

# Is China trying to control the narrative about Xinjiang on Google News?

A study of the Belgian media landscape as aggregated by Google News

CROSSOVER



This project has received funding from the European Union's programme on the financing of Pilot Projects and Preparatory Actions in the field of "Communications Networks, Content and Technology" under grant agreement LC-01682253

## Version tracker

Reviewer	Date	Notes
Check First	Feb 14, 2023	Initial version
Check First	Feb 15, 2023	Addition of the paragraph on Google's answer (in the conclusions)

<b>1. Introduction</b>	<b>2</b>
<b>2. Key findings</b>	<b>3</b>
2.1 Xinjiang: articles written by Chinese state-controlled sources outnumber top western media outlets	3
2.2 Other observed keywords	3
<b>3. A Chinese strategy? Diluting challenging journalism about Xinjiang in a sea of nondescript articles</b>	<b>5</b>
3.1. Chinese government-controlled outlets racing to the top	6
3.1.1. China Internet Information Center in French	6
3.1.2. People daily in French	6
3.1.3. New China News Agency (Xinhua)	6
3.1.4. China Radio International	7
3.1.5. Chine Direct	7
3.1.6. Ministry of Foreign Affairs of PRC	7
3.2. A very talkative set of state-controlled outlets	8
3.2.1. A high number of articles about Xinjiang originate from China	8
3.2.2. Chinese state-controlled outlets on a publishing frenzy	8
3.2.3. Articles from Chinese state-controlled outlets are widely displayed on Google News	10
3.3 What are Chinese and Western outlets reporting about Xinjiang?	15
3.3.1. Article body word clouds	17
3.3.2. Article title word clouds	19
<b>4. What amount of foreign vs. domestic content are Belgians exposed to?</b>	<b>21</b>
4.1. Analysing publishers top level domains	21
4.2. Analysing the articles' origin countries using top level domains	22
4.3. Comparing both analyses	22
<b>5. How about other topics than Xinjiang on Google News?</b>	<b>24</b>
5.1. Which publishers are linked to on Google News?	24
5.2. French and Dutch sources are very present in the overall results	25
5.3. Belgian media outlets are still widely favoured in terms of total articles displayed	26
5.4. A special mention: the dominance of VRT and HLN content served to Dutch-speaking Belgian users	28
<b>6. Conclusions</b>	<b>29</b>
<b>7. Methodology</b>	<b>30</b>
7.1. Exportation of all the data for the following keywords and languages:	30
7.2. Analysis of the results for each keyword calculating the following:	30
7.3. Enrichment of the dataset with metadata	30
7.4. Creation of data visualisation for the values	31
7.5 Availability of datasets used for the investigation	31

# 1. Introduction

Google News is a popular tool presenting articles from publishing outlets (called “sources” by Google). Our research shows that most outlets promoted by Google News are actual media organisations but we also found relatively dubious sources alongside legitimate ones. This study shows which outlets were promoted when querying a selection of keywords monitored by CrossOver, and which ones were pushed before the eyes of Belgian users?

In an effort to measure foreign influence in the Belgian media space as presented by Google News, CrossOver developed a classification approach based on the publishers top level domains to approximate their country of origin. This process showed that Belgian sources are a minority among selected publishers, meaning that Google News presents articles from an array of international publishers where Belgian outlets only represent a small fraction of selected websites. However this is generally counter-balanced by exposing Belgian articles more frequently than foreign sources in the results.

CrossOver found that the Google News service, as accessed from Belgium, presented the users with accurate content from legitimate sources most of the time on the searched topics<sup>1</sup>. On specific search terms however, an unusual amount of content originating from state-controlled Chinese outlets was found, pushing official Chinese narratives. **In the case of the query “Xinjiang”, the two most active publishers on the topic are Chinese state-controlled outlets.**

Another noteworthy item is the fact that, depending on the searched topic, Belgian media outlets are less represented in the Google News results than French or Dutch media organisations in terms of total article occurrences although the service is accessed from Belgium. However, Google News’ ranking system seems to favour Belgian media organisations as their articles are presented more often and for longer periods than those from neighbouring countries.

---

<sup>1</sup> See [methodology](#)

## 2. Key findings

### 2.1 Xinjiang: articles written by Chinese state-controlled sources outnumber top western media outlets

Our research established that when searching for the keyword “Xinjiang”, the autonomous Uyghur province in People's Republic of China (PRC), French speaking Belgian users of Google News were presented with content produced by Chinese state-controlled actors. These stories depict the province mainly in a positive fashion<sup>2</sup> and with little to no mention of Uyghurs and no mention at all of human rights issues. **13% of all displays of articles about Xinjiang were written by Chinese state-controlled outlets.**



CrossOver found that six Chinese outlets produced an overwhelming quantity of articles returned in Google News' results for the search “Xinjiang”, seemingly in an attempt to dwarf the other publisher's productions.

**Among 209 publishers covering Xinjiang on Google News, six state-controlled Chinese outlets produced 27% of the total number of articles written on the subject.** The most prolific Chinese outlet about Xinjiang is the French language version of “China Internet Information Center”, an outlet directly run by the Chinese government.

This could be interpreted as a deliberate strategy from these Chinese state-controlled outlets to flood the offer of articles about Xinjiang with content, produced in French, avoiding the topics of human rights of the Uyghur community altogether. Instead, they offer articles about tourism, agriculture or infrastructure, **effectively diluting western coverage of the region.**

### 2.2 Other observed keywords

For the most part, Google News presents users with content originating from the mainstream media from Belgium, France and the Netherlands (depending on the considered language). According to our dataset, the service generally does not seem to promote dubious publications on other search terms than “Xinjiang”<sup>3</sup>. Some debatable sources can however be found in the results such as political content presented as “news”, emanating from PTB, a left-wing Belgian

<sup>2</sup> See sentiment mapping in section [3.2.2](#)

<sup>3</sup> See [methodology](#) for a list of queried search terms.

political party; another example is the French far-right *Boulevard Voltaire*<sup>45</sup> website. However present on Google News, such outlets appear marginally on the service and are not displayed often in search results.

Although Google News was accessed from Belgium during our data collection, content produced by French and Dutch outlets were a majority on most topics when only article quantity is considered. This can be a result of the fact that the amount of French and Dutch media outlets is simply greater than the quantity of Belgian outlets. This being said, Google News' ranking of articles seemed to favour Belgian outlets. Even if fewer articles are produced by Belgian sources, they are exposed for longer and more frequently than foreign ones.

Another observation was the fact that in Dutch, two media outlets are taking the lion's share of Google News users attention. VRT and HLN are responsible for nearly half of all content displayed on the platform.

