

Taxonomy Overview

meta-chara- cteristic	dimen- sion	subdimension	feature	code	characteristics		
					0	1	
detection	complexity features	headline	length of the headline	headlength	low	high	
			textual gap between title and body	headgap	no	yes	
			sentence-like headline	headsent	no	yes	
		corpus	relatively shortness of body	corpshort	no	yes	
			simplicity in sentence structure	corpsimpl	no	yes	
			lexical and contentual poorness	corplex	no	yes	
			relatively high amount of typographical errors	corperror	low	high	
			relatively low demand on reader's education level	compeduc	no	yes	
			comprehension	level of semantic incoherence	compincoh	low	high
		informativity		extremity of information quantity	infoextrem	low	high
				lack of new information	infonewinfo	no	yes
			lack of topical redundancy	infotopredun	no	yes	
	psychology features		mobilization	level of emotional polarization	mobpolar	low	high
		level of sensationalism		mobsensat	low	high	
		arousal of (negative) affects		mobaaffect	low	high	
		level of topicality		mobtopical	low	high	
		subjectiveness	tendency to subjective statements	subjtenden	low	high	
			level of personal motives	subjmotiv	low	high	
	stylistic features	vocabulary	kind of discourse	subjdiscours	knowledge-based	opinion-based	
			usage of exaggerated vocables	vocexagg	no	yes	
			amount of first-/second-person pronouns	vocpronoun	low	high	
		phrasing	amount of excessive emotional adverbs	vocadverb	low	high	
			usage of emphatic wording	phrasemph	no	yes	
			level of informality of language	prhasinformal	low	high	
		authenticity	vagueness of phrasing	authvague	low	high	
			authenticity/referencing of information	authrefer	frequently referenced	poorly referenced	
categorizing	themes of fake news	political & economic		thempoleco	no	yes	
		social		themsoc	no	yes	
		pseudoscientific		themscience	no	yes	
		historical		themhisto	no	yes	
		gossip/rumor		themgoss	no	yes	
		extreme		themextrem	no	yes	
		worldview		themworld	no	yes	
		commercial		themcommer	no	yes	
	content type	clickbait		typclick	no	yes	
		manipulated content		typmanipul	no	yes	
		fabricated content		typfabric	no	yes	
		false context		typfalse	no	yes	
		imposter content		typimpost	no	yes	
		social bot content		typbot	no	yes	
		conspiracy theory		typconspir	no	yes	
		one-sided		typonesid	no	yes	
		propaganda		typpropa	no	yes	
		trolling		typtroll	no	yes	
grade of veracity	mostly true		vtrue	no	yes		
	mixture of true and false		vtruefalse	no	yes		
	mostly false		vfalse	no	yes		
	no factual content		vnofact	no	yes		

Feature Description

code	feature	description
COMPLEXITY		
headline		
headlength	length of the headline	Unreliable sources try to convey as much information as possible in the title to draw the reader's attention. Thus they use a higher amount of plaintext/words.
headgap	textual gap between title and body	Unreliable sources like clickbaits tend to have a lower textual similarity between the title and the body of the article.
headsent	sentence-like headline	Titles of fake content often present sentence-like claims about people and entities associating them with actions to convey as much information as possible.
corpus		
corpshort	relatively shortness of body	Unreliable sources tend to have a lower level of plaintext/number of words in relation to real articles.
corpsimpl	simplicity in sentence structure	Sentences of fake news exhibit a lower complexity in structure and a relatively low amount of words.
corplex	lexical and contentual poorness	Fake articles incline to have less diversity at the lexical and content level.
corperror	relatively high amount of typographical errors	Unreliable sources empirically exhibit a higher amount of typographical errors.
comprehension		
compeduc	relatively low demand on reader's education level	Reliable articles tend to be written in a more complex way regarding readability measures like number of syllables per word or number of words per sentence. A higher complexity increases the effort to apprehend the content and thus works as an indicator for the demanded education level of the reader.
compincoh	level of semantic incoherence	Real articles contain sentences with correlated pairs of content-words and sentences that are correlated with each other. Unreliable sources exhibit a lower level of sentence/word correlations which is called semantic incoherence.
informativity		
infoextrem	extremity of information quantity	Fake articles often correspond with either a considerably low amount of information or a remarkable overload of information.
infonewinfo	lack of new information	The body of fake news articles adds relatively little new information, but serves to repeat and enhance the claims made in the title.
infotopredun	lack of topical redundancy	Valid articles about a particular topic contain several direct or indirect references to this subject. One can interpret those as a kind of contentual redundancy which fake sources are usually missing.
PSYCHOLOGY FEATURES		
mobilization		
mobpolar	level of emotional polarization	Unreliable sources incline to use a more emotionally persuasive language in general.
mobsensat	level of sensationalism	Fake articles tend to be written in a hyperbolic way to attract the reader's

