2.7 Content Production Cost Calculation Details

Table 4 in the article presents the cost of content production and savings induces by applying the content writing machine. In this section, we elaborate on the calculations for Table 4. We took available working times and salary statistics for the human reviser / SEO expert necessary (i.e., as stated in the Table's footer: 39 hours available working time per week, 1,567 hours per year; 45,000 € of salary per year) and calculated the times, cost and possible output per year when using the manual way vs. the machine for text generation. To estimate the time spent per unit of content, information was provided by both the company and the experiment participants. Based on this information, we can calculate expected outputs and labor cost per year. The calculation in Table 4 in the main manuscript are based on the following inputs, with the values from Table 4 appearing in quotations:

- "Human labor time for content production" = empirically determined

- "Server cost per unit (€)" = empirically determined

- For human groups: "Produced content units" = 1,567 hours per year / "Median (hours)" For example, for the column "Company (real)" in Table 4: $1{,}567/9.5 \approx 164.95$

- For machine to keep total costs at 45,000 € (i.e., the same as the human costs): "Produced content units" = 45,000 € / ("Labour cost per unit (€)" + "Server cost per unit (€)"). For example, for the column "Revised Machine" in Table 4: $45{,}000/(15.79+5.00) \approx 2{,}164.03$

- "Production level (%)" = ("Produced content units" (e.g., of the revised machine) / "Produced content units" Company (real))*100-100. For example, for the column "Revised Machine" in Table 4: $(2{,}164.03/164.95)*100-100 \approx 1{,}211.95$

- "Labor cost per unit (€)" = 45,000 € / "Produced content units". For example, for the column "Company (real)" in Table 4: 45,000/164.95 ≈ 272.81

- "Cost for 164.95 units (€)" = ("Labour cost per unit (€)" + "Server cost per unit (€)")*164.95. For example, for the column "Company (real)" in Table 4: 272.81+0*164.95 ≈ 45,000

- "Cost for 2,164.03 units (€)" = ("Labour cost per unit (€)" + "Server cost per unit (€)")*2,164.03. For example, for the column "Company (real)" in Table 4: 272.81+0*2,164.03 ≈ 590,369

- "Produced content units" ("Possible real financial impact (2015 to 2019)") = empirically determined

- "Cost (€)" = 439*("Labour cost per unit (€)" + "Server cost per unit (€)"). For example, for the column "Company (real)" in Table 4: 439*(272.81+0) ≈ 119,765

- "Possible savings (€)" = "Cost (€)" of the Company (Real) – "Cost (€)" of specific comparison group. For example, for the column "Revised Machine" in Table 4: 119,765-9,127 ≈ 110,638

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, Essegaier 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 attribtues, 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, Gomze AN, Kaiser L, Polosukhin I (2017) Attention is all you need. *31st Conference on Neural Information Processing Systems (NIPS 2017)*. 1-15.