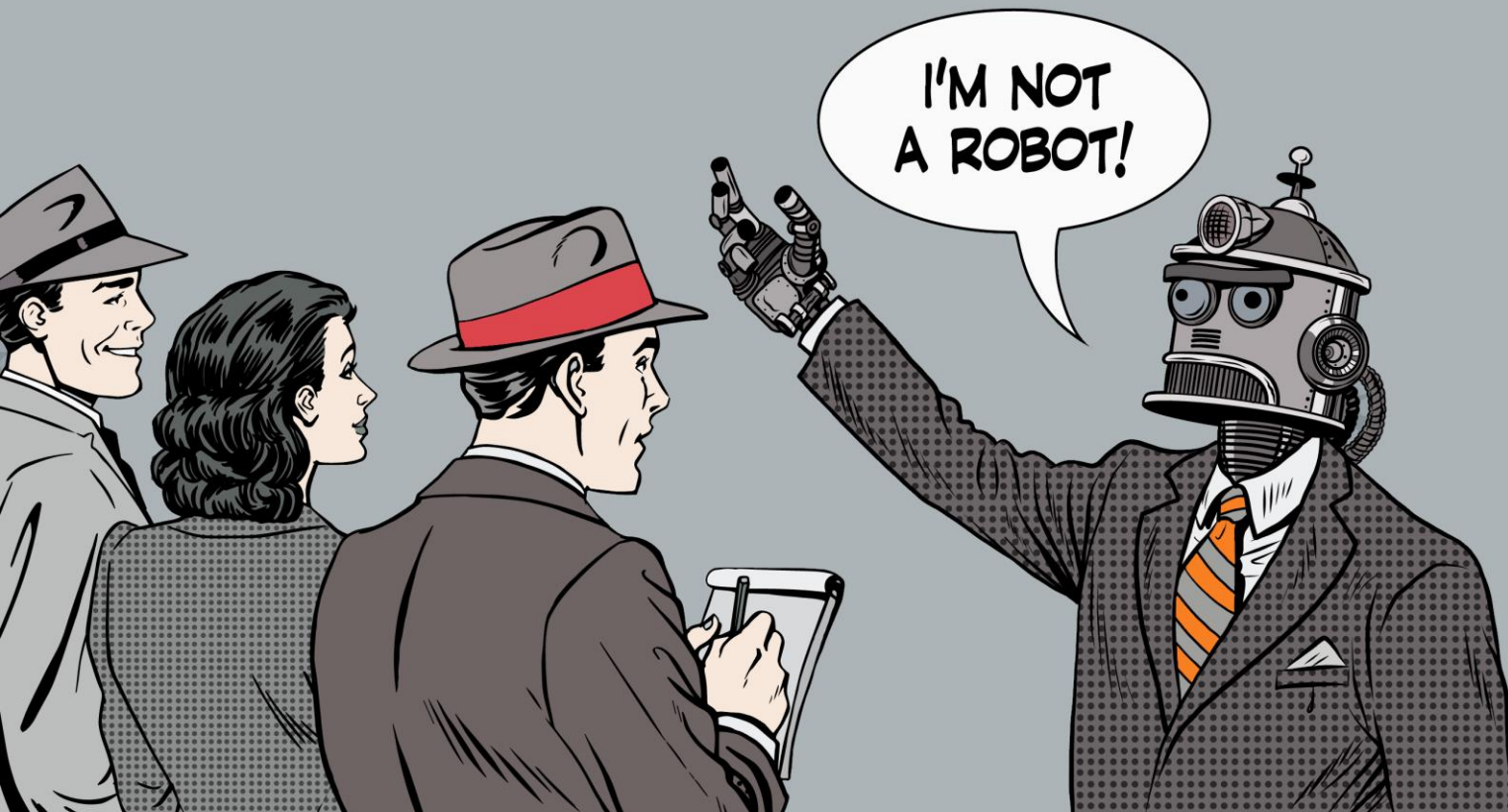


Automated journalism: an overview



Laurence Dierickx
Université Libre de Bruxelles

2018

What is automated journalism?

