
The Rumour Mill: Making the Spread of Misinformation Explicit and Tangible

Nanna Inie

Lix Technologies
IT University of Copenhagen
Copenhagen, Denmark
nanna@lix.com

Jeanette Falk Olesen

Aarhus University
Aarhus, Denmark
jfo@cc.au.dk

Leon Derczynski

IT University of Copenhagen
Copenhagen, Denmark
ld@itu.dk

Abstract

Spread of misinformation presents a technological and social threat to society. With the advance of AI-based language models, automatically generated texts have become difficult to identify and easy to create at scale. We present the “Rumour Mill”, a playful art piece, designed as a commentary on the spread of rumours and automatically-generated misinformation. The mill is a tabletop interactive machine, which invites a user to experience the process of creating believable text by interacting with different tangible controls on the mill. The user manipulates visible parameters to adjust the genre and type of an automatically generated text rumour. The Rumour Mill is a physical demonstration of the state of NLP technology and its ability to generate and manipulate natural language text, and of the act of starting and spreading rumours.

Author Keywords

Misinformation; NLP; rumour spread; fake news

CCS Concepts

•Human-centered computing → Human computer interaction (HCI); •Computing methodologies → Natural language generation;

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Copyright held by the owner/author(s).
CHI'20, April 25–30, 2020, Honolulu, HI, USA
ACM 978-1-4503-6819-3/20/04.
<https://doi.org/10.1145/3334480.XXXXXXX>



Figure 1: The Rumour Mill makes the process of creating rumours by means of NLP technology explicit for users by letting them interact with tangible controls that alters parameters for the AI-generated text.

Introduction and contribution

Identifying and verifying misinformation, fake news and rumours are complex processes, among other things because AI-generated¹ texts can be so believable that they can be indistinguishable from human-written text. This presents profound social issues, as the spread of misinformation, fake news, and rumours can bring potentially serious consequences. With the Rumour Mill we make the rumour process explicit by inviting users to manipulate tangible controls, which alter parameters of an AI-generated rumour. In this way, the Mill is an opportunity to enlighten users, as they get to see for themselves what is possible in creating believable rumours by means of AI technology. The prototype makes recent advances in AI and text generation visible and accessible to a broad audience, while also making the act of starting and sharing a rumour overt and tangible.

Background

The creation and spread of intentionally false and misleading claims undermines social processes of sharing information and of making collective decisions. After society shifted to the internet as its main news source, it became much easier to become a news publisher. The subsequent proliferation of information sources has made it easy for inauthentic information to subvert, distract from, and undermine mainstream, authoritative news. This brings risks. Misinformation has played a role in recent political processes; the 2016 US Election and the Brexit vote both showed signs of interference and misinformation [3, 8].

People spread misinformation and rumours for a variety of reasons. Some share false stories in order to gain attention; others in order to blend in with a group sharing stories

¹Strictly speaking, the recent slew of text generation models are not AI because they are not intelligent. Rather, we use this term to situate our methods within the AI research domain.

(particularly in the wake of disasters [1, 6]); and malicious actors will also seed and propagate false narratives, to manipulate opinion or simply to sow dissent and wreak havoc. However, this is always a covert process.

A further challenge presented by the diversification of where people gather information from is that the origin of a claim becomes less clear, and thus so does its veracity. To address this, many fact-checking agencies have appeared around the world to organise against and monitor false claims.² However the work here is tireless; there are many false claims around, and the cost of creating a new one is very low. In contrast, the effort involved in verifying a claim or rumour is high, involving expert work and having few technological tools to help [9].

Compounding this, automatic tools for generating coherent and believable text have recently become much more proficient [7]. Some are rated by humans as better at producing believable, quality text than human authors [11]. The risk this technology presents has been considered so great that the release of some models is embargoed to give research time to catch up with identifying such text [7]. The extra load presented to fact-checkers by this renewable, tireless source is potentially huge.

By making the rumour seeding process publicly visible and explicit, this no longer is a covert process that is conceived of by the rumour's spreader, but instead a public process; and by making the rumour a tangible object labelled with its source, the spread of the rumour and identification of its nature become explicit. At the same time, people interacting with these rumours should know that they are automatically generated, educating them about what is possible today and giving the exposure needed for people to become

²E.g. the IFCN, <https://www.poynter.org/ifcn/>



Figure 2: The interaction panel of the mill displaying the input parameters.

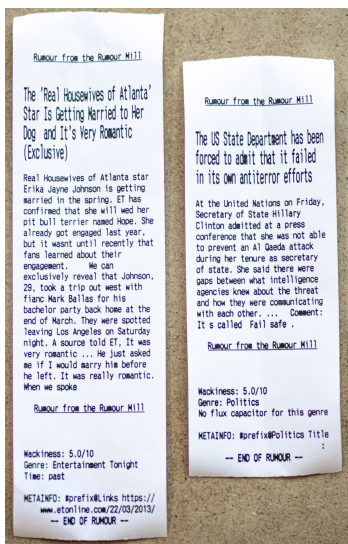


Figure 3: Example rumours printed by the Rumour Mill.

better at spotting and querying automatically-generated or other suspicious claims on their own.

Demo set-up

A wooden and transparent Plexiglas box contains a computer and printer and offers a number of input controls. The computer manages rumour generation and delivers rumours on the printer. The input controls allow users to specify or tune the rumour they would like to create, through different modalities (i.e. a potentiometer, a 12-step switch and a toggle switch). The rotary lever is used to initiate rumour generation, explicitly evoking the metaphor of milling, cf. a rumour mill, co-opting the idiom for a “process in which a group or network of persons originate or promulgate gossip and other unsubstantiated claims.”³ The rumour is then printed on a thermal printer, together with the input settings, and can be thrown away, brought home as a souvenir, or shared on a bulletin board next to the machine.

