

1.8 Machine Generated Sample Content

To demonstrate the versatility of our approach, Tables W1.8.1, W1.8.2, and W1.8.3 present abbreviated samples of machine-generated content for keywords from varied industries (“best e bike insurance,” “aerospace component manufacturer,” and “state library bookshop”) that have not yet been revised by a human. For comparison reasons each of Tables W1.8.1, W1.8.2, and W1.8.3 contain a real example of top 10 and worst 10 search engine ranked content (i.e., in the examples presented here we used Google’s search engine rankings 290-300). As in Table 2, the texts are shown with their associated quality score components. We can see that while lower ranked real content is both off-topic and performs poor in terms of quality score components (except for uniqueness, for reasons as already discussed in section 1.4 of this Appendix), the machine generated content typically matches those of the top 10 samples very well and points to subtle differences not immediately transparent to human readers.

Table W1.8.1: Example Generated Piece of Content vs. Top & Worst ranking Content

Stage	Quality Score	Generated Content
GPT-2 fine-tuned (not revised)	$s_a = .65$	Best e bike insurance
	$s_k = .73$	Best value electric bike insurance
	$s_d = .84$	
	$s_n = .58$	We offer a range of products to suit every budget and circumstance. Our products can
	$s_r = .62$	provide additional protection for theft, damage, accident, negligence, fraud and more. Our
	$qs_g = .144$	customer service reps are here to tailor a program to your needs.
		Bicycle Roadside Assistance
		If you find yourself with a disabled bike well come pick you up. [...]
Real top10 ranked content	$s_a = .58$	The importance of e-bike insurance
	$s_k = .61$	
	$s_d = .96$	Electric bikes (commonly referred to as e-bikes) are expensive, high potential risks on the
	$s_n = .50$	road typically not fully covered by homeowners, renter or auto insurances. Its a risky move
	$s_r = .79$	for your e-bike to be uninsured. Fortunately, our bicycle insurance is designed specifically
	$qs_g = .133$	for e-bikes to give you the peace of mind you deserve. Lets take a closer look at why e-bike insurance is a good move. E-bikes can be expensive. [...]
Real worst10 ranked content	$s_a = .19$	Cycle brands in mumbai
	$s_k = .08$	
	$s_d = .96$	And summer seasons are the perfect time to shape yourself into an envious body, dont you
	$s_n = .17$	think so? You would be thinking why are we talking [Continue reading]
	$s_r = .21$	Top 20 Sportswear Brands for Men and Women in India Metro Cycle Company in Charni
	$qs_g = .001$	Road, Mumbai. Know more about latest cycles and new cycles in India 2020. [brand name] [brand name] ([Phone number]) is the leading Bicycle Dealer In India. Incense sticks nowadays are made up of aromatic materials that produce aroma when burned. Cycles in India Get all Cycles information like cycle prices, reviews, photos, etc. . There are more cycle brands in India but we have listed only 25, drop your comment for your best cycle. , one of the largest cycle manufacturers in India is now emerging as best Bicycle Companies online. [...]

Human revision in our field experiment reported below: = identifiers like brand names and phone numbers are replaced by a tag (e.g., [brand name]) to retain confidentiality; headlines are printed in bold to ease reading;