---

<sup>4</sup> Boulevard Voltaire is an ultraconservative opinion website founded in 2012 by Robert Ménard and Dominique Jamet which describes itself as "open to all sensibilities of the conservative right"

<sup>5</sup>[https://www.lemonde.fr/economie/article/2022/07/08/comment-l-extreme-droite-a-infiltrer-les-medias\\_6133900\\_3234.html](https://www.lemonde.fr/economie/article/2022/07/08/comment-l-extreme-droite-a-infiltrer-les-medias_6133900_3234.html)

### 3. A Chinese strategy? Diluting challenging journalism about Xinjiang in a sea of nondescript articles

As CrossOver has demonstrated in its previous investigation “[Are YouTube Algorithms Addicted to State-Controlled Media?](#)”, China’s official world view is widely distributed by its official state-run media outlets and sometimes amplified by platforms such as Youtube. Google News is no exception when considering Xinjiang. On the topic of the autonomous Uyghur province, a particular effort was put into the production of many French language articles on websites hosted in China. Six Chinese state-controlled websites are responsible for 27% of all articles retrieved from Google News when searching for the keyword “Xinjiang” between 1 February 2022 and 1 February 2023.

In terms of articles displays, eg. how frequently a specific article was shown during the study period, Chinese state-controlled outlets made up for 13% of the total when considering the top 50 sources. This means that although Chinese state-controlled sources accounted for only 3% of the number of sources considered (6 out of 50), they represent 13% of the article displays.

The human rights situation in Xinjiang, China, has been widely criticised by the international community. Reports<sup>6</sup> indicate that the Chinese government has been engaging in widespread human rights violations, including mass detainment of Uyghurs and other minority groups in internment camps, forced labour, and cultural assimilation



policies. There have also been allegations of widespread surveillance<sup>7</sup> and repression of religious and cultural expression. The Chinese government has defended its policies in Xinjiang as necessary to combat terrorism and extremism<sup>8</sup>, but the evidence suggests that these policies are primarily aimed at suppressing the ethnic and cultural identity of the region's minority populations.

<sup>6</sup> <https://news.un.org/en/story/2022/08/1125932>

<sup>7</sup> <https://www.hrw.org/news/2021/11/24/mass-surveillance-fuels-oppression-uyghurs-and-palestinians>

<sup>8</sup> In official statements, the Chinese Government has emphasised that “Xinjiang-related issues are in essence about countering violent terrorism and separatism” and that it is doing so “in accordance with law”. E.g., Remarks “A People-centred Approach for Global Human Rights Progress”, by H.E. Wang Yi, State Councillor and Foreign Minister of the People's Republic of China, at the High-level Segment of the 46th Session of UN Human Rights Council, 22 February 2021.

## 3.1. Chinese government-controlled outlets racing to the top

Out of a total of 1050 unique articles shown by Google News as results to the query “Xinjiang”, 288 were published by the following outlets, all state owned or controlled:

### 3.1.1. China Internet Information Center in French



**URL:** french.china.org.cn

**HTTPS certificate:** No

**Description:** This website describes itself as a central point to learn about China for foreigners, aggregating content from multiple news agencies across the country. The website’s impressum states “*The authorised government portal site to China, China.org.cn is published under the auspices of the*

*State Council Information Office and the China International Publishing Group (CIPG) in Beijing.*”

### 3.1.2. People daily in French



**URL:** french.peopledaily.com.cn

**HTTPS certificate:** No

**Description:** People daily in French is the French language version of China’s largest newspaper, portraying itself as the “*organ of the central committee of the Chinese communist party*”.

### 3.1.3. New China News Agency (Xinhua)



**URL:** french.xinhuanet.com

**HTTPS certificate:** Yes

**Description:** French language version of the website of the official Chinese news agency Xinhua, subordinate to the State Council of China and under the control of the Chinese Communist Party.

### 3.1.4. China Radio International



**URL:** [French.cri.cn](http://French.cri.cn)  
**HTTPS certificate:** Yes

**Description:** French language website of China Radio International, a state-owned media outlet controlled by the Chinese Communist Party through the Chinese Media Group.

### 3.1.5. Chine Direct



**URL:** [chinedirect.net](http://chinedirect.net)  
**HTTPS certificate:** Yes

**Description:** French language website presenting itself as a resources to understand China's culture, history or political and societal affairs. Ran by the Chinese presse agency Xinhua.

### 3.1.6. Ministry of Foreign Affairs of PRC



**URL:** [www.fmprc.gov.cn](http://www.fmprc.gov.cn)  
**HTTPS certificate:** Yes

**Description:** Official site of the Ministry of Foreign Affairs of the People's Republic of China

Note: https certificates have become a common practice in modern web browsing. Navigating to a website which does not use this safe protocol will present the user with a warning message discouraging the user to continue using this site.

## 3.2. A very talkative set of state-controlled outlets

### 3.2.1. A high number of articles about Xinjiang originates from China

The following table shows Google News' top 15 publishers in terms of produced and listed articles to be found in the results when searching the keyword "Xinjiang" during our study.

#	Publication url	Total unique articles	% total articles
1	french.china.org.cn	105	10,00 %
2	french.peopledaily.com.cn	93	8,86 %
3	www.zonebourse.com	57	5,43 %
4	french.xinhuanet.com	46	4,38 %
5	www.lefigaro.fr	28	2,67 %
6	www.rtf.be	24	2,29 %
7	www.liberation.fr	23	2,19 %
8	www.letemps.ch	21	2,00 %
9	www.sportytrader.com	21	2,00 %
10	www.lemonde.fr	19	1,81 %
11	www.la-croix.com	19	1,81 %
12	french.cri.cn	18	1,71 %
13	www.rfi.fr	17	1,62 %
14	www.chine-magazine.com	16	1,52 %
15	www.courrierinternational.com	15	1,43 %
...			
	<b>Total</b>	<b>1050</b>	<b>100%</b>

