

Fully Automatic Journalism: We Need to Talk About Nonfake News Generation

Anya Belz

Computing, Engineering and Mathematics

University of Brighton

Lewes Road, Brighton BN2 4GJ, UK

a.s.belz@brighton.ac.uk

1 Dangerously good text generation?

It would have been hard this year not to be aware of the furore around OpenAI’s ‘too dangerous to release’ GPT-2 text generator,¹ with one carefully selected text about unicorns in particular cropping up everywhere (Figure 1). Text generation using large neural language models trained on large datasets (Zellers et al., 2019; Radford et al., 2019) is reaching headline grabbing levels of quality, with coverage focusing on the dangers of misuse, in particular for fake news generation.

There are factors that make these models less dangerous than the headlines imply: one, the content of generated texts is not fully controllable beyond general topic, with a lot of unpredictability in the output. Two, texts generated with models like GPT-2 are easy to spot, because they contain tell-tale inconsistencies and non sequiturs (e.g. ‘four-horned unicorns’ and ‘two centuries old mystery’ in Figure 1). Besides, the very techniques that make the models work so well for generation, also make it easy to automatically detect text generated by them, using tools like GLTR (Gehrmann et al., 2019) and Grover (Zellers et al., 2019).

Meanwhile, the first fully automatic news generators are going into production without much media furore or awareness among readers, driven by very different, fully controllable and as yet undetectable, text generation technologies.

2 Fully automated news is a reality

Automation has been making inroads in newsrooms for over a decade (Bloomberg and AP being early adopters), but it’s only now that we’re seeing the first fully automatic news agencies. Monok (Sweden) runs a news generation system² that is an extractive-abstractive multi-document

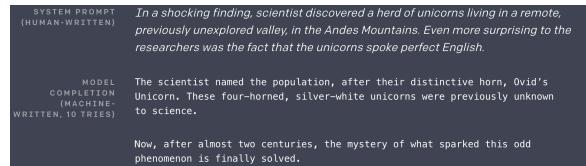


Figure 1: Sample prompt and beginning of completion generated with GPT-2 (after 10 tries).¹



Figure 2: Published news articles generated by Monok (2019-06-17), and Radar (Nilsson, 2019).

multi-modal summarizer, combining neural pre-processing for coreference, named entities, etc., with a coherence-preserving rule-based extractor, and a neural abstractor which rewrites the text so as to create a sufficiently new piece of writing. The resulting articles are validated via third-party plagiarism detectors, and augmented with scraped videos, social media embeds, Wikipedia links, etc.³ For sample published output see Figure 2.

Radar AI (UK) uses a system based on Arria NLG Studio⁴ which uses a template-scripting language to control text generation from structured data. Radar uses periodic-release public data sets (e.g. NHS) as input; for each of these, its editors write templates structured into decision trees that control which templates are used, which data is selected and where it is inserted, depending on date, identity and location of client newspaper, etc. Following this one-off effort, article generation is fully automatic for a given data source. For sample published output see Figure 2, right.

¹<https://openai.com/blog/better-language-models/>

²www.monok.com/en

³Details of technologies as provided by Monok and Radar.

⁴<https://www.arria.com/studio/studio-overview/>

3 Issues with nonfake news generation

Radar and Monok are benefiting from the coming of age of language generation and summarisation technologies, and utilise them to address specific needs in the news production sector, in the former case enabling regional newspapers to survive in an age of falling ad revenues, and in the latter case providing a cost-effective way to create high-quality factual content that does well in SEO terms. The underlying methods start from structured data or existing news articles, and are not easily adapted for fake news generation.

This doesn't mean that nonfake news generation doesn't come with its own risks. There is potential for intentional misuse: automatically generated news (AGN) can be arbitrarily selective in content, slanted to further specific agendas, and shot off to any number of destinations at the touch of a button, potentially overwhelming news output, and drowning out more balanced accounts. As with fake news, automatic detection would offer some leverage (based on spread patterns and text properties, in the absence of fake content that can be fact-checked). Unfortunately GLTR-like automatic detection doesn't work for summarisation and template generation, because these incorporate (paraphrased) word strings written by humans. Figure 3 shows GLTR results for a Monok article deemed slightly more likely to be human-authored than the similar, actually human-authored, article above it (more red and purple, more human).

Humans and search engines don't do much better. Both Monok's and Radar's articles are published without human postprocessing (although in the latter case, editors add local information), indicating a level of quality that readers find hard to distinguish from human-written. Monok escapes all Google penalties for automatically generated and non-unique content, meaning it passes for human-written and new, despite being neither.

Even with bona fide use, bias and misrepresentation can creep into generated articles from source data, and automatic selection of input risks omission of contextualising information resulting in misrepresentations or even untrue entailments.

But perhaps the elephant in the room is transparency. Automation has some hand in producing most of the news we read now, but you would never know it from attributions. Newspapers are running full AGN under human bylines, some editors believing readers "couldn't care less who

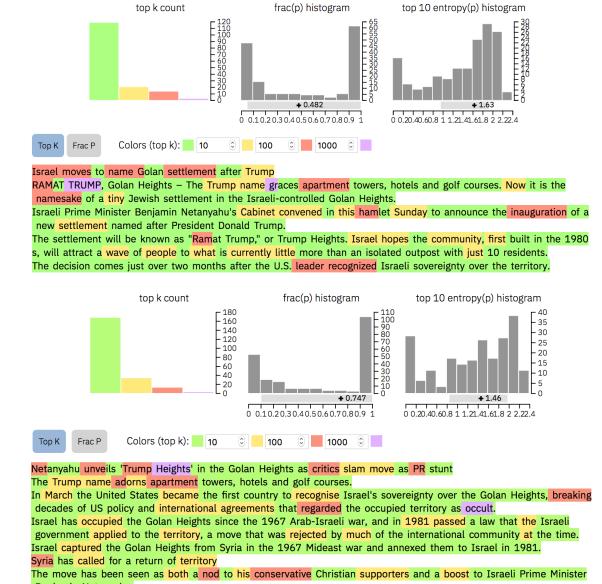


Figure 3: GLTR-2 verdict on AGN (bottom, Monok) vs. human-authored news (top, Fox News).

wrote it" (Nilsson, 2019). Even automation pioneer AP has nothing on AGN, its use or attribution, in the 2019 edition of its iconic AP Stylebook, just a chapter on (human-authored) data journalism.

The industry still assumes "no one wants AI to fully replace humans in the newsroom, and no one is actively working towards that end" (Chiusi and Beckett, 2019), but start-ups like Monok and Radar show that full AGN is entering general news production. Developing tools for detecting AGN, addressing data biases, and contextualising reported facts can mitigate against some of the adherent risks. More importantly, journalism needs to develop guidance and best practice to ensure the transparent, ethical and responsible use of AGN, if further erosion of concepts like truth, authorship, and responsibility for content is to be avoided.

References

- F. Chiusi and C. Beckett. 2019. [The journalism AI global survey: What we've learned so far](#).
- S. Gehrmann, H. Strobelt, and A. M. Rush. 2019. GLTR: Statistical detection and visualization of generated text. *ArXiv*, abs/1906.04043.
- P. Nilsson. 2019. [Robo-reporter writes front-page news](#). *Financial Times*, 4 June 2019.
- A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8).
- R. Zellers, A. Holtzman, H. Rashkin, Y. Bisk, A. Farhadi, F. Roesner, and Y. Choi. 2019. Defending against neural fake news. *ArXiv*, abs/1905.12616.