- A **technological process**: from data to text
- A **subfield of computational linguistics** (*first experiences of NLG in 1970s*)
- An **editorial process**: human intentions designed into computer processing

A matter of choices made by human beings



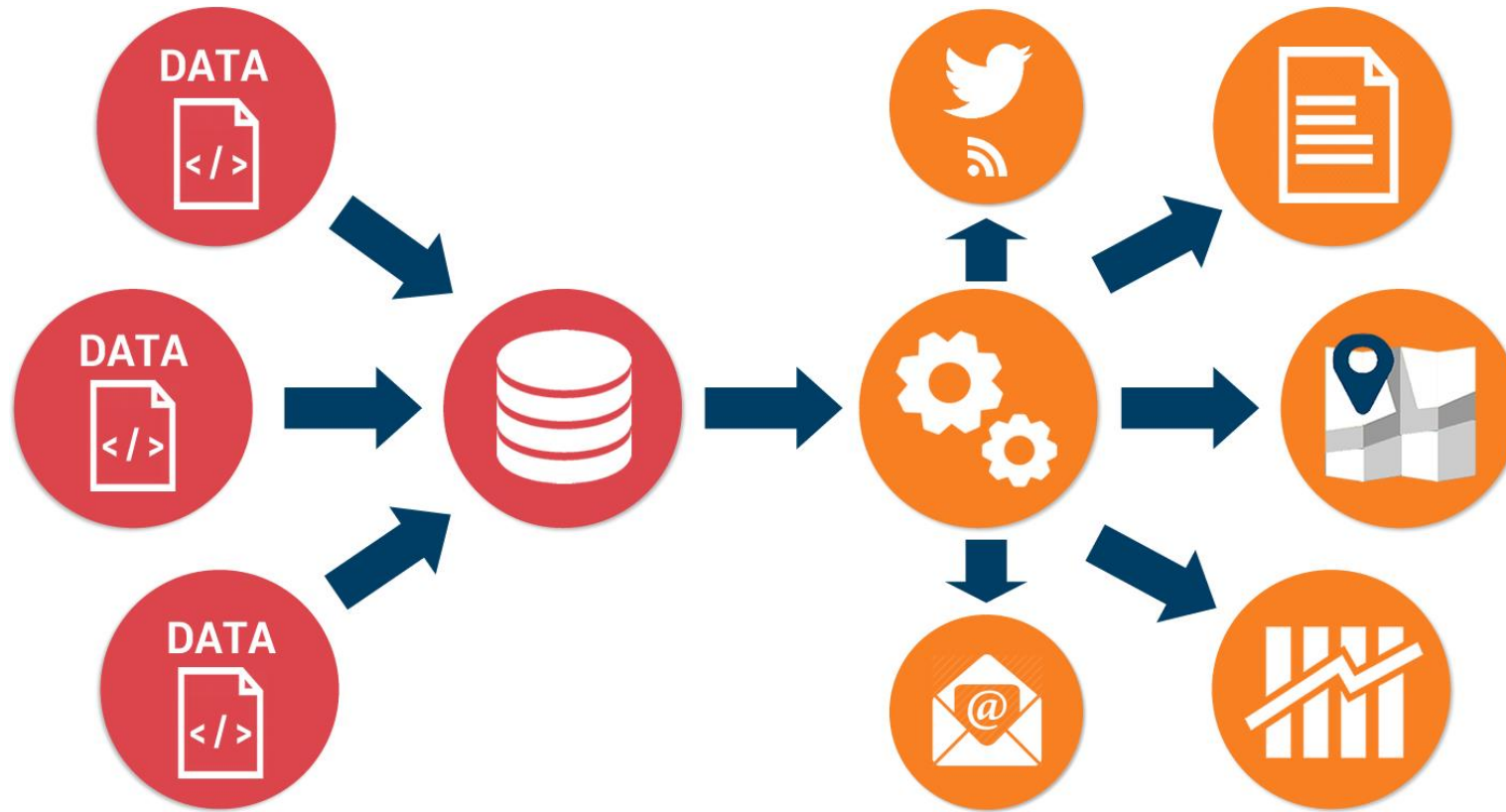
What is not automated journalism?

- A technology that will act by itself
- An uncontrolled process full of bias (*bias or errors are from humans or from bad data quality, a computer software never repeat the same mistake... if it is detected!*)
- A 3D robot even if the metaphor is widely used (*except in Asia*)



Automated journalism: an overview
Laurence Dierickx

How does it work?