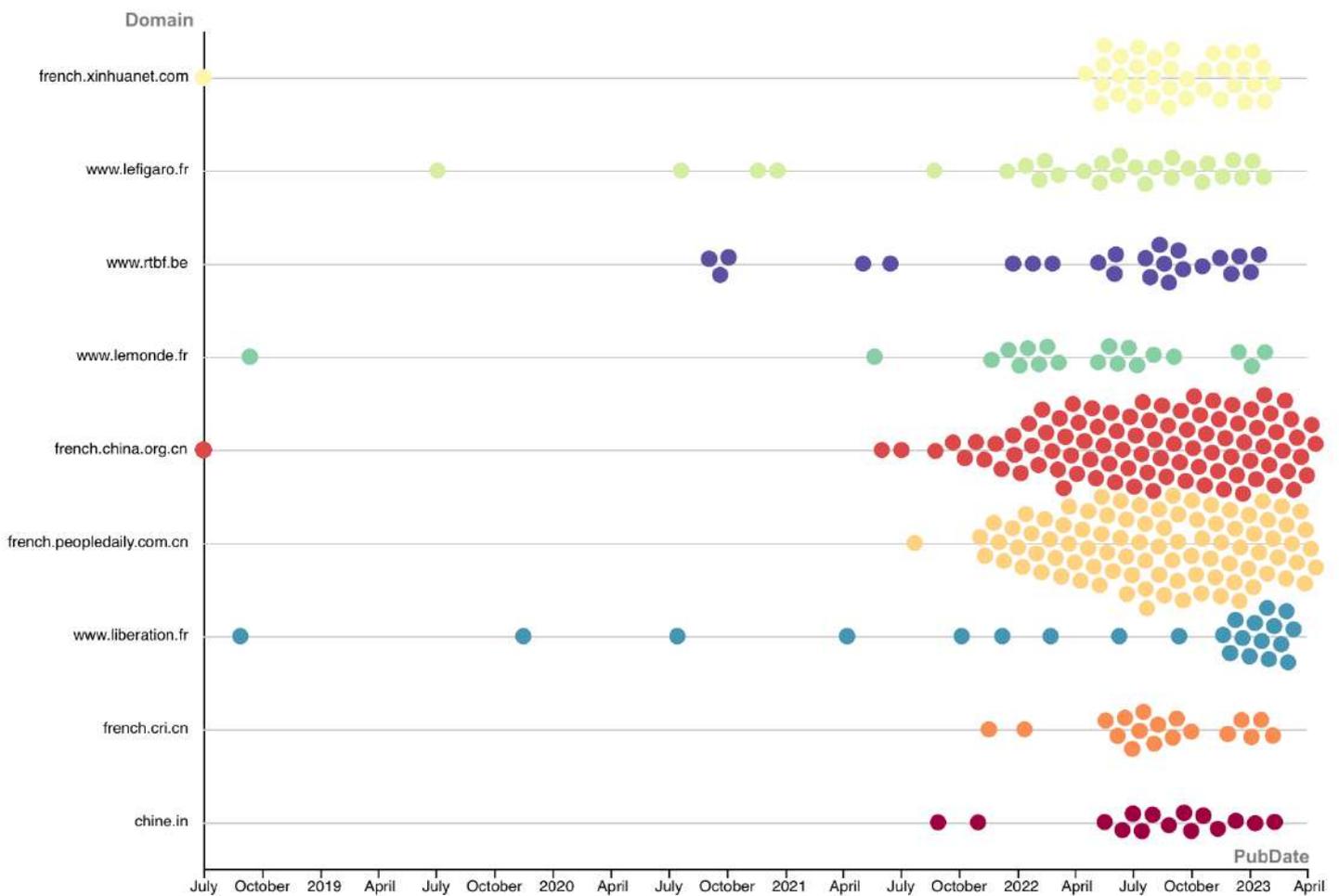
*China Internet Information Center* is by far the most prolific outlet on the topic of Xinjiang, followed by another Chinese state-controlled outlet, the French language version of *People's Daily*, PRC's main newspaper. As a point of comparison, the first non-Chinese *mainstream media* publication appearing in our dataset is the French newspaper *Le Figaro*, with about a quarter of the articles listed for *China Internet Information Center*. The first Belgian outlet is the public service radio and television service RTBF, showing less than a fifth of the number of articles listed for *China Internet Information Center*.

### 3.2.2. Chinese state-controlled outlets on a publishing frenzy

Another way to visualise the effort of these Chinese state-controlled media is to represent their publication frequency. The following chart shows the number of articles about Xinjiang listed on Google News as a function of their publication date. The highest ranked Chinese

publications in terms of number of articles listed are compared to the following most represented western media outlets.

Not only can we observe that Google News can link to old articles from the summer of 2018, but also that the publication rate of Chinese state-controlled outlets increased in late 2022 as the topic of Xinjiang was more covered by western outlets like *Le Monde*, *Libération* or *RTBF*. Incidentally, this period corresponds to the publication of a report from Amnesty International on the repression of Uyghurs by Chinese authorities<sup>9</sup>.



<sup>9</sup> <https://crossover.social/online-chinese-media-offensive-also-reaches-belgium/>

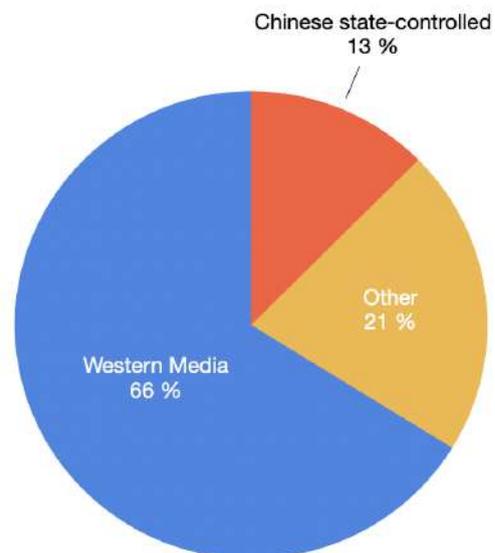
As regards the completeness of the contents relayed by Google News, we could not determine if the service chose to exhaustively publish all articles from any given publisher on the topic.

### 3.2.3. Articles from Chinese state-controlled outlets are widely displayed on Google News

Any given article can show up multiple times across search results on Google News. For this study, we queried the platform twice a day and kept count of which articles were shown to Belgian users in French. During the year-long study, CrossOver counted 48700 article displays from 209 sources. For concision, we considered the top 50 publishers<sup>10</sup> in terms of article displays and categorised each source in the following fashion:

Chinese state-controlled outlets	A clear link to the Chinese government is stated on the website
Western media outlets	Well established journalistic outlets from France, Belgium, Switzerland, Canada
Other	Non-journalistic outlets, NGOs and other outlets

Again, Chinese state-controlled outlets are very present in those results, accounting for 13% of all articles displayed although they only represent 3% of all listed sources (6 out of 50 outlets).



<sup>10</sup> The top 50 sources accounted for 91% of all article displays (43,658 article displays out of a total of 47,800)

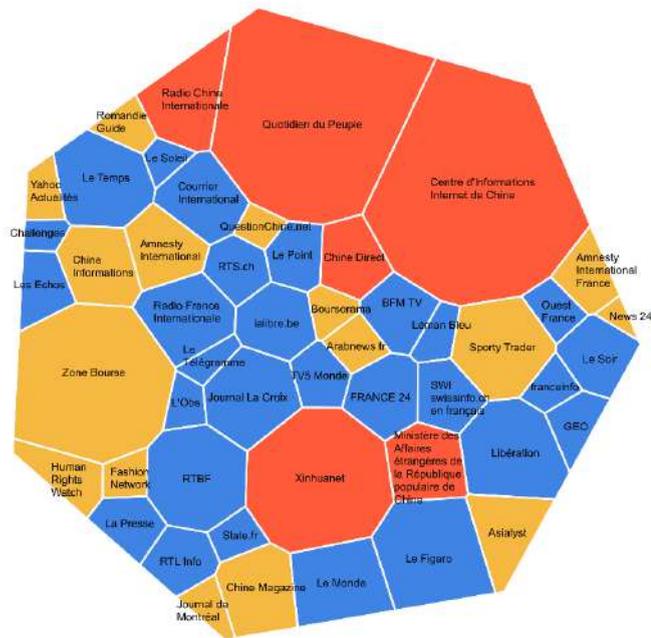
The following table shows the top 50 sources in terms of articles displayed during the study. It lists the sources' domain, name, number of article displays, their category and proportion of article displays as a percentage of the total number of articles displayed for the top 50 sources.

