

Ethos Guidelines (3rd Iteration):

Sentence-based ethos annotation on Social Media

with labels *Ethos Supports* and *Ethos Attacks*

Developed on the PolarIs1 corpus*

Authors of guidelines: blinded for review

Annotators of development corpus: blinded for review

Blinded for review

September 2022

1 Theoretical background

Ethos is defined as the character of the speaker. In its original form, this meant building one's own ethos through speeches made at the beginning of public address [Aristotle, 1991]. For our research purposes, ethos is considered as the character of a speaker (a person, a group of persons or an organisation) which can be *attacked* or *supported* by *other* speakers performing ethotic moves. By looking at them, we can investigate issues such as hate speech on social media, captured as ethotically attacking each other on Twitter or Reddit (for an example of an exchange on social media that we analyse, see Fig. 1).

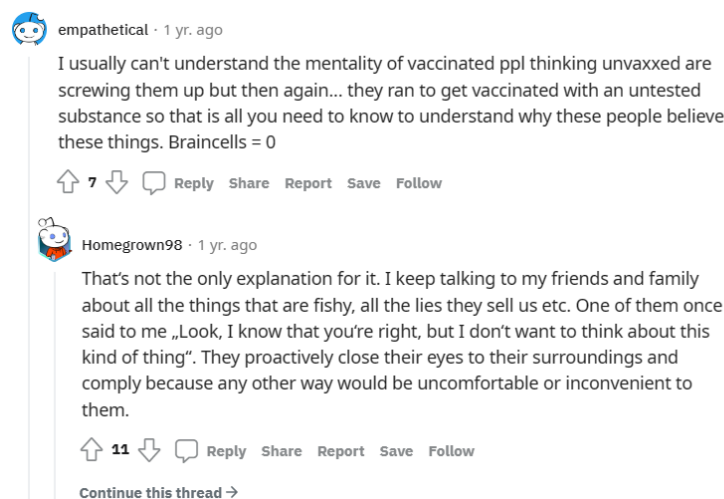


Figure 1: A typical exchange on Reddit extracted into our corpus for the annotation of ethos supports and attacks.

In this study we do not consider cases when a speaker establishes one's own ethos, i.e. instances of speech in which a speaker supports their own ethos, e.g., *I've been living in this town for 20 years and I understand its needs*; or *I've been a leader of teams for my whole life and I know how to manage people*, but

*Available at: <https://github.com/kaatkaa/PolarIs-Corpora>

we do consider utterances in which one speaker is **supporting**, e.g. Ex. (1-a)-(1-c), or **attacking** someone's else ethos such as in Ex. (1-d)-(1-f).

- (1)
- a. bigbadjohn365: *You win the internet today*
 - b. walks_sc: *Yes you are very correct*
 - c. Cosmic_Engineer: *your vigilance is commendable (and should be the norm)*
 - d. SherylBR549: *you're misinterpreting data to promote your own bias*
 - e. 69L46: *Reading / watching CNN is like eating junk food out of the trash*
 - f. walks_sc: *Keep living in fear*

2 Data resources and processing

This annotation scheme has been developed to investigate ethotic appeals in Social Media (hereafter **SM**) discussions on Polarisation Issues (thus we use the abbreviation **PolarIs** as a name of our corpus). PolarIs corpus consists of conversations taking place on two social media, Reddit and Twitter. As a consequence, language you will encounter is very often different from 'school English', filled with colloquialisms, abbreviations and references to affordances of both platforms. The data has been automatically separated into text units that were structured to allow you annotation which will later be used for automated analytics of large datasets.

The purpose of your work is to annotate the file according to instructions in Section 3. Once the file is fully annotated, data gathered in it will allow us to summarise and investigate in analytics. Your attention to linguistic markers of ethotic moves is what allows for construction and training mining algorithms which can analyse large datasets so it's important you decide carefully and consistently on your annotation. You will work on a CSV file, where you will see conversations on selected topics separated into shorter text spans which you will analyse individually. Download the file onto your computer and save it after each work session to avoid losing any progress.

When you open the file, you will see rows with text spans and descriptions of their properties in corresponding columns (see Fig. 2):

- **Column B:** *full_text_id* is an individual, chronological number for a given text unit;
- **Column C-D:** *comment_url* and *comment_id* are an individual url and id for a given text unit;
- **Column E:** *conversation_id* is an individual id of a comment starting the Reddit sub-thread to which a given each text unit was responding;
- **Column F:** *source* is a name of the user who uttered a given text unit;
- **Column G:** *full_text* is an utterance that was posted by the *source* user;
- **Column H:** *content* is a meaning of the utterance is Column G, i.e. what the *source* meant when posting his *full_text*.