The mill is made to be large enough that it is awkward to operate covertly. This brings rumour production into an act that is not only tangible for the instigator, but also visible for those nearby. Printed output is marked clearly as being a rumour, both making it easy to source and thus refute, as well as ensuring we are not creating potentially harmful printouts – an important ethical point, and best-practice in reporting of misinformation [4].

Power is supplied externally, and rumour generation occurs remotely, as the hardware required to generate this amount of text in real-time is large, noisy, difficult to transport, and warm (therefore not so welcome in the mill itself). However, the mill can store a cache of rumours locally to allow it to operate through temporary network drops.

³From https://en.wiktionary.org/wiki/rumour_mill

Technology

Rumours produced via the Rumour Mill are generated dependent on three user-specified parameters: 1) Rumour wackiness: The user specifies how wacky or conventional the story should be; 2) Rumour genre: The user specifies a genre of the rumour; and 3) “When”: The user specifies if the rumour should be from the past, the present, or the future. These buttons are placed on a panel as the main interface of the mill (Figure 2).

These parameters are converted to control codes and hyperparameters for a CTRL language model [5]. The “wackiness” corresponds to the “temperature” of the language model, where a low-wackiness story will make low-risk word choices that more closely resemble the training data of the CTRL language model, and a high-wackiness story will have more unusual words and thus express less-likely terms, actions, entities and phrases. The genre is described through control codes. The selected available genres are: politics; conspiracy theory; science news; CNN business; Entertainment Tonight; Daily Mail health; Fox Sports; Independent world news; celebrity gossip; CHI tweets; Russia Today; and a Random setting. The “When” button sets the time period from which a story comes and is determined by including a date in CTRL’s Links control code: Contemporary stories come from within the twelve months up to today, past stories from the decade before that period, and future ones from twelve months from tomorrow.

The rumour text generation itself is a two-step process. When a rumour is requested by “milling” (turning the lever of the mill) and after a user has manipulated the parameters, firstly a headline is requested from a GPT-2 [7] instance. GPT-2 is a high-quality language model that can be fine-tuned to produce a particular kind of text. We tuned it to produce claims and headlines, based on titles from

Snopes, Politifact, and Emergent [10, 2]. This fine-tuning ensures that headline-style text is generated. Secondly, the headline is used as a seed for another language model, CTRL [5]. The CTRL model gives a number of control codes that will condition the generated text. In this case, the text is tuned to fit the user-input parameters and to continue on from the headline. CTRL will thus generate a fluent story that matches the requested style, based on the claim from GPT-2. The output from a milling is one printed rumour, including headline and news blurb.

Conclusion

Automatically generated text can be used for creating and spreading misinformation, fake news and rumours which can be extremely difficult to identify and verify. The Rumour Mill is a playful object of commentary on the process of artificially generating and manipulating believable rumours. In interacting with The Rumour Mill, the piece provides an overt experience of what is possible with AI-generated texts.

Acknowledgements

Verif-AI project (Independent Danish Research Foundation); MultiStance project (ITU Copenhagen); industrial researcher fellowship, CIBIS project (Innovation Fund Denmark).

REFERENCES

- [1] Ahmer Arif, Kelley Shanahan, Fang-Ju Chou, Yoanna Dosouto, Kate Starbird, and Emma S Spiro. 2016. How information snowballs: Exploring the role of exposure in online rumor propagation. In *Proc. CSCW*.
- [2] William Ferreira and Andreas Vlachos. 2016. Emergent: a novel data-set for stance classification. In *Proc. NAACL*. 1163–1168.
- [3] Philip Howard and Bence Kollanyi. 2016. Bots, #StrongerIn, and #Brexit: computational propaganda during the UK-EU referendum. *SSRN 2798311* (2016).
- [4] American Press Institute. 2014. Fact checking: A studio workshop. (2014).
- [5] Nitish Shirish Keskar, Bryan McCann, Lav R Varshney, Caiming Xiong, and Richard Socher. 2019. CTRL: A conditional transformer language model for controllable generation. *arXiv preprint arXiv:1909.05858* (2019).
- [6] Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. 2010. Twitter under crisis: Can we trust what we RT?. In *Proceedings of the first workshop on Social Media Analytics*. ACM, 71–79.
- [7] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language Models are Unsupervised Multitask Learners. (2019).
- [8] Chengcheng Shao, Pik-Mai Hui, Lei Wang, Xinwen Jiang, Alessandro Flammini, Filippo Menczer, and Giovanni Luca Ciampaglia. 2018. Anatomy of an online misinformation network. *PLoS one* 13, 4 (2018).
- [9] Peter Tolmie, Rob Procter, David William Randall, Mark Rouncefield, Christian Burger, Geraldine Wong Sak Hoi, Arkaitz Zubiaga, and Maria Liakata. 2017. Supporting the use of user generated content in journalistic practice. In *Proc. CHI*. ACM, 3632–3644.
- [10] William Yang Wang. 2017. “Liar, Liar Pants on Fire”: A New Benchmark Dataset for Fake News Detection. In *Proc. ACL*. 422–426.
- [11] Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. 2019. Defending Against Neural Fake News. *arXiv preprint arXiv:1905.12616* (2019).