WHAT TO SAY ?
HOW TO SAY IT?

What are the requirements?



Structured data

Rights to use the data

Reliable source

Data must be up to date

Good data quality = accurate, complete,...

Relevance for a journalistic use

What are the covered domains?



**Limitations are consequences of the requirements.
Everything cannot be automated!**

Is it error free?

3 reported cases

2008

Google News

Outdated financial data published with the date of the day while it was much older, announcing the bankrupt of a company. Consequences : traders sells... (impact on markets).

2015

Associated Press

Netflix released second-quarter earnings at the same time as its stock underwent a 7-to-1 split. Value were no split: misleading interpretation (71, fall of individual share).

2017

L.A. Times

Quakebot (2014): massive earthquake (magnitude 6.7), panic on social networks! This earthquake had happened well... but in 1925! Error came from wrong encoding (by a human) in the database (U.S. Geological Survey..)

Bad data value

Human errors

HUMAN CONTROL NEEDED

Which European media do use it?

PRESS AGENCIES

NEWSPAPERS

MAGAZINES

PURE PLAYERS



WE DON'T KNOW EVERYTHING!

CONFIDENTIAL DISCLOSURES
(Germany)



Readers are not ready?

Newsrooms are not ready?

Which European media do use it?

CHABLIS 8907

POPULATION EN 2012 : 17 355 habitants PRÉSIDENT DU CONSEIL GÉNÉRAL : André VILLIERS

PARTAGER  

Résultats du premier tour des élections départementales (22 mars 2015) : canton de Chablis

Le binôme constitué de Mme CHARPIGNON Sylvie et de M. GENDRAUD Patrick (Union de la Droite) est arrivé en tête du premier tour des élections départementales, dimanche 22 mars, dans le canton de Chablis avec 45,53 % des suffrages exprimés. Le tandem de l'Union de la Droite devance le binôme du Front National formé par M. BERTHOLLET Charles et Mme LAFAYE Alexandra, également qualifié pour le second tour, qui a obtenu 34,12 % des voix. M. SACKPEY Gilles et Mme VASSEUR Mireille de Divers gauche sont éliminés. Le taux d'abstention a atteint le score de 44,41 % dans ce canton.

Ces textes ont été écrits en collaboration avec Data2Content, une marque de la société Syllabs

Résultats du Premier tour

BALLOTAGE Binôme Union de la Droite

MME CHARPIGNON SYLVIE ET M. GENDRAUD PATRICK

45,53 %

BALLOTAGE Binôme Front National

M. BERTHOLLET CHARLES ET MME LAFAYE ALEXANDRA

34,12 %



Résultats du premier tour des élections départementales (22 mars 2015) : canton de Chablis

Le binôme constitué de Mme CHARPIGNON Sylvie et de M. GENDRAUD Patrick (Union de la Droite) est arrivé en tête du premier tour des élections départementales, dimanche 22 mars, dans le canton de Chablis avec 45,53 % des suffrages exprimés. Le tandem de l'Union de la Droite devance le binôme du Front National formé par M. BERTHOLLET Charles et Mme LAFAYE Alexandra, également qualifié pour le second tour, qui a obtenu 34,12 % des voix. M. SACKPEY Gilles et Mme VASSEUR Mireille de Divers gauche sont éliminés. Le taux d'abstention a atteint le score de 44,41 % dans ce canton.

Ces textes ont été écrits en collaboration avec Data2Content, une marque de la société Syllabs

RÉSULTATS DES ÉLECTIONS DÉPARTEMENTALES

Saisissez une ville ou un code postal

PUBLICITE

Automated journalism: an overview
Laurence Dierickx

ULB

Why do they use it?

To speed up content production

High speed level production: in France, more than 1 billion articles produced in one night (elections)

To extend media coverage

To cover events not covered before (f.e NTB sports coverage in Norway)

To provide a service to the readers

Real time information about hot topics

To make something that was not done before

Dealing with large amounts of data can be tricky. Automation can help in a fast and accurate way.

To assist journalists

By freeing them of repetitive and time-consuming tasks.

Not to replace human journalists

Nothing in this way is observed (employees) but there is no zero risk! Freelancers are more fragile.

Automation systems can be easily multilingual

What are the obstacles for development?