Table W1.8.2: Example Generated Piece of Content vs. Top & Worst ranking Content

Stage	Quality Score	Generated Content
GPT-2 fine-tuned (not revised)	$s_a = .40$	Aerospace component manufacturer
	$s_k = .40$	
	$s_d = .88$	With a revenue of \$8 billion, [brand name] is the third-largest aerospace company in the world. It provides products for both global aerospace and defense, manufacturing nacelle systems, avionics systems, control surfaces, electrical systems, mechanical and aerospace systems, and flight simulator systems.
	$s_n = .92$	
	$s_r = .98$	The company also provides aftermarket services, like draping and electrical systems, in addition to product development.
	$qs_g = .127$	[brand name] specializes in fabricating complex components and sub-assemblies for military and commercial uses. Examples of products it creates are for use in communication systems, missiles and aviation defense, and in the aerospace industry generally. [...]
Real top10 ranked content	$s_a = .31$	[brand name]: World Leaders in Aerospace
	$s_k = .38$	
	$s_d = .99$	Since Flight Began [brand name] comprises one of the largest, most experienced concentrations of world-class aerospace companies, the world's AEROSPACE ALLEY! Our firms grew up where aerospace was born. Today, our precision components take to the skies every day, in every corner of the globe, as they have since the inception of powered flight.
	$s_n = .25$	
	$s_r = .60$	[brand name] member companies are united by a single goal; be the world leader in providing customers with aerospace components of unsurpassed quality, at competitive prices and always, on-time delivery. To achieve that goal, we work with our peers and with regional, national and international experts to ensure continuous improvement of our products and services. [...]
	$qs_g = .017$	
Real worst10 ranked content	$s_a = .07$	Aerospace Industry
	$s_k = .05$	
	$s_d = .91$	Aerospace Industry
	$s_n = .17$	
	$s_r = .13$	TOTAL: 8 HORNET Series INSPIRER Series GRANDER 5Ax Series DBC2000mm DBC1500mm GRANDER 5MG Series DBC3100mm DBC2500mm Industry Aerospace Industry GRANDER 5Ax Series GRANDER 5MG Series HORNET Series INSPIRER Series DBC 1500mm DBC 2000mm DBC 2500mm DBC 3100mm Automobile HE Series HT Series VTW Series VA Series Medical Sphere VTH Series VTP Series VTJ Series VTG Series Large Molding DBC 1500mm DBC 2000mm DBC 2500mm DBC 3100mm Electronics Products HF Series VA Series VF Series VTT Series Parts Processing HE Series HF Series VTT Series VTW Series VF Series VA Series VH Series VE Series VK Series VP Series VTH Series TS Series CNC Lathes T Series CNC Lathes TEL: [phone number] FAX: [fax number] Email: [email] Address: [address] TEL: [phone number] Email: [email] Address: [address] 2018 [brand name] All rights reserved.
	$qs_g = .000$	

Human revision in our field experiment reported below: [] = identifiers like brand names and phone numbers are replaced by a tag (e.g., [brand name]) to retain confidentiality; headlines are printed in bold to ease reading;

Table W1.8.3: Example Generated Piece of Content vs. Top & Worst ranking Content

Stage	Quality Score	Generated Content
GPT-2 fine-tuned (not revised)	$s_a = .25$	State library bookshop
	$s_k = .22$	open seven days a week
	$s_d = .88$	[...]
	$s_n = .92$	\$10 kids 12 and under
	$s_r = .98$	Free playtime on the playroom terrace
	$qS_g = .042$	Discover the wonders of Queensland
		The [brand name] Historical Library is Queensland's first permanent, accessible, fully-staffed library. The Library is home to a beautifully curated selection of Queensland's best contemporary and classic books, and offers free access to more than 220,000 unique readers each day. The Library is Brisbane's only permanent non-profit library, and is funded entirely by the State Government. [...]
Real top10 ranked content	$s_a = .27$	Readings State Library
	$s_k = .25$	Readings State Library is the first permanent commercial bookshop permitted to be established in the building in the Library's 150 year history. It is a unique space that caters to students, travellers, book lovers, CBD workers and library regulars. As part of the State Library Victoria's Vision 2020 redevelopment project, our shop was relocated to a welcoming new space beautifully designed by Australasian design firm Architectus with Scandinavia's Schmidt Hammer Lassen Architects. [...]
	$s_d = .96$	
	$s_n = .42$	
	$s_r = .72$	
	$qS_g = .019$	
Real worst10 ranked content	$s_a = .11$	About the Author
	$s_k = .03$	Tasmanian writer Katherine Johnson is the author of four novels. Paris Savages based on a true story of human zoos will be published in October 2019. Katherine's third novel, Matryoshka or Russian dolls (Ventura Press 2018), is a story of secrets, refuge and loves lost and found. Her previous novels include The Better Son (Ventura Press 2016), set in northern Tasmanias caves, and Pescadors Wake (Fourth Estate 2009), the story of the danger and heartbreak of lives at the mercy of the sea during a three-week Southern Ocean chase. Katherine is the recipient of The University of Tasmania Prize and the Peoples Choice Award (Tasmanian Literary Prizes) and HarperCollins Varuna Awards. The Better Son was longlisted for the Australian Indie Book Awards 2017 and the Tasmania Book Prize (Premiers Literary Prizes 2017). [...]
	$s_d = .96$	
	$s_n = .25$	
	$s_r = .74$	
	$qS_g = .000$	

