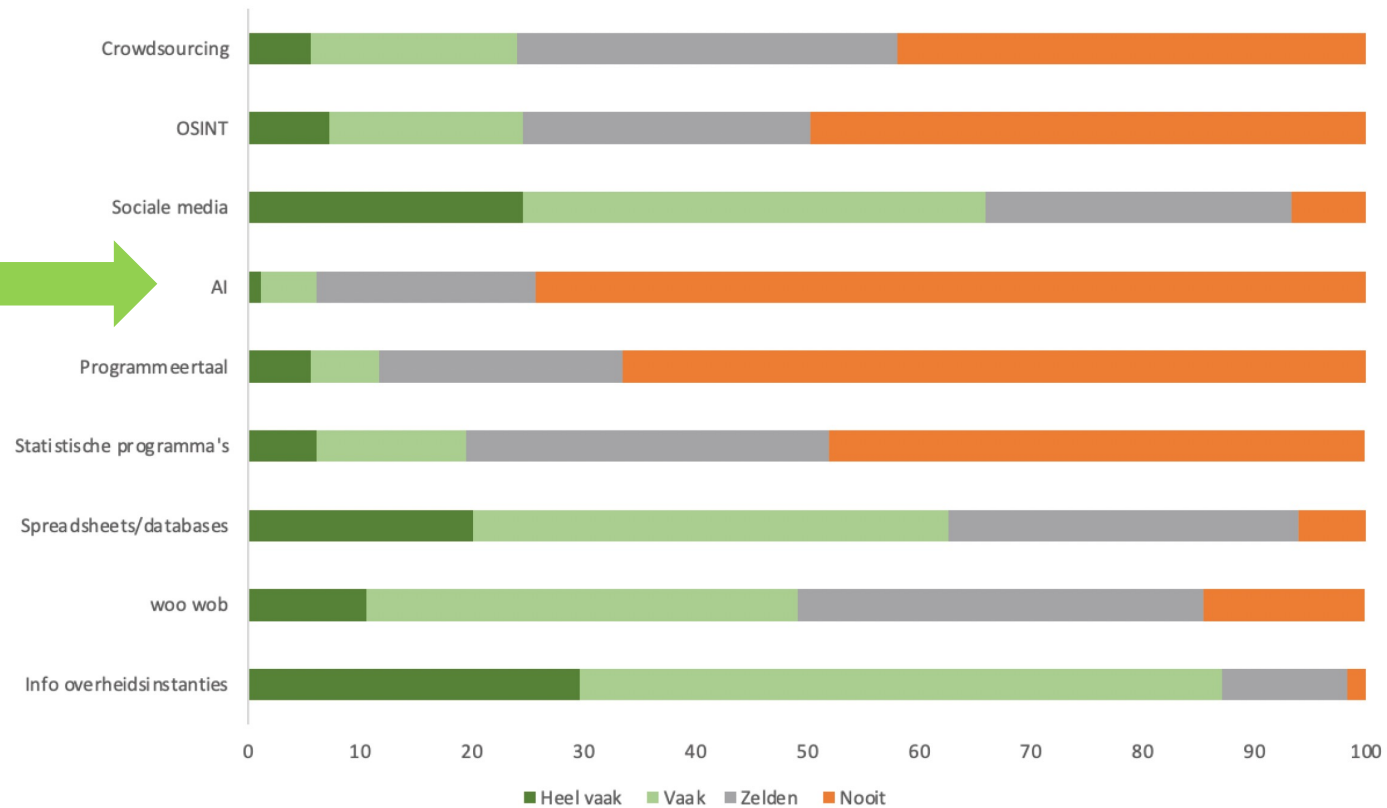
AI-onderzoeksjournalistiek in het algemeen: keuze voor bepaalde AI-methode (supervised learning, bijvoorbeeld), handmatig labelen van onderzoeksdata door ‘menselijke’ journalisten

Generatieve AI: Abstract, mede door controle van output

Designing Journalism



> FOR SOCIETY



Bron: 'Onderzoeksjournalistiek in Nederland en Vlaanderen anno 2022' (VVOJ-rapport)

Doel

Inzicht krijgen in de **stappen die je kunt zetten bij de keuze** van een AI-tool en **welke afwegingen** daarbij kunnen horen.