1

Data

Available, structured, reliable

2

Costs

Between 20 & 50k *Le Monde* // projects supported by Google DNI (UK 708k, Sweden 46k) // public funds (Netherlands, T.U, 700k)

3

Resources

Need internal skills

Who are the humans behind?

1

IT service

Inside the media (f.e. NTB, Norway)

2

Tech start-up (not claiming themselves as media cies)

Mostly in France (Syllabs) and in Germany (Retresco, AX Semantics)

3

Journo-dev

Marginal cases

How do journalists react?

1

Enthusiastic

Free-up time, repetitive and boring tasks for the machine

2

Pessimistic

Fear that automated systems will replace journalists

3

Fatalist

If automation can destroy jobs, it can also create new ones

Do the audiences make the difference?

According to studies, the answer is no but...

- audiences find generated contents less pleasant to read
- automated stories are considered as less well written
- in the same time, those stories are judged more accurate and reliable

Tool for journalists?

Case study (Belgium): Bxl'air bot

 **Bxl'air bot**

[Données du jour](#)
[Pollucarte](#)
[Statistiques](#)
[Rapports mensuels](#)
[Téléchargements](#)
[Normes et dépassements](#)



Bulletin du 14 mars 2018

[Pas d'alerte pollution](#)

7h03. L'indice de la qualité de l'air bruxellois est bon, communique Bruxelles Environnement. La teneur de l'air en particules fines de type PM10 est de 11 µg/m³. Elle se trouve en-dessous de la recommandation de l'Organisation mondiale de la santé (OMS), ainsi que sous la norme européenne. Le taux moyen de particules fines de type PM2.5 est de 10 µg/m³. Il est en-dessous de la norme recommandée par l'OMS. Le taux moyen de black carbon (carbone suie) est de 1,4 µg/m³. La concentration d'ozone dans l'air est de 19 µg/m³. Le taux moyen de dioxyde d'azote est de 50 µg/m³. Ce taux dépasse de 10 µg/m³ la norme européenne et celle recommandée par l'OMS (moyenne annuelle). Le ciel est partiellement nuageux, pour une température de 4,5 degrés.

Ce texte a été généré de manière automatique à partir de données publiques extraites en temps réel.

[Twitter](#) [Partager 25](#)

Tool for journalists?

www.bxlairaibot.be (*Alter Echos*)

- Retrieve and store data about air quality in Brussels
- Real time open data (HTML) collected from different sources
- Provide first data analysis: text, charts, maps
- Monitoring (1 year), raw material for investigative work (causes & consequences)



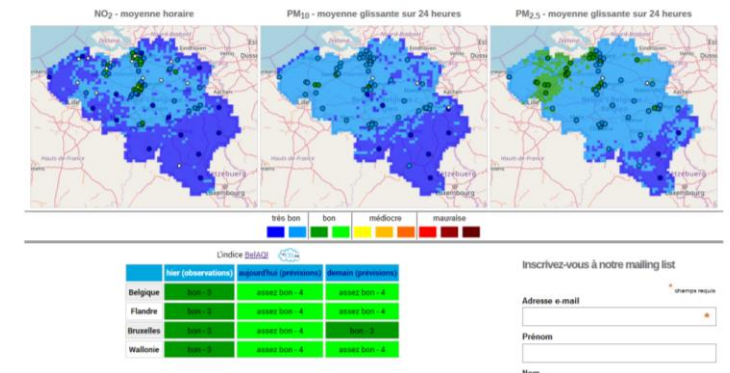
Le site "Qualité de l'air" de Bruxelles Environnement vous permet de suivre en permanence et en temps réel l'évolution des concentrations en polluants dans l'air ambiant en Région de Bruxelles-Capitale.