Now, your task is to fill in the following columns (see Fig. 3). Read Section 3 to understand how they should be annotated – below you will find only a summary of the annotation task:

- **Column I:** *No_ethos* should be filled with 1 if given text unit did not constitute an ethotic move;
- **Column J:** *Contains_ethos* should be filled with 1 if given text unit constitute an ethotic move;
- **Column K:** *Support* should be filled with 1 if given text unit constitute an ethotic support;

	A	B	C	D	E	F	G
1		full_text_id	comment_url	comment_id	conversation_id	source	full_text
3	1	1	/r/climatechange/comments/ilrxe43	wxmvmx9	snoovision	Excellent	Excellent location for solar
4	2	2	/r/climatechange/comments/ilrziau	wxmvmx9	shanem	India has	India has proven this out for many years.
5	3	2	/r/climatechange/comments/ilrziau	wxmvmx9	shanem	India has	Credit where it's due, the experiment was done by non Americans
6	4	3	/r/climatechange/comments/ils717g	wxmvmx9	Carniboy	good idea	good idea, next step ban the almonds, they are bad for you anyway!
7	5	4	/r/climatechange/comments/ilt4eyi	wxmvmx9	Diafotisi	Based.	Based.
8	6	5	/r/climatechange/comments/iltj5r5	wxmvmx9	UnfairAd7	LOLOL!	LOLOL!
9	7	6	/r/climatechange/comments/iltj7ah	ilrziau	UnfairAd7	Its just as	Its just as dumb, there.
10	8	7	/r/climatechange/comments/ils7app	ils717g	Medeski	While I kn	While I know almonds are terrible for the amount of water needed, how are they bad for you?
11	9	8	/r/climatechange/comments/ils83en	ils7app	Carniboy	Most of t	Most of the calories are fat, high in omega 6 and inflammatory, the protein is not very bioavailable, they are high
12	10	9	/r/climatechange/comments/ils9pdf	ils83en	Harassed	Don't the	Don't they also have cyanide?
13	11	9	/r/climatechange/comments/ils9pdf	ils83en	Harassed	Don't the	Or is that an old wives tale?
14	12	10	/r/climatechange/comments/ilsdty8	ils9pdf	leisureche	Mmmm c	Mmmm cyanide 🤔
15	13	11	/r/climatechange/comments/ils9pdf	ils9pdf	clinicalpsy	According	According to Wikipedia, the free encyclopedia that anyone can edit, yes indeed - Almonds have cyanide in them.
16	14	11	/r/climatechange/comments/ils9pdf	ils9pdf	clinicalpsy	According	So, if I have to choose between almonds or cocoa beans being removed from the economy, I'll choose almonds b
17	15	12	/r/climatechange/comments/ils9pdf	ils9pdf	NewyBlue	Smell the	Smell the same.
18	16	13	/r/climatechange/comments/ils9pdf	ils9pdf	NewyBlue	Cyanide C	Cyanide C/N carbon dioxide CO2

Figure 2: The structure of data before the annotation.

- **Column L:** Attack should be filled with 1 if given text unit contains an ethotic attack;
- **Column P:** Source should be filled with the name of the author of a given text unit;
- **Column Q:** Target should be filled with the name of the target of an ethotic attack or support in a given text unit.