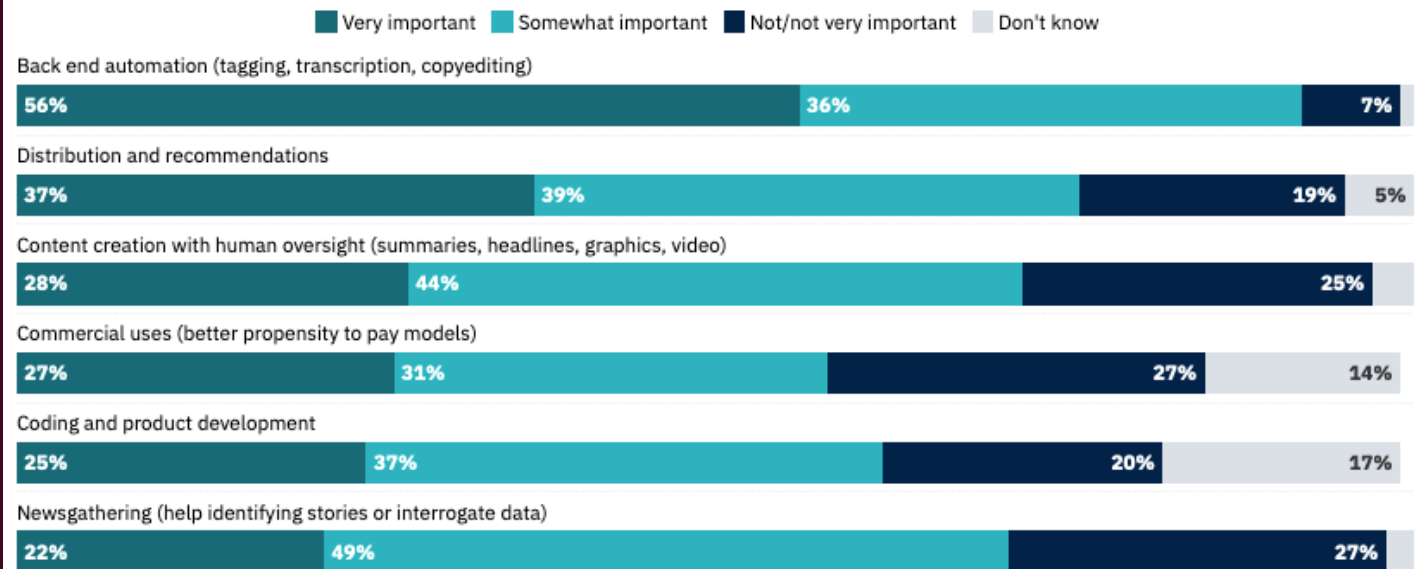
Designing Journalism



We zoeken binnen de journalistiek momenteel naar **praktische taken** waarbij AI een rol kan spelen.

AI uses which will be key for news publishers in 2024

Back end automation and content creation are now much more important than two years ago. Data from a survey of 314 media leaders in 56 countries for 'Journalism, media and technology trends and predictions 2024'.



Q11. To what extent will the following uses of Artificial Intelligence (AI) and Generative AI be important to your company in 2024? Base: 296.

Source: Journalism, media and technology trends and predictions 2024.

[Get the data](#) • [Embed](#)



reutersinstitute.politics.ox.ac.uk/journalism-media-and-technology-trends-and-predictions-2024

Designing Journalism



Verhogen van de efficiëntie en de productiviteit als belangrijkste doelstellingen binnen de journalistiek.



Designing Journalism



Er komt mogelijk nóg meer content (en nieuwsmijding).

Nóg meer datahonger van grote techbedrijven

‘Generative search’ leidt mogelijk tot mindere vindbaarheid van journalistieke content.

61%

of Gen Z and 53% of Millennials reported that they are using AI tools in place of search engines (like Google) to find information about a topic.

<https://www.theverge.com/pressroom/2024/4/26/24139468/artificial-intelligence-survey-data>

Designing Journalism



> FOR SOCIETY



How news publishers consider fighting news avoidance

Data from a survey of 314 media leaders in 56 countries for 'Journalism, media and technology trends and predictions 2024'.

Widespread acceptance of these



Less enthusiasm about these



Q19. The following is a list of approaches that some media companies are looking at to combat news avoidance/fatigue. How important, if at all, does your company consider each of these initiatives to be this year? Base: 300

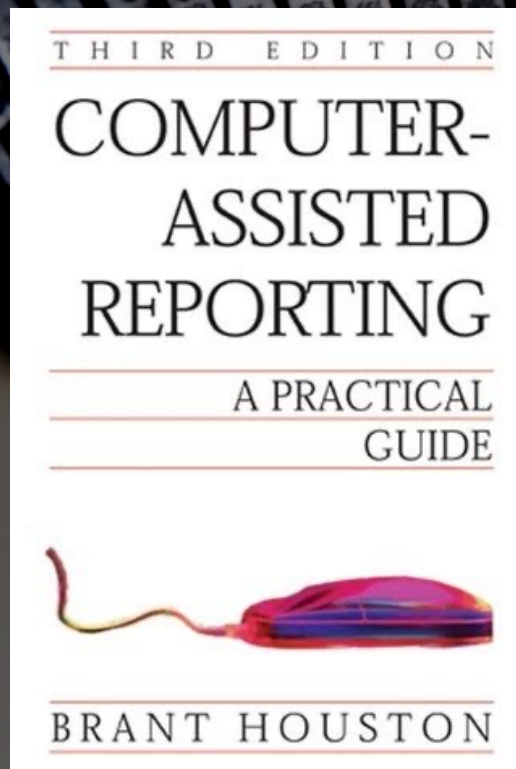
Source: Journalism, media and technology trends and predictions 2024.

[Get the data](#) • [Embed](#)



reutersinstitute.politics.ox.ac.uk/journalism-media
and-technology-trends-and-predictions-2024