#	Domain	Source	Qualification	Displays	% total
1	www.la-croix.com	Journal La Croix	Western Media	3695	8,46 %
2	www.lemonde.fr	Le Monde	Western Media	2957	6,77 %
3	french.china.org.cn	Centre d'Informations Internet de Chine	Chinese state-controlled	2757	6,31 %
4	www.amnesty.org	Amnesty International	Other	2240	5,13 %
5	www.courrierinternational.com	Courrier International	Western Media	2215	5,07 %
6	www.liberation.fr	Libération	Western Media	2191	5,02 %
7	www.letemps.ch	Le Temps	Western Media	2131	4,88 %
8	www.rtb.be	RTBF	Western Media	1885	4,32 %
9	french.peopledaily.com.cn	Quotidien du Peuple	Chinese state-controlled	1865	4,27 %
10	www.geo.fr	GEO	Western Media	1637	3,75 %
11	www.lefigaro.fr	Le Figaro	Western Media	1400	3,21 %
12	asialyst.com	Asialyst	Other	1359	3,11 %
13	www.amnesty.fr	Amnesty International France	Other	1151	2,64 %
14	www.rfi.fr	Radio France Internationale	Western Media	984	2,25 %
15	www.bfmtv.com	BFM TV	Western Media	960	2,20 %
16	www.rts.ch	RTS.ch	Western Media	953	2,18 %
17	fr.fashionnetwork.com	FashionNetwork.com FR	Other	935	2,14 %
18	www.hrw.org	Humans Rights Watch	Other	849	1,94 %
19	www.lesechos.fr	Les Échos	Western Media	846	1,94 %
20	www.lapresse.ca	La Presse	Western Media	843	1,93 %
21	www.france24.com	FRANCE 24	Western Media	664	1,52 %
22	www.lepoint.fr	Le Point	Western Media	658	1,51 %
23	french.xinhuanet.com	Xinhuanet	Chinese state-controlled	632	1,45 %
24	novastan.org	Novastan France	Western Media	508	1,16 %
25	www.zonebourse.com	Zonebourse.com	Other	472	1,08 %
26	www.ouest-france.fr	Ouest France	Western Media	468	1,07 %
27	www.lalibre.be	lalibre.be	Western Media	462	1,06 %
28	www.bbc.com	BBC	Western Media	412	0,94 %
29	www.chine-magazine.com	Chine Magazine	Other	396	0,91 %
30	www.lesoir.be	Le Soir	Western Media	393	0,90 %
31	www.questionchine.net	QuestionChine.net	Other	374	0,86 %
32	www.commodafrica.com	Commodafrica	Other	344	0,79 %
33	www.forbes.fr	Forbes France	Western Media	327	0,75 %

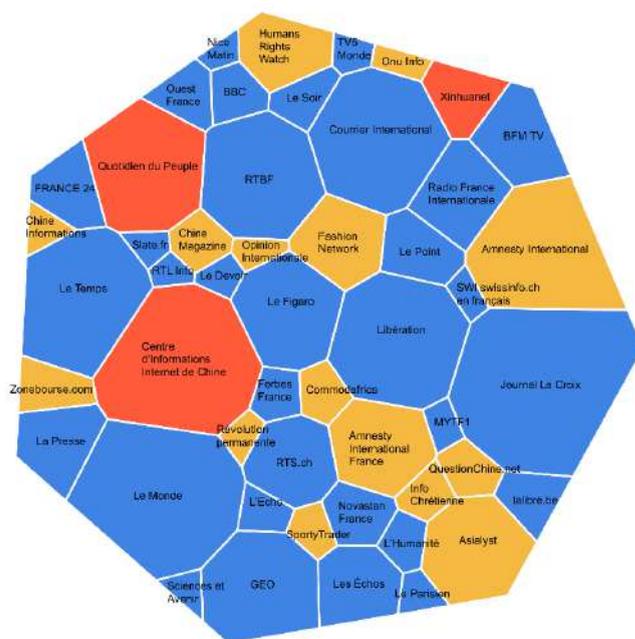
#	Domain	Source	Qualification	Displays	% total
34	www.infochretienne.com	Info Chrétienne	Other	304	0,70 %
35	www.humanite.fr	L'Humanité	Western Media	298	0,68 %
36	www.lecho.be	L'Echo	Western Media	272	0,62 %
37	www.sportytrader.com	SportyTrader	Other	256	0,59 %
38	information.tv5monde.com	TV5 Monde	Western Media	255	0,58 %
39	news.un.org	Onu Info	Other	223	0,51 %
40	chine.in	Chine Informations	Chinese state-controlled	213	0,49 %
41	www.ledevoir.com	Le Devoir	Western Media	209	0,48 %
42	www.leparisien.fr	Le Parisien	Western Media	208	0,48 %
43	opinion-internationale.com	Opinion Internationale	Other	198	0,45 %
44	www.tf1.fr	MYTF1	Western Media	195	0,45 %
45	www.revolutionpermanente.fr	Révolution permanente	Other	189	0,43 %
46	www.swissinfo.ch	SWI swissinfo.ch en français	Western Media	183	0,42 %
47	www.slate.fr	Slate.fr	Western Media	182	0,42 %
48	www.sciencesetavenir.fr	Sciences et Avenir	Western Media	178	0,41 %
49	www.nicematin.com	Nice Matin	Western Media	167	0,38 %
50	www.rtl.be	RTL Info	Western Media	165	0,38 %

The following Voronoi representations show the share of publishers by article count as an area (top) and their respective amount of displayed articles (bottom). We can observe that even if the Chinese state-controlled outlets take up a large area on the left Voronoi, their exposure to Belgian users is reduced by the effects of Google News' algorithm.