[derniers 14 jours](#)

Particulate Matter (PM10) Daily mean concentrations (30x30 or 24x30 GMT) last 14 days

code	city	01/03	02/03	03/03	04/03	05/03	06/03	07/03	08/03	09/03	10/03	11/03	12/03	13/03	14/03
41B011	Sint-Agatha-Berchem	45	50	59	18	15	15	15	15	11	10	6	12	NA	
41MEU1	Neder-Over-Heembeek	43	49	52	17	15	19	17	12	16	11	10	8	9	NA
41N043	Voorhaven (Haren)	49	51	NA	NA	NA	15	17	17	17	13	16	15	NA	
41R001	Sint-Jans-Molenbeek	49	47	56	17	NA	15	19	17	11	11	8	10	NA	
41R012	Ukkel	51	42	51	14	NA	13	10	13	8	7	5	6	NA	
41W011	Sint-Lambrechts-Woluwe	55	45	58	17	13	19	16	12	12	10	8	5	9	NA
40AL01	Antwerpen-Linkeoever	74	70	81	19	16	15	20	14	14	14	10	12	NA	
40AL02	Doel (Engelvestenweg)	91	75	88	20	16	16	16	16	18	18	12	12	NA	
40AL03	Verebroek	87	72	84	20	16	16	16	16	16	17	10	13	NA	
40AL04	Kallo (Liefkenshoektunnel)	83	72	79	19	16	16	16	16	16	16	15	13	NA	
40AL05	Kallo (plus Kallo)	91	72	83	18	16	16	16	16	16	16	15	13	NA	
40GK06	Diepenbeek (Zinniastraat)	52	56	65	16	16	16	16	16	16	16	16	16	NA	
40HB23	Hoboken	43	56	70	17	17	20	15	19	12	13	11	11	NA	
40MN01	Menen	45	53	64	16	20	16	16	16	16	16	12	16	NA	
40OB01	Oostrozebeke	53	62	74	16	16	16	16	16	16	16	16	16	NA	
40OB02	Wielzebeke	52	62	77	16	16	16	16	16	16	16	16	16	NA	
40RL01	Roeselare (Brugsesteenweg)	53	61	73	16	16	16	16	16	16	16	16	16	NA	
40SA04	Hoeveren	49	61	82	19	16	16	16	16	16	16	16	16	NA	
40SZ01	Steenokkerzeel	47	61	76	16	16	16	16	16	16	16	16	16	NA	
40SZ02	Steenokkerzeel	41	52	63	16	14	19	16	12	15	11	11	7	10	NA
42MR02	Antwerpen (Luchtbal)	56	64	80	16	16	16	16	16	16	16	16	16	NA	

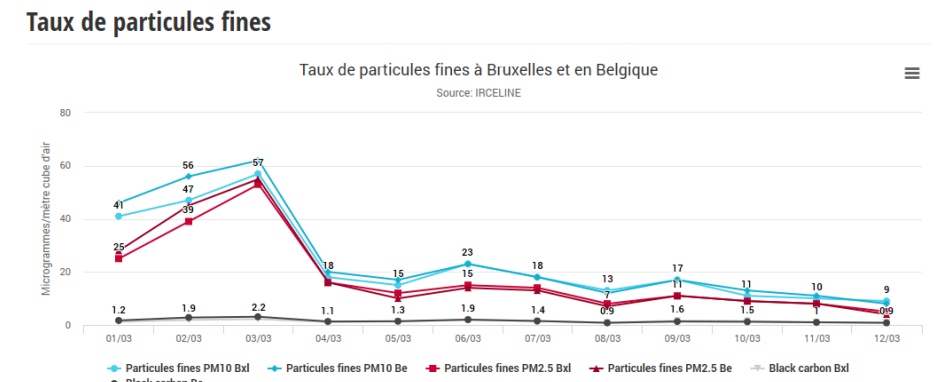
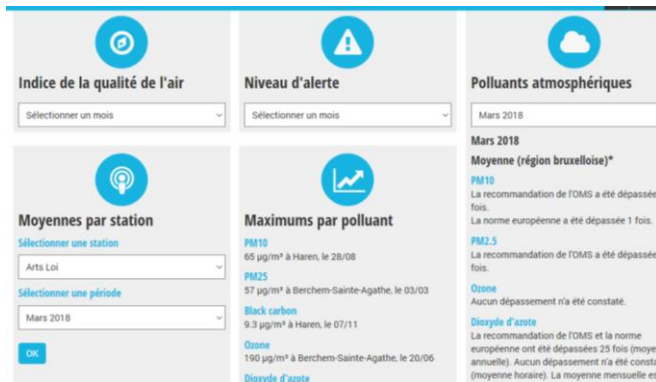