Designing Journalism



Copy from the web and paste to Excel

	A	B	C	D	E	F	G	H	I
	Donor	Party	Donor	Donor Status	Date accepted	Donation			
1	Mr Wendy Alexander MSP	Labour	Scottish Industry Forum	Unincorporated Association	25/11/2002	£10,000.00			
2	Mr Wendy Alexander MSP	Labour	Scottish Industry Forum	Unincorporated Association	31/03/2003	£2,000.00			
3	The Rt Hon Michael Acrean QC MP	Conservative	Mr Henry Kewick	Individual	29/06/2001	£5,000.00			
4	The Rt Hon Michael Acrean QC MP	Conservative	Mr John Spurling	Individual	29/06/2001	£10,000.00			
5	The Rt Hon Michael Acrean QC MP	Conservative	Mr Bruce Laidlaw	Individual	06/07/2001	£10,000.00			
6	The Rt Hon Michael Acrean QC MP	Conservative	Lord Hans of Pechomy	Individual	12/07/2001	£5,000.00			
7	The Rt Hon Hilary Benn MP	Labour	Mr Mark Paul McDonald	Individual	27/02/2007	£2,500.00			
8	The Rt Hon Hilary Benn MP	Labour	Mr Kevin G McGrath	Individual	31/03/2007	£10,000.00			
9	The Rt Hon Hilary Benn MP	Labour	Mr Glad Haysem	Individual	26/04/2007	£5,000.00			
10	The Rt Hon Hilary Benn MP	Labour	Lord Pamy Mitchell	Individual	06/05/2007	£2,000.00			
11	The Rt Hon Hilary Benn MP	Labour	Mr Richard Anthony Nelson	Individual	06/06/2007	£2,000.00			
12	The Rt Hon Hilary Benn MP	Labour	Mr Leonard Coltrane	Individual	13/06/2007	£2,500.00			
13	The Rt Hon Hilary Benn MP	Labour	Sir Sigmund Stenborg	Individual	17/06/2007	£2,000.00			
14	The Rt Hon Hilary Benn MP	Labour	Mr David Abraham	Individual	22/06/2007	£5,000.00			
15	The Rt Hon Hazel Blears MP	Labour	Wilkinson Star Ltd	Company	06/03/2007	£10,000.00			
16	The Rt Hon Hazel Blears MP	Labour	Mr's Isabel Harnock	Individual	16/05/2007	£2,500.00			
17	The Rt Hon Hazel Blears MP	Labour	Shorelands Business Services Ltd	Company	17/05/2007	£10,000.00			
18	The Rt Hon Hazel Blears MP	Labour	Mr Anthony Bailey	Individual	17/05/2007	£10,000.00			
19	The Rt Hon Hazel Blears MP	Labour	B M Creative Management Ltd	Company	26/05/2007	£5,000.00			
20	The Rt Hon Hazel Blears MP	Labour	Lozan Properties Ltd	Company	26/05/2007	£10,000.00			
21	The Rt Hon Hazel Blears MP	Labour	Unicef of Shop Distributors and Allied Workers	Trade Union	26/05/2007	£12,000.00			
22	The Rt Hon Hazel Blears MP	Labour	Mr Jim Ramsbottom	Individual	18/06/2007	£5,000.00			
23	The Rt Hon Hazel Blears MP	Labour	Mr Mahmood Khayem	Individual	26/06/2007	£5,000.00			
24	The Rt Hon Gordon Brown MP	Labour	Lord Sushantha Shattacharyya	Individual	6/05/2007	£25,000.00			
25	The Rt Hon Gordon Brown MP	Labour	Mr Paul Myers	Individual	06/05/2007	£9,750.00			
26	The Rt Hon Gordon Brown MP	Labour	Sir Sigmund Stenborg	Individual	14/05/2007	£10,000.00			
27	The Rt Hon Gordon Brown MP	Labour	Lord Alexander Leitch	Individual	21/05/2007	£5,000.00			
28	The Rt Hon Gordon Brown MP	Labour	Lord John Gageon	Individual	21/05/2007	£10,000.00			
29	The Rt Hon Gordon Brown MP	Labour	Community	Trade Union	21/05/2007	£10,000.00			
30	The Rt Hon Gordon Brown MP	Labour	Mr John Walsley	Individual	21/05/2007	£2,000.00			
31	The Rt Hon Gordon Brown MP	Labour	Carson Industries Plc	Company	21/05/2007	£10,000.00			

Onderzoeksjournalisten & AI (in Nederland)

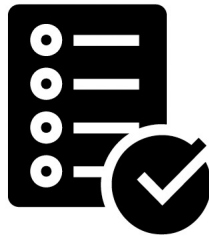
In het onderzoeksprogramma It's A Match onderzocht Fontys Journalistiek samen met werkveldpartners Pointer, Beeld & Geluid en Omroep Flevoland AI-gebruik en -kansen binnen Nederlandse onderzoeksjournalistiek. Drie inzichten uit interviews met nieuwsorganisaties (N=25):

- 1) AI wordt voorzichtig omarmd in de onderzochte nieuwsredacties voor onderzoeksjournalistieke inzet.
- 2) De journalisten die we spraken, afgezien van een kleine groep koplopers, uiten onzekerheid over de keuze van AI-tools en hebben moeite om AI-gedreven onderzoek volledig te begrijpen.
- 3) Tijd (om 'echt te verdiepen' in AI-mogelijkheden en structurele ondersteuning werden genoemd als mogelijke oorzaken.

“Ik had niet de luxe om een week lang uit te zoeken hoe ik dit moest doen.”

(Regionale onderzoeksjournalist)

How ready do you think your company
is to take advantage of AI's potential?