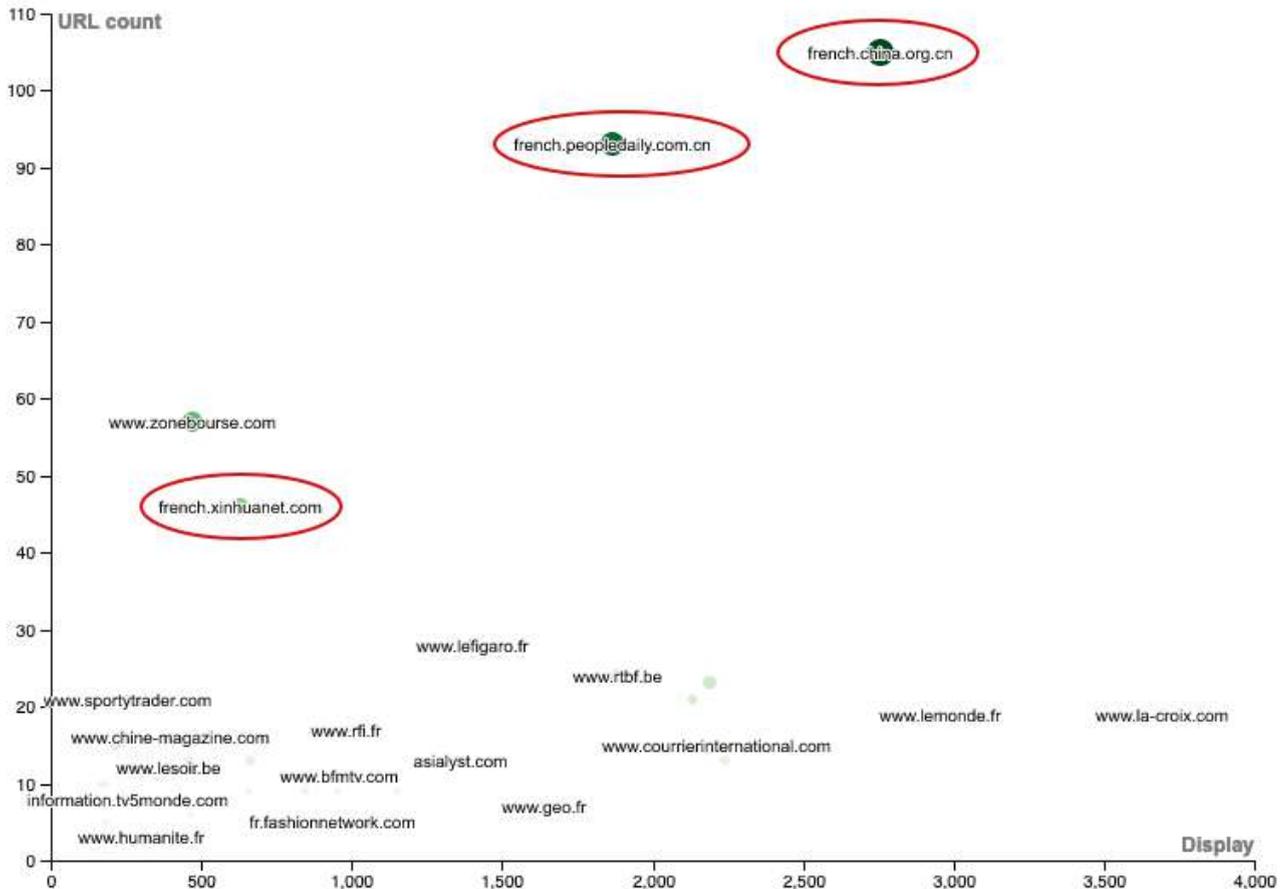
Top 50 sources by uniques urls



Top 50 sources by display



As a way to visualise both previous sections of this paper, the following graph shows the relationship between the number of unique articles returned by Google News on the topic over the number of times throughout the year these articles were displayed in the search results.



In an attempt to understand if the number of produced articles influences the exposure of a source to Google News users, we produced the above graph. This representation shows how often we observed content from specific sources as compared to the number of articles from these sources that were aggregated after performing a search for “Xinjiang” on Google News. The furthest to the right an outlet is located on the graph, the most displayed it was. For example, even if a mere 19 articles from the French outlet *La Croix* were linked to, the publication was the most present in the search results, with 3695 impressions during the considered period. *China Internet Information Center* however, got 2757 impressions “only” even if 105 articles were linked to from Google News.

This suggests that Google News’ ranking does not only consider the sheer amount or publication date of articles produced on a topic by a given outlet, but sorts them using other factors. Alas, these possible other factors are poorly documented. The intricacies of Google

News' ranking are unknowable. On its support page<sup>11</sup>, Google describes the selection process of the articles as follows:

*“Computer algorithms determine what shows up in Google News. The algorithms determine which stories, images, and videos show, and in what order. In some cases, people like publishers and Google News teams choose stories.”*

The same support page also mentions that personalised results can be presented to a user logged into Google's services. As CrossOver's methodology does not rely on user profiles or preference settings in Google News, no personalisation aspect can be inferred from our results. Our data was collected each time as a fresh, not logged into Google services user.

Additionally, in a blogpost on its developer forum<sup>12</sup>, Google explains how the company assesses the transparency of the sources Google News is aggregating, both at the article level and at the site level. The blogpost states that:

*“An important way news sources build trust with their audience is by providing information about their articles and site. This includes information like clear dates and bylines, as well as information about authors, the news source, company or network behind it, and contact information. This type of transparency helps ensure that readers can easily learn information about both the content they are reading, viewing, or listening to and the creators of that content.”*

The company also suggests that its policy is not enforced in the same manner, depending on the country of origin of the articles:

*“We consider different regional and country-level expectations and practices around transparency.”*

### 3.3 What are Chinese and Western outlets reporting about Xinjiang?

In short, French and Belgian publishers mostly cover human rights issues and Uyghurs repression by the Chinese government, when Chinese controlled outlets entertain their audience about tourism and business issues. The topics addressed in the pieces were analysed by Crossover in an automated fashion. The NLP entities extraction tool Diffbot was used to extract the topics addressed in the full body of the articles. These terms were gathered on word clouds. Additionally, as users are exposed to a list of articles' titles when browsing news, we also analysed the words used in the titles. The reason for this being that most users would read the titles but not necessarily all articles.