code	feature	description
		attention i.e. with a high usage of all-caps-words, exclamation marks or a general sentiment wording.
mobaaffect	arousal of (negative) affects	Fake content uses a higher degree of words related to emotional actions, states and processes.
mobtopical	level of topicality	Legitimate sources tend to report about past events whereas fake articles focus on highly recent topics.
subjectiveness		
subjtenden	tendency to subjective statements	Fake articles tend to be written from a more personal view.
subjmotiv	level of personal motives	Creators of fake content are often driven by personal motives like rising profit, promoting ideology or psychological aims (and suchlike).
subjdiscours	kind of discourse	Words and expressions of fake articles relate to a more argumentative discourse aiming to convince the reader of a specific point of view.
STYLE FEATURES		
vocabulary		
vocexagg	usage of exaggerated vocables	Unreliable sources more often use hyperbolic words as superlatives and subjectives.
vocpronoun	amount of first-/second-person pronouns	First-person and second-person pronouns are empirically more often used in fake news than in legitimate articles.
vocadverb	amount of excessive emotional adverbs	Fake articles indicate a higher amount of adverbs like 'often', 'undoubtedly', 'heavily' to lure the reader to the content.
phrasing		
phrasemph	usage of emphatic wording	Unreliable sources use a high level of exclamation marks, swear words, visual references and are slightly more prone to emotional tones and higher polarity.
prhasinformal	level of informality of language	The language of unreliable sources tends to be less formal than reliable articles.
authenticity		
authvague	vagueness of phrasing	Fake articles use a higher amount of vague phrasing or hedging words (like 'possibly', 'usually', 'tend to be') to achieve a more indirect form of expression.
authrefer	authenticity/referencing of information	Legitimate sources are considerably better referenced than unreliable articles.

Themes and Categories

code	theme/category	description
THEMES		
thempoleco	political & economic	About specific politicians, or legal, political or economic actions.
themsoc	social	About social events, activists, public benefit and minority organizations, as well as dangers or threats to human and animal health.
themscience	pseudoscientific	Content that calls on supposedly scientific research or reputable institutions without identifying concrete sources or by manipulating them to create a false theory.
themhisto	historical	About historical events or the distant past of public figures.
themgoss	gossip/rumor	About celebrities, actors and public figures without a political or activist profile.
themextrem	extreme	About drastic, catastrophic or brutal events.
themworld	worldview	About religion, faith, and spiritual figures as well as various non-religious ideologies, views, and beliefs
themcommer	commercial	Like false product reviews, advertising campaigns or commercial clickbait aimed at accumulating views, likes and comments (such as manipulated photos of animals).
CATEGORIES		
typclick	clickbait	Sources that provide generally credible or dubious factual content but deliberately use exaggerated, misleading or unverified headlines/thumbnails to lure readers to open the intended Web page. The goal is to increase their traffic for profit, popularity, or sensationalization. Once the reader gets there, the content rarely satisfies their interests.
typmanipul	manipulated content	Manipulating information or image to deceive the recipient. The recipient receives information prepared in such a way that he is not aware of its misuse.
typfabric	fabricated content	Stories that lack any factual ground, 100 % false. The intention is to deceive and cause harm. One of the most severe types of fabrication adopts the style of news articles so the recipients believe it is legitimate. Could be text or visual media.
typfalse	false context	It consists of real information but in a false context. The recipient is aware that the information is true, but he does not realize that the context has been changed.
typimpost	imposter content	Genuine sources that are impersonated with false, made-up sources to support basically a false narrative. It is actually very misleading since a source or author is considered a great criteria of verifying credibility. (use of journalists name/ logo /branding of mimic URLs)
typbot	social bot content	A social bot is a computer algorithm that automatically produces and posts content and is able to interact with legitimate users and other bots, trying to emulate and possibly alter their behavior.
typconspir	conspiracy theory	Stories without factual basis which usually explain important events as secret plots by government or powerful individuals. By definition their truthfulness is difficult to verify. Evidence refuting the conspiracy is regarded as further proof of the conspiracy.
typonesid	one-sided	Stories that are extremely biased toward a person/party/situation/event driving division and polarization. The context of this type of news information is extremely imbalanced (i.e. left or right wing), inflammatory, emotional and often riddled with untruths. They contain either a mixture of true and false or mostly false, thus misleading information designed to confirm a particular ideological view.

Grades of Veracity

code	grade	description
GRADES OF VERACITY		
vtrue	mostly true	The post and any related link or image are based on factual information and portray it accurately. This lets the authors interpret the event/info in their own way, so long as they do not misrepresent events, numbers, quotes, reactions, etc., or make information up. This rating does not allow for unsupported speculation or claims.
vtruefalse	mixture of true and false	Some elements of the information are factually accurate, but some elements or claims are not. This rating should be used when speculation or unfounded claims are mixed with real events, numbers, quotes, etc., or when the headline of the link being shared makes a false claim but the text of the story is largely accurate. It should also only be used when the unsupported or false information is roughly equal to the accurate information in the post or link. Finally, use this rating for news articles that are based on unconfirmed information.
vfalse	mostly false	Most or all of the information in the post or in the link being shared is inaccurate. This should also be used when the central claim being made is false.
vnofact	no factual content	This rating is used for posts that are pure opinion, comics, satire, or any other posts that do not make a factual claim. This is also the category to use for posts that are of the "Like this if you think..." variety.