13%

well prepared

45%

somewhat prepared

33%

poorly prepared



7%

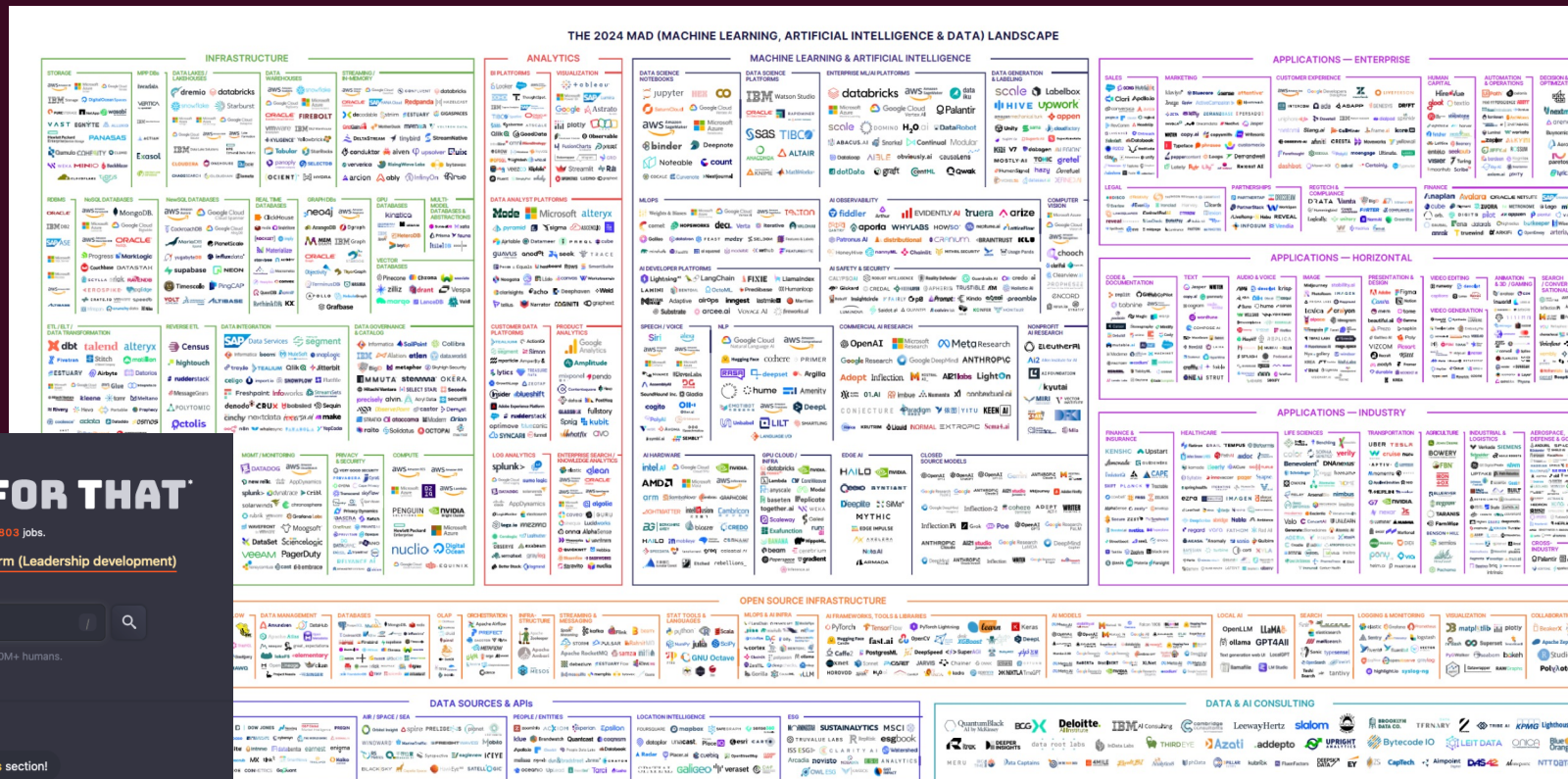
not prepared at all

“We zitten gevangen in een magisch doolhof. Binnenin is het moeilijk om hype van realiteit te onderscheiden.”

Ezra Eeman – NPO Innovatie

Tijdens AI in Media World News Media Congress 24

Designing Journalism



*TAAFT for short
THERE'S AN AI FOR THAT
18,183 AIs for 14,939 tasks and 4,803 jobs.
Spotlight: Ren AI | Leadership & Accountability Platform (Leadership development)

Find AIs using AI

#1 AI aggregator. Updated daily. Used by 30M+ humans.

Home For You

Check out our brand new events section!

Just Launched

Clip Studio
Short videos
Free + from \$14/mo
3.3 ★ 2.3

TaroTap
Tarot card readings
Free + from \$9.9/mo
1.7 ★ 5.0

Blend AI
Text & Image genera...
Free + from \$7/mo
1.292 ★ 4.4

Relately
Communication impr...
Free
1.4

SoBrief
Book summaries
Free + from \$3.75/mo
1.9 ★ 5.0

Pulse Hero
User feedback analy...
Free + from \$29/mo
1.1

ElectionMap

leania.ai
v1.10 released 1d ago

Ren AI | Leadersh...
v1.10 released 20h ago

mad.firstmark.co

“Journalistiek begint niet met de vraag: “We hebben een tool. Het begint altijd met een verhaal of voorbeeld van een bepaald soort misbruik of gebrek. Als je kennis hebt van AI, dan weet je welke tools je op welk moment in je onderzoekswerk moet gebruiken.”

**(Onderzoeksjournalist
nationaal medium)**

Stappenplan voor toolkeuze

- ‘Context van de tool begrijpen’
- ‘Geschiktheid van de tool kritisch evalueren’
- ‘Documenteren hoe tools worden gebruikt tijdens onderzoek’
- ‘Kritisch reflecteren’

Context begrijpen

Wie/wat is de maker van deze tool?

Wat zijn (mogelijke) redenen voor het (gratis) beschikbaar stellen van deze tool?

Welke beperkingen heeft deze tool?