Human revision in our field experiment reported below: [] = identifiers like brand names and phone numbers are replaced by a tag (e.g., [brand name]) to retain confidentiality; headlines are printed in bold to ease reading;

Appendix References

- Baayen RH, Shafaei-Bajestan E (2019) Analyzing linguistic data: A practical introduction to statistics. Package 'languageR'. Version 1.5.0. CRAN. Accessed May 20, 2019, <https://cran.r-project.org/web/packages/languageR/languageR.pdf>
- Benoit K, Watanabe K, Wang H, Nulty P, Obeng A, Müller S, Matsuo A, (2018) “quanteda: An R package for the quantitative analysis of textual data.” *Journal of Open Source Software*. 3(30). <https://doi.org/10.21105/joss.00774>
- Berger J, Sherman G, Ungar L (2020b) TextAnalyzer. Accessed November 11, 2020, <http://textanalyzer.org>
- Bronnenberg BJ, Kim JB, Mela CF (2016) Zooming in on choice: How do consumers search for cameras online? *Marketing Science*. 35(5):693-712.
- Danaher PJ, Mullarkey GW, Essegai S (2006) Factors affecting website visit duration: A cross-domain analysis. *Journal of Marketing Research*. 43(2):182-194.
- Edelman B, Zhenyu L (2016) Design of search engine services: Channel interdependence in search engine results. *Journal of Marketing Research*. 53(6):881-900.
- Flanigan, AJ, Metzger, MJ (2007) The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society*. 9(2):319-342. <https://doi.org/10.1177/1461444807075015>
- Jerath K, Ma L, Park YH (2014) Consumer click behavior at a search engine: The role of keyword popularity. *Journal of Marketing Research*. 51(4):480-486.
- Kamoen N, Holleman B, Bergh H (2013) Positive, negative, and bipolar questions: The effect of question polarity on ratings of text readability. *Survey Research Methods*. 7(3):181-189.
- Liu J, Toubia O (2018) A semantic approach for estimating consumer content preferences from online search queries. *Marketing Science*. 37(6):930-952.
- Maechler M, Rousseeuw P, Croux C, Todorov V, Ruckstuhl A, Salibian-Barrera M, Verbeke T, Koller M, Conceicao ELT, Palma MA (2020) Basic robust statistics. Package 'robustbase'. Version 0.93-6. CRAN. Accessed May 20, 2020, <https://cran.r-project.org/web/packages/robustbase/robustbase.pdf>
- Pennebaker JW, Booth RJ, Boyd RL, Francis ME (2015) Linguistic inquiry and word count: LIWC2015. Austin, TX: Pennebaker Conglomerates. Accessed November 1, 2020, www.LIWC.net.

Pitler E, Nenkova A (2008) Revisiting Readability: A unified framework for predicting text quality. *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. 186-195.

Radford A, Narasimhan K, Salimans T, Sutskever I (2018) Improving language understanding by generative pre-training. OpenAI.

Roberts C (2010) Correlations among variables in message and messenger credibility scales. *American Behavioral Scientist*. 54(1):43-56.

Rocklage MD, Rucker DD, Nordgren LF (2018) Persuasion, emotion and language: the intent to persuade transforms language via emotionality. *Psychological Science*. 29(5):749-760.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. *31st Conference on Neural Information Processing Systems (NIPS 2017)*. 1-15.