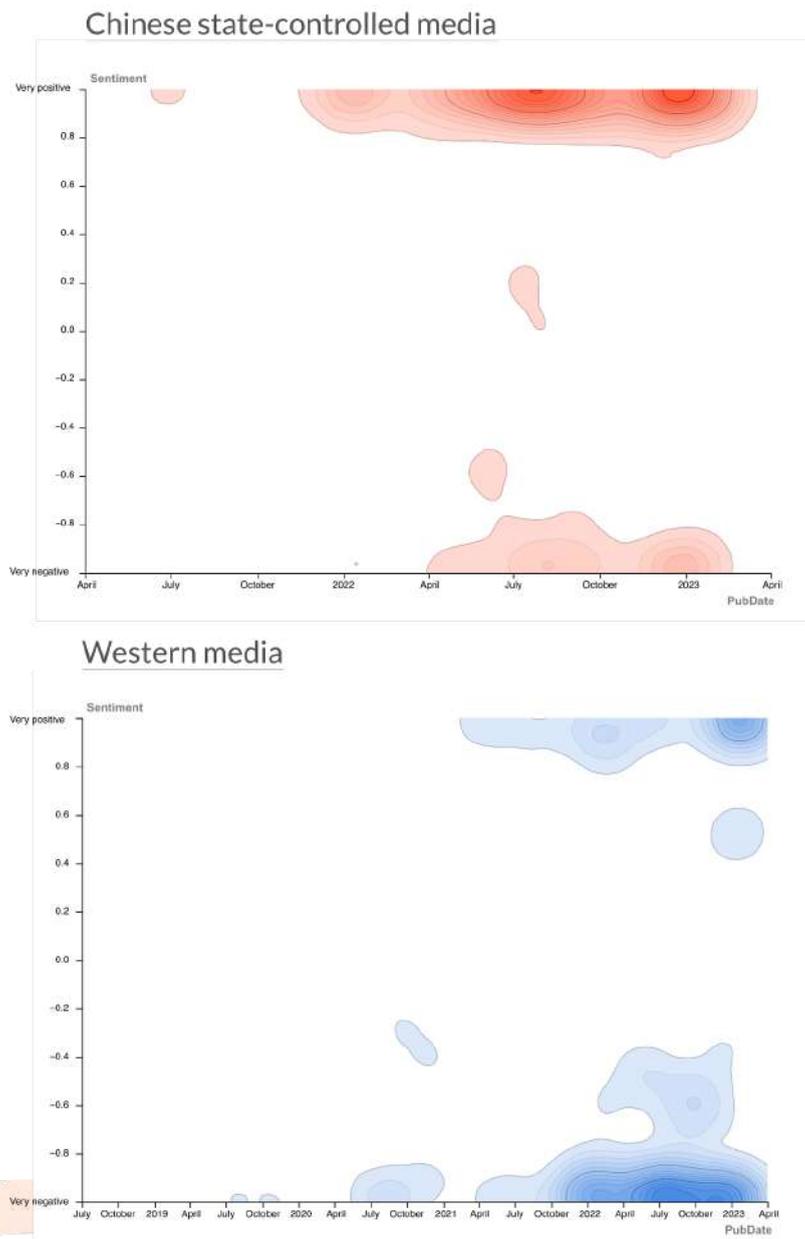
---

<sup>11</sup> <https://support.google.com/googlenews/answer/9005749?hl=en>

<sup>12</sup> <https://developers.google.com/search/blog/2021/06/google-news-sources?hl=en>

To illustrate this observation, we ran a NLP entities extraction algorithm<sup>13</sup> on the articles produced by the four Chinese state-controlled outlets<sup>14</sup> we've identified. We compared it to a similar mapping for the four most displayed Western media<sup>15</sup> outlets on the list. The following heatmap shows a representation of the sentiment of the Chinese (top) and Western (bottom) articles. The sentiment is plotted on the y-axis, 0 being very negative and 1 very positive.

The top heatmap shows that the sentiment expressed in Chinese state-controlled outlets in their articles about Xinjiang is mostly positive (darker areas along the top), as opposed to mostly negative for the articles written by Western media (as shown on the bottom heatmap).



<sup>13</sup> <https://docs.diffbot.com/reference/introduction-to-natural-language-api>

<sup>14</sup> Centre d'Informations Internet de Chine, Quotidien du Peuple, Xinhuanet, Radio Chine Internationale

<sup>15</sup> Le Figaro, RTBF, Libération, Le Monde









## 4. What amount of foreign vs. domestic content are Belgians exposed to?

As we've established, the quantity of available articles from a publisher on a given topic does not solely determine which articles are presented to a Google News user when performing a search. CrossOver also measured the number of times a specific article appeared in the search results, e.g. their *exposition*. A useful indicator to determine the origin of a publisher is the website's top level domain (TLD). Most of the time, media outlets will use national TLDs (e.g. .be for Belgium, .cn for China) also called Country Code Top Level Domain or ccTLD. This indicator is not perfect as some publishers choose an other TLD than their domestic ccTLD (e.g. the Belgian outlet *lavenir.net* or the Chinese outlet *xinhuanet.com*) but these are exceptions and CrossOver determined that this indicator gave a solid representation of the "nationality" of a piece of content or publisher.

### What is a TLD (Top Level Domain)?

A TLD, or Top-Level Domain, is the highest level of a domain name in the Internet's domain name system (DNS). It is the last part of a domain name and is usually separated from the rest of the domain name by a dot (.). Examples of TLDs include .com, .org, .net, .edu, and many others. TLDs are used to categorise and organise domain names, and they can provide information about the type or purpose of a website.

A ccTLD (country code Top-Level Domain) is a type of TLD that represents a specific country or territory on the Internet. It consists of two letters that correspond to the ISO 3166-1 alpha-2 code of a country. Some examples of ccTLDs are .fr (France), .de (Germany), .cn (China), .jp (Japan), etc. ccTLDs are used to identify websites that are associated with a particular country or territory and are typically managed by the country's government or a designated entity.

### 4.1. Analysing publishers top level domains

In an effort to determine the proportions of national and foreign content Belgians are exposed to on Google News, CrossOver used a TLD analysis model on two indicators: publishers and articles. For all considered keywords, we listed all publishers contributing content matching a given keyword. For example, when performing the search "Xinjiang" twice a day during the year-long study, we collected content from 209 different publishers as returned by Google News. These publishers were bucketed according to their ccTLD, producing 17 different categories, each corresponding to a ccTLD, hence giving a solid indication of their country of operation.

## 4.2. Analysing the articles' origin countries using top level domains

Using an analogue classification to the one explained above, CrossOver established a list of the probable origin countries of all articles presented to users during the study based on the ccTLD present in the linked url. This allowed us to determine the probable origin country of the most viewed content on Google News. As the same article can be displayed multiple times over the course of any given search period, certain articles would appear more frequently in our data collection.

## 4.3. Comparing both analyses

Google News' algorithms favour certain articles and/or publishers, listing them more often than others in search results.

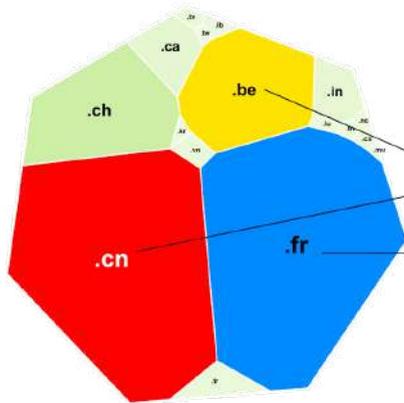
As an example, the following diagram representation shows which articles from which ccTLDs benefit from the most exposure on Google News after a search for the keyword "Xinjiang" in the French language. The left Voronoi graph shows the proportion of articles from given ccTLDs as an area of the total. It clearly shows that Google News aggregated mostly French and Chinese content to provide coverage on the topic of "Xinjiang" to Belgian users as out of the 668 articles present in the results, 229 were bearing a .fr ccTLD and 230 a chinese .cn ccTLD.

The right graph shows the ccTLD proportion for all articles presented to the user for the search "Xinjiang". The French outlets benefited from a greater exposure to Belgian users. However the Chinese articles were less presented to users when compared to the amount of available articles. This echoes our findings as presented in [section 3.2.2](#) as we demonstrated that the sheer amount of articles produced by an outlet isn't enough to guarantee that an outlet will be presented in the same proportions to users.

In addition, we can observe that Belgian outlets only constitute a minority of the news offering on the service, even though Google News was accessed from Belgium.

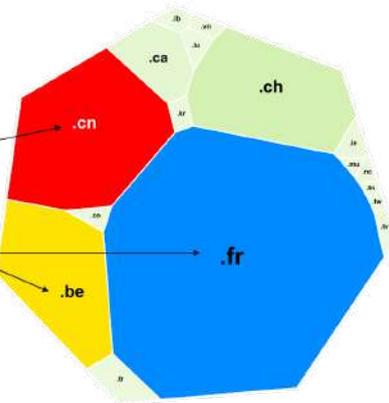
## Number of articles per ccTLD

Keyword : Xinjiang  
Language : FR  
Country : Belgium



## Article displays per ccTLD

Keyword : Xinjiang  
Language : FR  
Country : Belgium



Another noteworthy datapoint is the relationship between the left and right diagram for specific ccTLD. The yellow area occupied by the Belgian ccTLD on the publisher's graph is roughly the same on the articles diagram, meaning that the proportion of Belgian articles shown to the public is somewhat proportional to the importance of Belgian outlets among publishers. However, looking at the red .cn ccTLD, we can note that this is not true for Chinese publishers. Even if they represent a large portion of the published articles, their articles are presented to the public in a smaller proportion.