Context begrijpen: voorbeeld

ChatGPT-eigenaar: OpenAI

Gratis en betaald account.

Gratis betekent meestal: betalen met je (gebruikers)data.

Richtlijn Fontys Journalistiek:
geen persoonsgegevens en geen
onder embargo zijnde informatie
'openbaar maken'.

Updated: November 14, 2023

Terms of use

Content

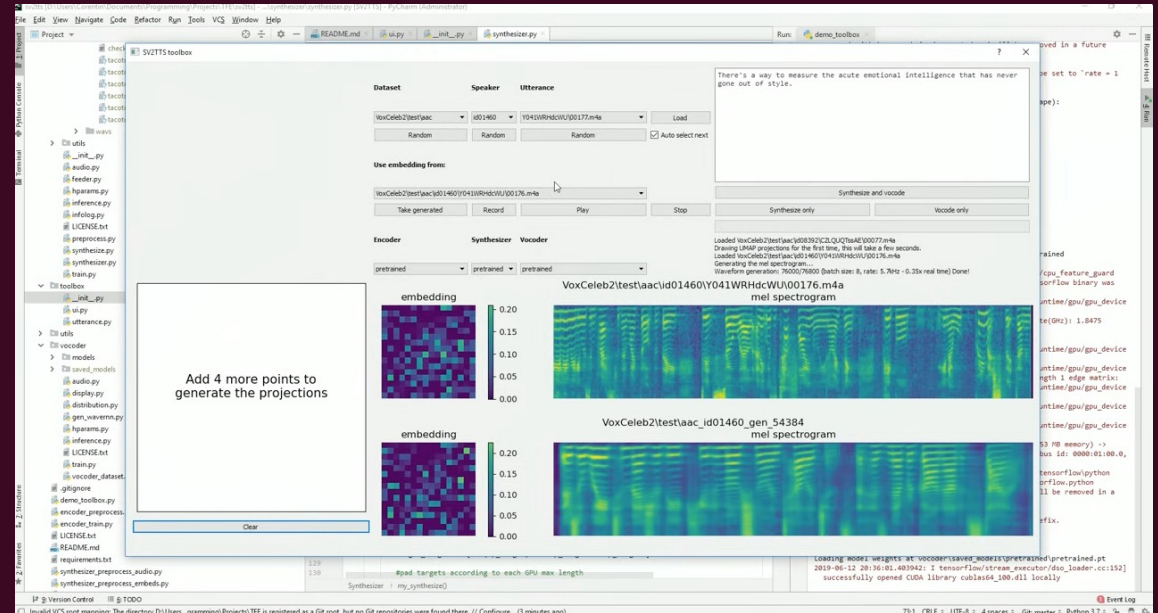
Your content. You may provide input to the Services ("Input"), and receive output from the Services based on the Input ("Output"). Input and Output are collectively "Content." You are responsible for Content, including ensuring that it does not violate any applicable law or these Terms. You represent and warrant that you have all rights, licenses, and permissions needed to provide Input to our Services.

Our use of content. We may use Content to provide, maintain, develop, and improve our Services, comply with applicable law, enforce our terms and policies, and keep our Services safe.

Context begrijpen: voorbeeld

Corentine Jemine ontwikkelde tijdens zijn masterscriptie aan de Universiteit van Luik een stemkloon-tool.

Hij stelde zijn hele systeem op Github open-source beschikbaar voor verder gebruik en ontwikkeling. Github is een platform waar programmeurs code met elkaar delen.



<https://github.com/CorentinJ/Real-Time-Voice-Cloning>

Context begrijpen: voorbeeld

Pointer wilde in 2022 onderzoek doen naar nepprofielen op datingapp Tinder. Maar hoe verzamel je die profielen?

Via Github stuitten ze op Frederik Mees, een student Computer Science, die met een Tinderbot profielinformatie verzamelt.

Pointerjournalist Jerry Vermanen contacteert Frederik, waarna ze de tool testen, updaten en 33.793 profielen verzamelen.

14 september 2022

Hoe we een computer hebben leren Tinderen om nepprofielen te herkennen

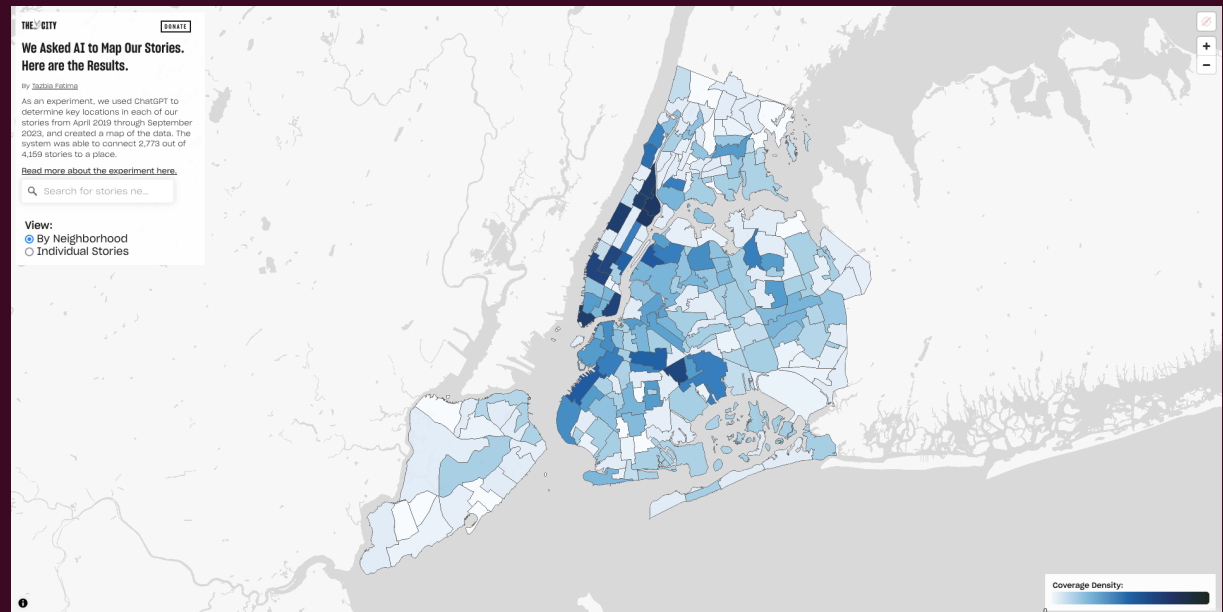
<https://pointer.kro-ncrv.nl/hoeweeneencomputerhebbenlerentinderenomnepprofielenteherkennen#>