	A	B	C	D	E	F	G	H	I	J	K	L	P	Q
1		full_text_id	comment_url	comment_id	conversation_id	source	full_text	content	No_Ethos	Contains_ethos	Support	Attack	Source	Target
162	49	23	/r/climatechan/ilrq6v	wxb6t8	Travisallen	Get rid of ultra wealthy people.	N and their private jets.				1		1	Travisallen5050 politicians
163	50	23	/r/climatechan/ilrq6v	wxb6t8	Travisallen	Get rid of ultra wealthy people.	N flying all the time.				1		1	Travisallen5050 "world" leaders
164	51	23	/r/climatechan/ilrq6v	wxb6t8	Travisallen	Get rid of ultra wealthy people.	N You know we have a ti		1				1	Travisallen5050
165	52	24	/r/climatechan/ilrg64f	wxb6t8	mikebotao	The best thing you can do is talk	The best thing you car			1				mikebotao
166	53	26	/r/climatechan/ilrlyrt	wxb6t8	jasiekbielecki	One can find a job where he 1) th	One can find a job wh			1				jasiekbielecki
167	54	27	/r/climatechan/ilsw1r	ilqfqk1	Ganajin	So the big takeaway is to have fe	So the big takeaway is			1				Ganajin
168	55	27	/r/climatechan/ilsw1r	ilqfqk1	Ganajin	So the big takeaway is to have fe	I think we are getting			1				Ganajin
169	56	28	/r/climatechan/ilr81wh	ilqfqk1	Coff33qu3st	Did you read that? I know it's fro	Did you read that?			1				Coff33qu3st
170	57	28	/r/climatechan/ilr81wh	ilqfqk1	Coff33qu3st	Did you read that? I know it's fro	know?				1		1	Coff33qu3st Ganajin
171	58	29	/r/climatechan/ils5aa3	ilpx4v2	jcraig87	Pork is terrible for the environme	Pork is terrible for the		1					jcraig87
172	59	29	/r/climatechan/ils5aa3	ilpx4v2	jcraig87	Pork is terrible for the environme	Weirdly enough espar		1					jcraig87
173	60	30	/r/climatechan/ilqphbv	ilqn6yd	Maximum	> Influencing elected officials is se	> Influencing elected c		1					mum8673
174	61	30	/r/climatechan/ilqphbv	ilqn6yd	Maximum	> Influencing elected officials is se	Absolutely this.		1					mum8673

Figure 3: Annotated data with columns describing the presence of ethos ('No ethos' in Column I and 'Contains ethos' in Column J) and the distinction between ethos supports and ethos attacks ('Support' in Column K and 'Attack' in Column L).

3 Guidelines

- ♣ **Rule 3.1 Text Units for annotating ethos.** You need to firstly make sure you identify the text unit you will be annotating. In **Column G** (see Fig. 3), you will find text units of what has been said on SM and this should be annotated. Additionally, you should look at two neighbouring columns: in **Column F**, you will find its source speaker and in **Column H** – its meaning.

Each text unit has two numbers for its identification. In **Column A** you will find a chronological number and in the **Column B** you will find an individual id number. You might use both when you need to identify the exact text unit (it works better than using Excel row numbers).

- ♣ **Rule 3.2 Context for annotating text units with ethos.** Context of the text unit described in **Rule 3.1**, which might be necessary to reconstruct its (non-)ethotic meaning, can be found directly in the spreadsheet annotated or externally on SM:

1. Direct Context: this is a post or posts which precede a given post. For instance, for G448 the direct context is what is above, i.e. in G447, G446, G445 etc. You should look up as many rows above that is necessary for you to better understand what was said in G448.

2. External Context: on rare occasions, you might need to look up for a piece of social media exchange (e.g. Fig. 1) in which a given text unit was posted in order to understand this text unit. For example in Fig. 4, it is difficult to understand who is responding to whom, as there are many users talking to each other. In such cases, you have to search the Internet for a fragment in SM that contains this text unit, and see who is responding to whom according to a tree-structure in which posts are ordered on SM.

	F	G	H	
1	n	source	full_text	sentence
1827		LackmustestTester	> We can look as to whether temperatures did stop in 2013. There has been no war. The global temperature did not increase comparing 2022 to 1988.	
1828		LackmustestTester	> We can look as to whether temperatures did stop in 2013. There has been no war. That's what the numbers tell.	
1829		Left_Insanity	What I find laughable about you trolls is that you jump into this sub, engage in sneaky ad hom attacks - "Thanks for st	
1830		Left_Insanity	What I find laughable about you trolls is that you jump into this sub, engage in sneaky ad hom attacks - "There is overwhelming evidence that climate change is real" - and you expect u	
1831		SftwEngr	I "believe" in only what can be shown reliably with good evidence, so no.	I "believe" in only what can be shown reliably with good evidence, so no.
1832		trufin2038	It does make a difference under 100ppm. But the effect is saturated	It does make a difference under 100ppm.
1833		trufin2038	It does make a difference under 100ppm. But the effect is saturated	But the effect is saturated
1834		GorillaP1mp	The literal title of the paper you reference claims "evidence of deep-ocean Heat up"	The literal title of the paper you reference claims "evidence of deep-ocean Heat uptake during surface-temperat
1835		Turrubul_Kuruman	Bigger list of known names to watch out for, due to participating in fraud, aggressive anti-science work, misrepresent	
1836		Turrubul_Kuruman	Bigger list of known names to watch out for, due to participating in fraud, aggressive anti-science work, misrepresent	

Figure 4: Sometimes the exchange between SM users and the meaning of what they say is not possible to be understood from what was retrieved into the spreadsheet such as in this figure. In such cases, you need to search Internet for original SM posts.