## 5. How about other topics than Xinjiang on Google News?

Our investigation was also based on another set of monitored keywords: “Covid”, “Vaccine”, “Russia” and “Ukraine” in both French and Dutch. We could not detect any widespread attempts to promote counter narratives about these terms. Nearly all the results originated from legitimate media outlets based in Belgium, France and the Netherlands for the most part<sup>17</sup>.

### 5.1. Which publishers are linked to on Google News?

When looking up “Covid”, “Vaccine”, “Russia” and “Ukraine”, we determined that Google News mostly presents the users with content originating from trusted media brands. The list of outlets we could gather are either national newspapers, specialised or leisure websites from Belgium, France and the Netherlands. A minority of outlets however fall out of these categories as for example the Belgian PTB far-left political party’s website about “Ukraine” or *Boulevard Voltaire* (a French far-right blog) about “Russia”.

---

<sup>17</sup> All raw data is available at <https://github.com/CrossOverSocial/google-news-dataset>



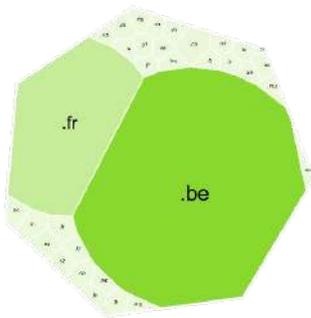
### 5.3. Belgian media outlets are still widely favoured in terms of total articles displayed

Looking at the total articles displayed though, Google News clearly favours Belgian media outlets. Even as there are generally fewer articles from Belgian publishers than from foreign sources, Google News presents Belgian users with an overwhelming majority of articles hosted on Belgian websites in Dutch. In French however, we can observe that on the topics of “Russia” and “Ukraine”, content from French publishers share the stage with Belgian publishers.

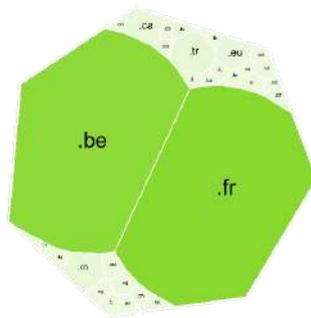
Article displays per ccTLD

Language: FR  
Country: Belgium

Keyword: COVID



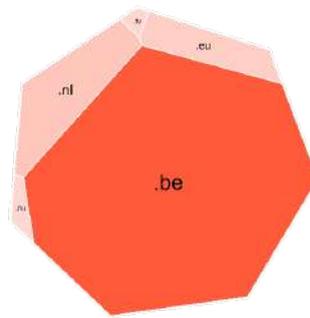
Keyword: Russie



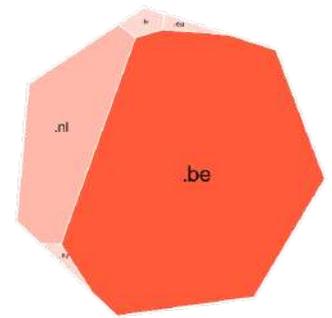
Article displays per ccTLD

Language: NL  
Country: Belgium

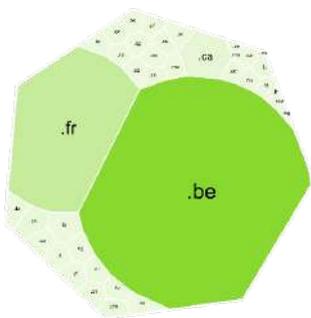
Keyword: COVID



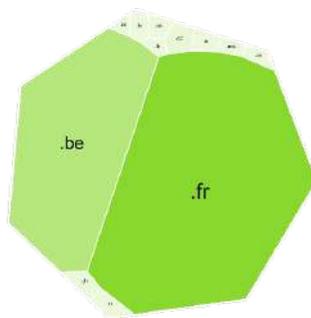
Keyword: Rusland



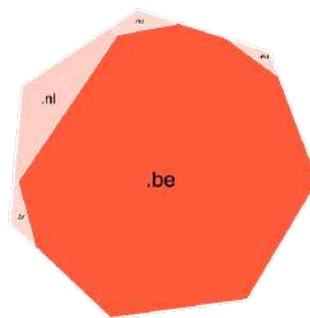
Keyword: Vaccin



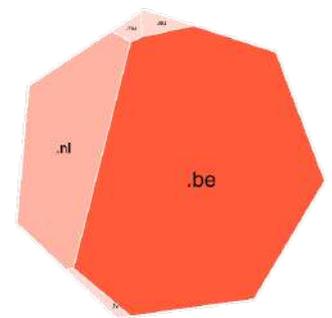
Keyword: Ukraine



Keyword: Vaccin



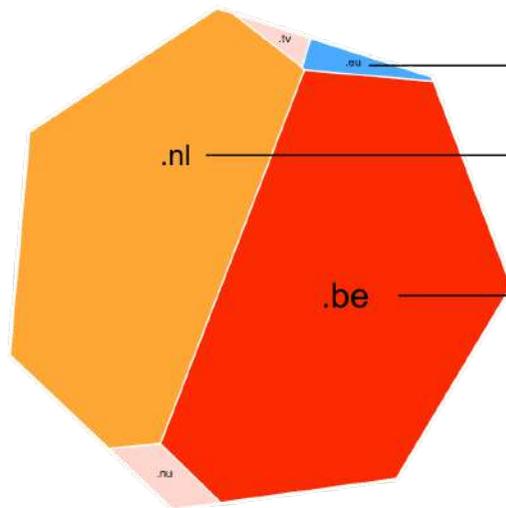
Keyword: Oekraïne



Again, this disproportionate exposition of Belgian content is particularly stark in Dutch, as shown on the diagram below. For all considered keywords, even if articles from Dutch and Belgian sources aggregated by Google News are present in comparable amounts, Belgian users of Google News mostly see articles from Belgian outlets.