Context begrijpen: voorbeeld

Journalisten van de New Yorkse hyperlocal The City wilden in kaart brengen of hun berichtgeving de hele metropool dekte.

Hiervoor lieten ze o.a. ChatGPT datapunten identificeren in gepubliceerde artikelen via ‘named entity recognition’.

Context begrijpen: voorbeeld



“You are a data scientist and mapper, Given this text, perform semantic analysis and return the geographical : ‘Neighborhood:’, ‘Street Name:’, ‘Landmark:’, ‘Geographical Coordinates:’. No other texts please.”

<https://tinyurl.com/TheCityGPT>

Context begrijpen: voorbeeld

ChatGPT is een Large Language Model (LLM) en staat bekend voor ‘hallucinaties’.

Met deze context in het achterhoofd werden alle resultaten handmatig gecheckt en geverifieerd via andere onderzoeksmethoden. Bijvoorbeeld door de data ook te laten analyseren door Geocoding API van Google.

Geschiktheid evalueren

Is de tool specifiek ontwikkeld voor journalistiek gebruik?

Kan ik de output van de tool handmatig checken en/of de juistheid van de resultaten inschatten?

Geschiedenis evalueren: voorbeeld

In een AI-project gebruikten journalisten uit Argentinië, Paraguay en de Filippijnen een bestaand beeld-onderschriftmodel om de namen en onderwerpen uit foto's en video's van fotojournalisten te halen.

Ze gebruikten in eerste instantie de tool Watson van IBM.

Geschiedheid evalueren: voorbeeld

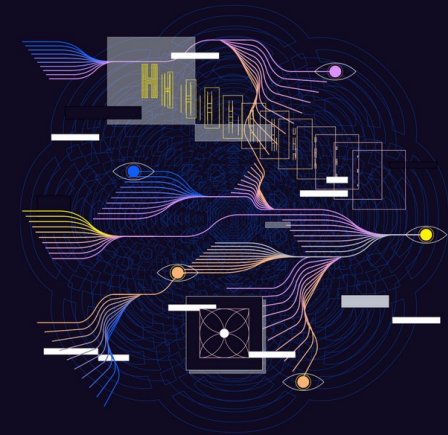


This image was labeled as "first date" by IBM Watson

iIMAGE
2TEXT

Making image recognition for
newsrooms more inclusive

ALJazeera GMA2 GRUPO OCTUBRE



De beschikbare tools waren niet gemaakt voor de journalistieke context. Dus ontwikkelden ze samen een eigen tool die nu beschikbaar is.

image2text.co

Geschiedenis evalueren: voorbeeld

Iedere output
checken in
generatieve AI door
ChatGPT of een
andere tool vragen
om stappen uit te
leggen. Hiermee
blijf je zelf
keurmeester. Check
ook alle output bij
openbare en
gesproken bronnen.

```
ChatGPT 4o v
Share 1

Je bent een opleider voor journalistiekstudenten, met een
specialisatie in computercode. Je werkt in Google Colab en hebt
een classifier gemaakt die tweets analyseert op de aanwezigheid
van transfergeruchten. Dit is je code:

import pandas as pd

# Bestand laden
file_path = '/content/drive/MyDrive/Test scraper geruchten/Scraper
test 3.xlsx'
df = pd.read_excel(file_path)

# Functie om tweets te classificeren
def classify_tweet(content):
    keywords_high = ["onderhandeling", "verkast", "zou rond zijn",
"lining up a move", "interesse van", "in beeld bij", "zou tekenen",
"bieding", "signs", "deal"]
    keywords_medium = ["in beeld", "intresse", "aan het bekijken",
"bekijken opties", "zou kunnen", "interesse", "interest", "nueva"]

    content_lower = content.lower()

    if any(keyword in content_lower for keyword in keywords_high):
        return 3 # Hoge waarschijnlijkheid van transfergerucht
    elif any(keyword in content_lower for keyword in
keywords_medium):
        return 2 # Matige waarschijnlijkheid van transfergerucht
    else:
        return 1 # Lage waarschijnlijkheid van transfergerucht

# Toepassen van classificatie op de kolom 'content'
df['score'] = df['content'].apply(classify_tweet)
```

Documenteren van je stappen en overwegingen

Welke denkstappen zet je? Beschrijf deze.

Noteer waar en hoe je voor bepaalde tools of werkwijzen kiest.

