

# Notes on Proppy

Zubair Abid

Analysis done on [google colab](#).

## Proppy's Data Format

This is based on a glance at the dataset itself, and not whatever is written on the zenodo entry.

1. Article Text
2. Location of the article
3. "Average Tone"
4. Article date
5. GDELT Article ID
6. Article URL
7. Article Author
8. Article Title
9. MBFC Factuality label <sup>1</sup> <sup>2</sup>
10. MBFC URL to Publication
11. Publication Name
12. MBFC notes on the Publication
13. MBFC bias label
14. Publication domain URL
15. **Propaganda label:** 1 is Propaganda, -1 isn't.

## Looking into it

- **Number of articles with MBFC available:** 19631/35986.

This isn't perfect. In a random sample, at least two articles from CNN were found – one labelled correctly as "CNN" and the other as "cnn.com",

---

<sup>1</sup>This is definitely the case, I checked against the entry for the Hartford Courant, Arab News, SHTFPlan, Frontpage Magazine, Korea Herald.

<sup>2</sup>This is also out of date. Arab News is marked as "HIGH", but it is currently on "MIXED" on the website. Also for SHTFPlan, the label is "MIXED" but the reasoning is given instead. This is also the case for Frontpage Magazine. For Korea Herald this is not the case as there is not reasoning given. In general, the reasoning is not given in non-propaganda – probably means that proppey's reasoning for propaganda is the MBFC reasoning.

so the latter did not have the label attached to it. Eg: “Freedom Outpost” is listed correctly 54% of the time.

Probably does not matter for propaganda scores, as non-listed sources can also be propaganda. If we need to correlate MBFC to it however this is an issue. **This might be resolvable by looking at Publication domain URL.**

There are positive propaganda scores associated with unknown sources on MBFC. Example: vdare.com, GDELT ID: 758670980, Row ID: 33946. Another example is a mis-done entry of frontpagemag.

Should clarify that not all “unknown” is incorrectly tagged; ~~something like “Freedom Outpost” is actually not in MBFC (and is considered propaganda in the two random samples it was seen in).~~ The Washington Standard is an example.

- **Number of articles tagged as “propaganda”:** 4021/35986.

Need to investigate: are all articles from a source called propaganda? Is it true the other way round? A quick check on “Freedom Outpost” seems to indicate the former.

Yes. That is the case. There are 10 sources always classified as propaganda, and 159 others that are never.

- **Number of sources classified as propaganda:** 10/169
- **Checking propaganda+ sources against MBFC ratings:**

This information was checked against MBFC as it is on the website on the 21st of February, and not on the Propy dataset (which has it slightly out-of-date).

Source	Factuality Label	Bias Label	Link
Breaking 911	Very Low	-	<a href="https://mediabiasfactcheck.com/breaking911/">https://mediabiasfactcheck.com/breaking911/</a>
SHTFPlan	Mixed	Extreme Right	<a href="https://mediabiasfactcheck.com/shtfplan-com/">https://mediabiasfactcheck.com/shtfplan-com/</a>
Clash Daily	Low	Extreme Right	<a href="https://mediabiasfactcheck.com/clash-daily/">https://mediabiasfactcheck.com/clash-daily/</a>
Personal Liberty	Low	Extreme Right	<a href="https://mediabiasfactcheck.com/personal-liberty/">https://mediabiasfactcheck.com/personal-liberty/</a>
Frontpage Magazine	Low	Extreme Right <sup>3</sup>	<a href="https://mediabiasfactcheck.com/frontpage-magazine/">https://mediabiasfactcheck.com/frontpage-magazine/</a>

Source	Factuality Label	Bias Label	Link
The Washington Standard	-	-	-
VDare	Low	Extreme Right <sup>4</sup>	<a href="https://mediabiasfactcheck.com/vdare/">https://mediabiasfactcheck.com/vdare/</a>
Freedom Outpost	- <sup>5</sup>	-	<a href="https://mediabiasfactcheck.com/fake-news/">https://mediabiasfactcheck.com/fake-news/</a>
Remnant	Low	Extreme Right	<a href="https://mediabiasfactcheck.com/the-remnant-magazine/">https://mediabiasfactcheck.com/the-remnant-magazine/</a>
Lew Rockwell	Mixed	Extreme Right <sup>6</sup>	<a href="https://mediabiasfactcheck.com/lew-rockwell/">https://mediabiasfactcheck.com/lew-rockwell/</a>

- **Checking MBFC Bias labels against Factuality labels**

This will be using the Proppy MBFC Labels, I guess.

It appears that anything that is Low or under on MBFC is considered to be propaganda. SHTFPlan and Lew Rockwell don't follow this rule, but their factuality scores might have changed since Proppy was compiled – we've seen others have.

The table below is compiled using Proppy's MBFC labels. The bias is more coarse-grained than the actual labels as they do not make sense in any publication related to the dataset.

Label	Number	High	Mixed	Low/Very Low	NIL	unknown
NIL	14	9	2	3	0	0
unknown	65	0	0	0	0	65
extreme right	6	0	1	5	0	0
right	8	3	3	2	0	0
right center	6	5	1	0	0	0
least biased	7	7	0	0	0	0
left center	60	56	3	0	1	0

<sup>3</sup>On propy, this is just "Right"

<sup>4</sup>On propy, this is just "Right"

<sup>5</sup>It is not on the site, but is listed under "Questionable Sources". On the other hand, you can see the data in the Proppy dataset itself.

<sup>6</sup>On propy, this is "NIL"

Label	Number	High	Mixed	Low/Very Low	NIL	unknown
left	3	1	2	0	0	0