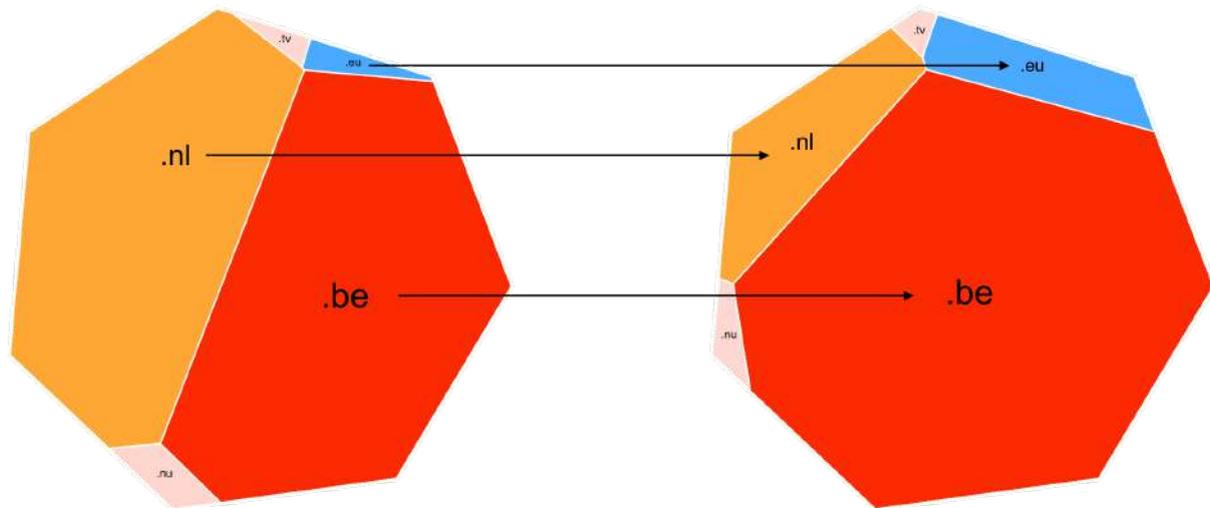
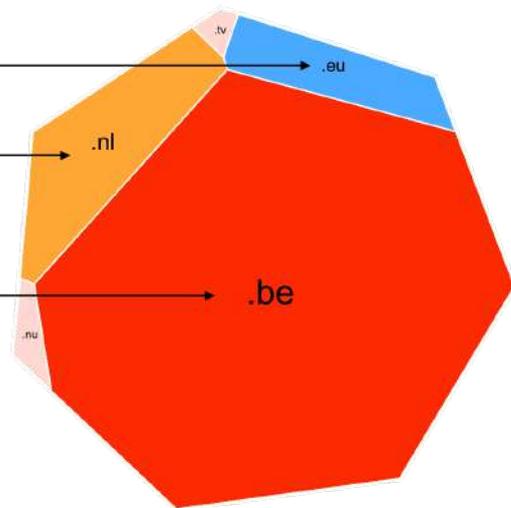
### Sources ccTLDs exposition

Language : NL  
Country : Belgium  
Keyword : COVID



### Articles ccTLDs exposition

Language : NL  
Country : Belgium  
Keyword : COVID





## 6. Conclusions

Examining a year worth of data, CrossOver aimed at evaluating what content Belgian users were exposed to when looking up a selection of neutral keywords on Google News. Most of the time, when searching for terms related to Covid or Russia and Ukraine, results showed news articles from legitimate media outlets to the users.

A search on Xinjiang however presented Belgians with a substantial proportion of pro-Chinese narrative about the autonomous Uyghur province. On the one hand, Google News showed western media outlets covering human rights issues and the repression of Uyghurs in this part of the world. On the other hand, six Chinese state-controlled outlets regularly appeared in the search results, producing positive narratives about the province covering tourism or the economy without any mention of human rights or repression. This set of six Chinese outlets produced a larger amount of articles than the six most displayed western media outlets.

As for exposure of Belgians to foreign content, we observed that Google News tends to favour domestic media outlets in the results by showing articles from Belgian sources more often in the search results. This result depends on the language. Belgian Dutch speakers were more likely to be exposed to domestic content than their French-speaking compatriots. The wider array of available French language outlets make up for a larger foreign content offer. This is reflected in the amount of articles from foreign sources shown.

*Note : we made several attempts to contact various organisations and agencies, including Google, China Internet Information Center in French, People daily in French, New China News Agency (Xinhua), China Radio International, Chine Direct, and the Ministry of Foreign Affairs of the People's Republic of China. At the time of publication, we only received an answer from Google. They pointed us to a [blog post](#) where we "can find more information on how [Google] use[s] different sources for pulling together Google News", which we already mention in the investigation. Google also redirects us to their "How News Works page", where they "have information on how [they] don't take [political positioning into account](#) from publishers". They finally provided us with "this [twitter thread](#) from Search Central about site diversity explains why sometimes many listings from the same source might pop-up in a Search."*

## 7. Methodology

Our investigation was conducted using a dataset built during the CrossOver project. For this particular paper, we analysed data collected from Google News' RSS feed, fetching twice a day which articles were returned by the Google News service for a specific search term<sup>18</sup>. The examined dataset was constituted between 1 February 2022 and 1 February 2023. The procedure for data analysis is described below.

### 7.1. Exportation of the data for the following keywords and languages:

- Covid (fr/nl)
- Vaccin (fr/nl)
- Russie (fr)
- Ukraine (fr)
- Xinjiang (fr)
- Rusland (nl)
- Oekraïne (nl)

### 7.2. Analysis of the results for each keyword calculating the following:

- Total unique URLs per TLD, ccTLD, domain, source
- Total display per TLD, ccTLD, domain, source

### 7.3. Enrichment of the dataset with metadata

We employed the external tool "Extract" from Diffbot to conduct a comprehensive analysis of sources related to the keyword Xinjiang, in order to gain a deeper understanding of the subject. This analysis entailed an automatic examination of the content of each publication, including its publication date, categorisation, tags, etc...

---

<sup>18</sup> see the overall [CrossOver methodology](#)

## 7.4. Creation of data visualisation for the values

TLD exposition per domain & TLD exposition per article

We are using the Treetop (Voronoi) diagrammatical visualisation to display the hierarchies and proportions of the Top Level Domains (TLDs) for the keywords. We only kept the country code TLDs (ccTLDs) from the original data file. The ccTLD list is sourced from <https://www.worldstandards.eu/other/tlds/>.

A limitation of this method is that the Top-Level Domain (TLD) does not always accurately reflect the country of origin of the publication. For example, Belgian publications such as [lavenir.net](http://lavenir.net) are not counted as Belgian due to the .net TLD, and some publications are using the .nu TLD, which means "now" in Dutch.

For the publication timeline graph, we used the "Extract" solution from Diffbot, a powerful tool that enables developers to quickly and easily get valuable structured data from any web page. This data includes the publication date of each article, which we extracted and used to generate the visualisation.

For the word cloud of the body of the articles, we employed the NLP entities extraction from Diffbot, an advanced natural language processing technology that allows developers to quickly and accurately identify and extract critical entities from web pages, such as people, places, and organisations.

To further support the analysis and to ensure accuracy, we then used the sentiment analysis result from Diffbot. This data point allowed us to quickly and accurately identify the sentiment of the content of the articles and measure how it changed over time.

Finally, to ensure the accuracy and reliability of our analysis, we used the Diffbot topic detection tool to identify the most commonly discussed topics in the articles. This enabled us to create an insightful word cloud visualisation, which allowed us to better understand the topics discussed in the articles and gain a deeper insight into the data. In addition, this tool allowed us to compare the frequency of topics discussed in the articles, helping us to identify any trends or patterns in the discussion.

## 7.5 Availability of datasets used for the investigation

All datasets used for this investigation are available on GitHub:

<https://github.com/CrossOverSocial/google-news-dataset>



**We observe and analyse the algorithms of social networks against disinformation**

 [crossover.social](https://crossover.social)  [@eucrossover](https://twitter.com/eucrossover)



 This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement LC-01682253