Deel dit met het veld!

Documenteren van je stappen en overwegingen: voorbeeld



Recognising 'bad actors' in data leaks with AI

Journalists are often required to find persons of interest in large sets of documents, which can be like searching for a needle in a haystack. With the help of AI we try to link relevant names in these documents to specialised databases in order to find out who deserves our attention.

A JournalismAI Project

Team

- *Luis Flores, Data Scientist, [The Guardian](#)*
- *Michel Schammel, Senior Data Scientist, [The Guardian](#)*
- *Dimitri Tokmetzis, Senior Investigative Journalist & Data Team Lead, [Follow the Money](#)*
- *Heleen Emanuel, Data Journalist & Creative Developer, [Follow the Money](#)*
- *Alet Law, Audience Development Manager, [Daily Maverick](#)*
- *Tinashe Munyuki, Retention Manager, [Daily Maverick](#)*

<https://tinyurl.com/BadActorsAI>

Documenteren van je stappen en overwegingen: voorbeeld

Object Detection met YOLOv8

YOLO (You Only Look Once) is een krachtige tool om vanalles te doen met afbeeldingen en video's. Zo kun je het gebruiken om objecten te volgen, poses van personen vast te leggen of om objecten te detecteren. In deze notebook gaan we aan de slag met het detecteren van objecten in afbeeldingen en video's. We gaan hiervoor gebruik maken van de YOLOv8 implementatie van Ultralytics. Deze implementatie is gebaseerd op PyTorch en is een van de meest accurate implementaties van YOLOv8. De implementatie is te vinden op [GitHub](#).



Het proces van het detecteren van objecten in afbeeldingen en video's kan worden opgedeeld in 3 stappen:

1. Het verzamelen van voorbeelddata
2. Het labellen van de data
3. Het trainen en valideren van het model
4. Het voorspellen van objecten in afbeeldingen en video's

In deze notebook behandelen we 2 cases:

1. Het detecteren van (omgekeerde) vlaggen en nazivlaggen
2. Het vinden van windmolens met satellietdata

1. Het detecteren van (omgekeerde) vlaggen en nazivlaggen

<https://tinyurl.com/PyItsAMatch>

Kritisch reflecteren

Waar zitten jouw eigen voorkeuren?

Waar zitten eigen menselijke bias en/of beperkingen?

Kritisch reflecteren: voorbeeld

Talking sense: using machine learning to understand quotes

The Guardian's data scientists have been working with other newsrooms on a global project to think about AI and journalism. Here they explain how they have been teaching a machine to understand what a quote is

In een samenwerking uit 2021 met het AFP probeerden datawetenschappers van The Guardian machine learning te gebruiken om citaten in documenten te identificeren en te labelen. Ze stuiten op hun eigen subjectiviteit...

Kritisch reflecteren: voorbeeld

Soms heb je
jouw eigen
‘lievelingetjes’



Extra: welke technische vaardigheid is er nodig?

Werkt de tool in je browser?

Of is deze te installeren als plugin?

Of vergt het een complexe installatie op je laptop?

Extra: welke technische vaardigheid is er nodig?

Blacklight:
Privacycontrole
Tool via website

The Markup

Blacklight

A Real-Time Website Privacy Inspector

Who is peeking over your shoulder while you work, watch videos, learn, explore, and shop on the internet? Enter the address of any website, and Blacklight will scan it and reveal the specific user-tracking technologies on the site—and who's getting your data. You may be surprised at what you learn.