- ♣ **Rule 3.3 Presence of Ethos.** You now need to decide WHETHER a text unit (R3.1) contains an appeal to ethos, i.e. a favourable or unfavourable reference to a speaker such as in Ex. (1-a)-(1-f) and Ex. (2-a)-(2-d); OR it does not mention any speaker, Ex. (2-e); OR it does mention a speaker but she is neither supported nor attacked, Ex. (2-f).

- (2)
- a. mrTrucker87: *The elderly did their part - they paid their taxes.*
 - b. Icefrisbee: *I don't disagree with you* (@Domified)
 - c. GaryGoodson17: *Fauci - a weathervane who creates his own wind.*
 - d. DamitCyrill: *BBC needs shit down.*
 - e. stmefreak: *It is going to be an interesting winter*
 - f. Icy-Ad-5551: *Is this Anthony Fauci*

Notice that a speaker means here any person, group of persons or organisation regardless of whether they participate in a discussion or they are simply a third party that is mentioned during the discussion. For instance in Ex. (2), Fauci (a person), the elderly (a group) or the BBC (an organisation) will be called 'a speaker' even if they do not post anything in the discussion on SM, but they are mentioned on Reddit or Twitter by users in the discussion. A speaker is also another SM user (such as Domified in Ex. (2-b)) with whom a given source SM user (Icefrisbee in this example) interacts directly in the discussion.

1. **Contains Ethos** is a label that you apply (marked as '1' in **Column J** in Fig. 3) when a text unit contains an appeal to ethos, i.e. a favourable (positive) or unfavourable (negative) reference to a person, a group of persons or an organisation. In these cases, we say that such a text unit is an ethotic sentence or an ethotic appeal.

Notice that a sentence is an ethotic appeal only if both (1) it mentions a speaker; and (2) she is mentioned in a favourable or unfavourable manner. This means that a reference to a speaker is not enough to apply 'Contain Ethos' label to a sentence. For example, even though Fauci is mentioned in Ex. (2-f), he is neither supported nor attacked by Icy-Ad-5551. Thus, (2-f) does not contain ethos (see **Rule 3.2.2**). This means that in Ex. (2) only (2-a)-(2-d) should be annotated with the label 'Contains Ethos'.

2. **No Ethos** is a label that you apply (marked as '1' in **Column I** in Fig. 3) when a text unit does not contain an appeal to ethos.

♣ **Rule 3.3 Polarisation of Ethos.** If you selected that a text unit contains ethos (**R3.2.1**), then you need to decide WHETHER the ethotic appeal/reference is positive/favourable (+ive) towards a speaker; OR it is negative/unfavourable (-ive) towards a speaker.

1. **Ethos Support** is a label that you apply (marked as '1' in **Column K** in Fig. 3) when a text unit contains a +ive appeal to ethos, i.e. there is a favourable reference to a person, a group of persons or an organisation. In these cases, we say that such a text unit is an ethotic support, see Ex. (1-a)-(1-c) and Ex. (2-a)-(2-b).
2. **Ethos Attack** is a label that you apply (marked as '1' in **Column L** in Fig. 3) when a text unit contains a -ive appeal to ethos, i.e. there is an unfavourable reference to a person, a group of persons or an organisation. In these cases, we say that such a text unit is an ethotic attack, see Ex. (1-d)-(1-f) and Ex. (2-c)-(2-d).

♣ **Rule 3.4 Speakers.** If you selected that a text unit contains ethos (**R3.2.1**), then you need to annotate also two types of speakers: an author of the ethotic appeal AND a referent of the ethotic appeal.

1. **Source** is a label which gives the name of an author of the ethotic appeal. You should insert it in **Column P**. In this study, the spreadsheet is pre-processed in such a way that this task comprises just of copying-and-pasting Column F into Column P.
2. **Target** is a label which gives the name of a referent of the ethotic appeal. You should insert it in **Column Q**. In Ex. (2), the Target speakers are: the elderly, Domified, Fauci, BBC, in (2-a)-(2-d) respectively.

Note that you have to resolve pronouns when you annotate targets which means that in cases where referents are mentioned as 'you', 'they' (see Ex. (1)), you should look up the context (**Rule 3.2**) in order to understand whom these pronouns mean and include in Column Q the reconstructed name of the Target speaker.

References

Aristotle. *On Rhetoric*. Oxford University Press, New York, 1991.