Automated journalism: an overview
Laurence Dierickx

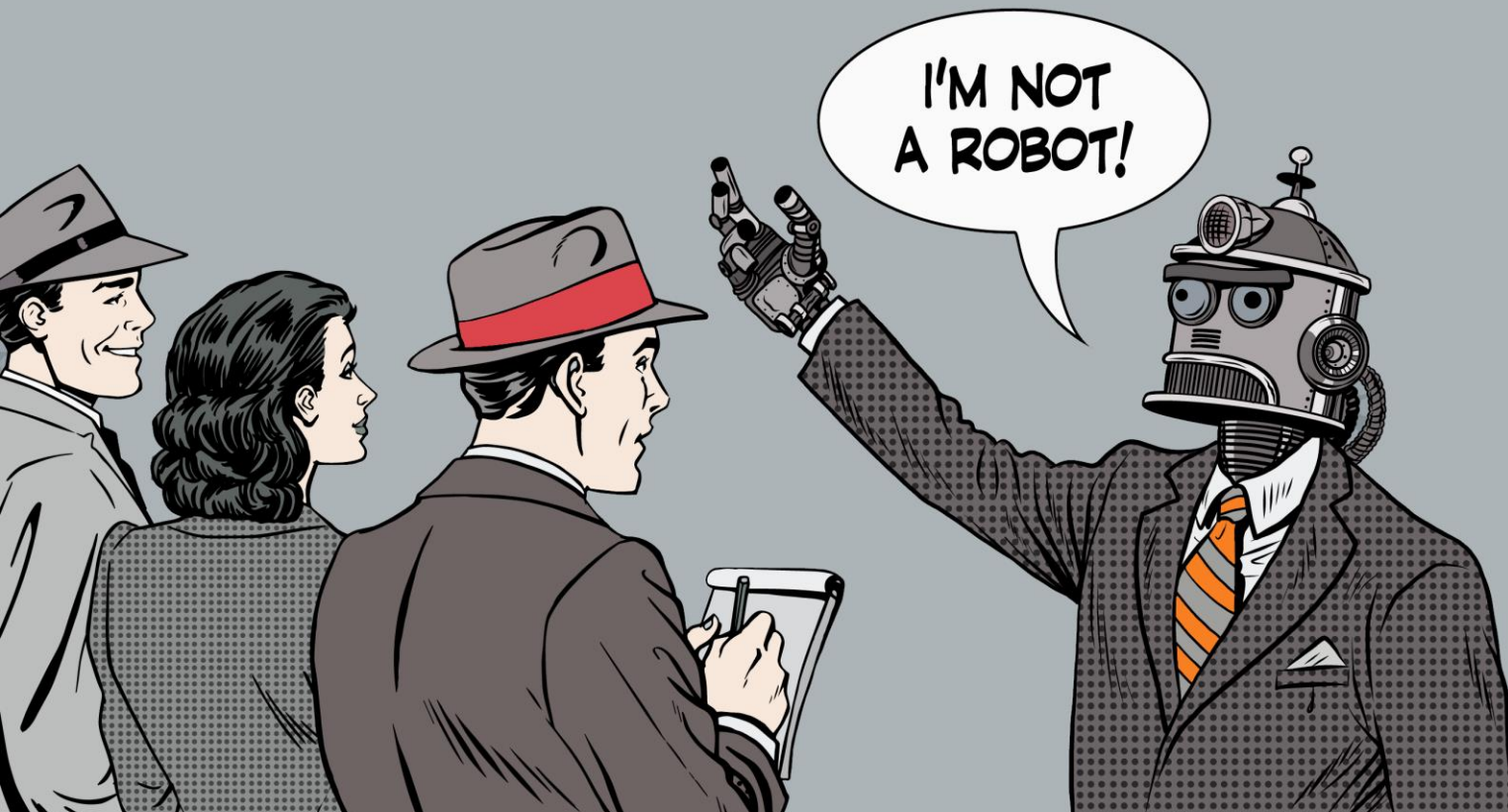
Tool for journalists?

Assets for the journalists

- Work by its own to make a long time job (gain of time)
- Accuracy (supposed a human control), objectivate by numbers (exceeding of EU norms)
- No need to get particular skills (at the opposite of datajournalism)
- Used by journalists only if they can make sense of it



Thank you!



@ohmyshambles
www.ohmybox.info