www.gld.nl Options ▾ Scan Site

Visited www.gld.nl on Oct. 7, 2024, 10:25 ET Learn more ▾

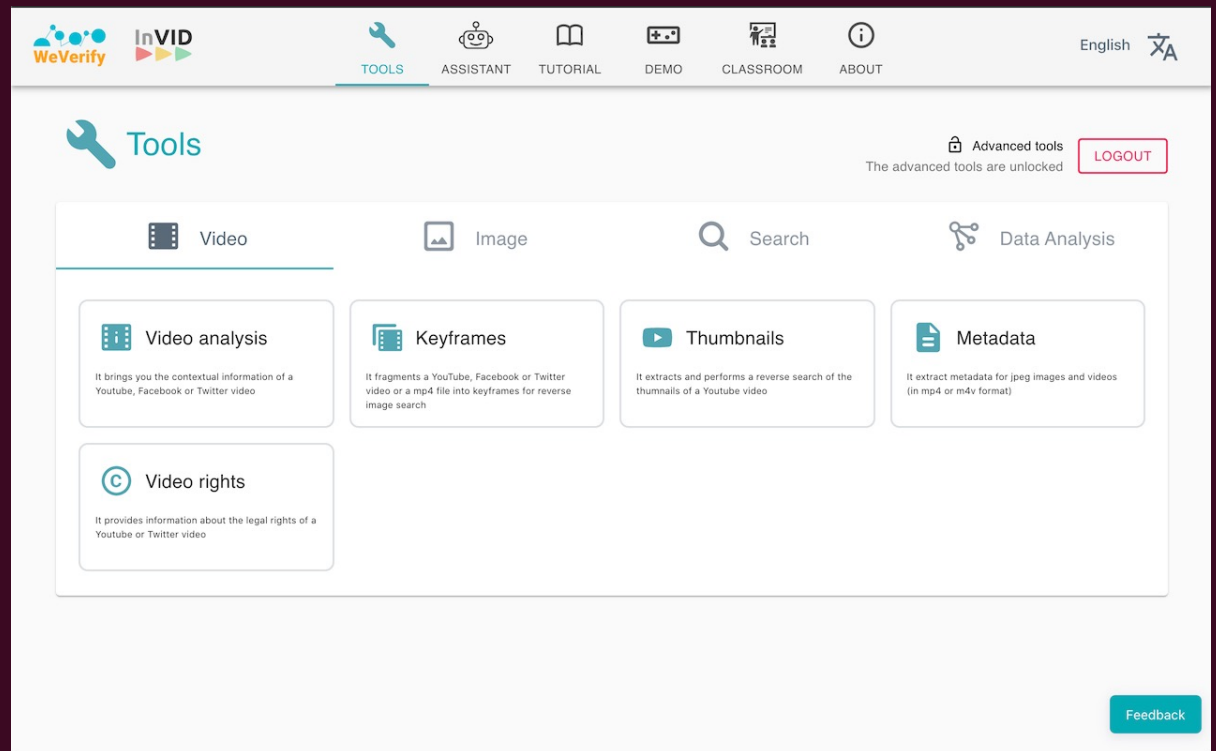
Blacklight Inspection Result

Blacklight works by visiting each website with a headless browser running custom software built by The Markup. To learn more, read our [methodology](#).

- 7** **Ad trackers found on this site.**
This is the **same as** the average of **seven** that we found on popular sites ▾
- 4** **Third-party cookies found.**
This is **more than** the average of **three** that we found on popular sites ▾
-  Tracking that evades cookie blockers wasn't found. ▾

Extra: welke technische vaardigheid is er nodig?

InVid: tools voor
beeldverificatie
**Plugin in
Google Chrome**





Extra: welke technische vaardigheid is er nodig?

Tabula: data
uit PDF's halen
**Installeren op
laptop**

Tabula



Tabula is a tool for liberating data tables locked inside PDF files.

[View the Project on GitHub](#)
tabulapdf/tabula

[Download for Windows](#) [Download for Mac](#) [View source on GitHub](#)

Current Version: [1.2.1](#)

Other Versions: [pre-releases & archives](#)

Need help? Open an [issue on Github](#).

Donate: Help support this project by [backing us on OpenCollective](#).

We'd love to hear from you! Say hi on Twitter at [@TabulaPDF](#)

Latest Version: Tabula 1.2.1
June 4, 2018

Tabula 1.2.1 fixes several bugs in the user interface and processing backend. (You can read about all the changes in the [release notes](#).)

Download Tabula below, or on the [release notes page](#).

Special thanks to [our OpenCollective backers](#) for supporting our work on Tabula; if you find Tabula useful in your work, please consider [a one-time or monthly donation](#).

How Can Tabula Help Me?

If you've ever tried to do anything with data provided to you in PDFs, you know how painful it is — there's no easy way to copy-and-paste rows of data out of PDF files. Tabula allows you to extract that data into a CSV or Microsoft Excel spreadsheet using a simple, easy-to-use interface. Tabula works on Mac, Windows and Linux.

Who Uses Tabula?

Tabula is used to power investigative reporting at news organizations of all sizes, including [ProPublica](#), [The Times of London](#), [Foreign Policy](#), [La Nación \(Argentina\)](#), [The New York Times](#) and the [St. Paul \(MN\) Pioneer Press](#).

Grassroots organizations like [SchoolCuts.org](#) rely on Tabula to turn clunky documents into human-friendly public resources.

And researchers of all kinds use Tabula to turn PDF reports into Excel spreadsheets, CSVs, and JSON files for use in analysis and database applications.

Download & Install Tabula

Windows & Linux users will need a copy of [Java](#) installed. You can [download Java here](#). (Java is included in the Mac version.)

- Download the version of Tabula for your operating system:
 - Windows: [tabula-win.zip](#)
 - Mac OS X: [tabula-mac.zip](#)
 - Linux/Other: [tabula-jar.zip](#), view README.txt inside for instructions
- Extract the zip file. (Instructions: [Windows](#), [Mac](#))
- Go into the folder you just extracted. Run the "Tabula" program inside.

Extra: welke technische vaardigheid is er nodig?

Zeeschuimer: data
verzamelen van
sociale platformen

**Installeren op laptop
+ plugin + in browser**

Wie kan je hierbij
helpen?



Zeeschuimer v1.9.0

Captured data objects

Active	Platform	Items	Actions		
☐	TikTok	44	Delete	.ndjson	to 4CAT
☐	Instagram (Posts)	0	Delete	.ndjson	to 4CAT
☐	LinkedIn	0	Delete	.ndjson	to 4CAT
☐	9GAG	0	Delete	.ndjson	to 4CAT
☐	Imgur	0	Delete	.ndjson	to 4CAT
☐	Twitter	0	Delete	.ndjson	to 4CAT
☐	Douyin	0	Delete	.ndjson	to 4CAT

[X Delete all items](#)

Connect to 4CAT

4CAT server URL: ?

Uploaded datasets

Platform	Items	Date	Dataset Link
tiktok	44	Wednesday, Mar 20, 2024	localhost
tiktok	324	Thursday, Feb 29, 2024	localhost

Designing
Journalism



> FOR SOCIETY

Takeaways

**Kritische keuzes zijn
noodzakelijk bij keuze én
controle van AI-tools.**

**Leren door te
experimenteren.**

Verder praten?



j.decooker@fontys.nl

Bezoek onze
Website



Volg onze
Instagram

Blijf up-to-date met ons
Whatsapp-Kanaal

