WORDS AS DATA: THE VULNERABILITY OF LANGUAGE IN AN AGE OF DIGITAL CAPITALISM

The security of the data that circulates the internet is dependent on much more than cryptographic key exchange. Data can represent all manner of information that might threaten personal and national securities and safety, and the very nature of the ways that information is transmitted and processed through the Web, they lose their original context. Just like any other type of data, linguistic data becomes vulnerable to manipulation and monetisation.

In this respect, words-as-data become detached from their original function as a means of human communication, and instead become vessels for the flow of advertising and cultural capital around the online and offline world. This has significant consequences.

We all need to communicate, access information and keep up with the modern marketplace, but in today’s digitally networked society, the words we enter into Web-based platforms such as search engines and social media have themselves been turned into valuable pieces of data. And when words are digitised for transmission and processing through the Web, they lose their original context. Just like any other type of data, linguistic data becomes vulnerable to manipulation and monetisation.

The computational manner in which linguistic data is processed is responsible for the sometimes amusing, but also sometimes dangerously stereotypical and controversial auto-predictions that appear when you start typing in the Google search bar. For example, the search term ‘doctor’ might be more often associated with someone like ‘doctor’, ‘boss’ or ‘CEO’, and this will be reflected in search results and auto-completions. It is also the reason why online translation services like Google Translate are often so bad. Google can at any time also interfere, censuring certain keywords so that they won’t be included in the construction of search results. This might be for political, commercial, legal or ethical reasons. Google is not just a neutral and democratic gatekeeper of the world’s information, and it is crucially important not to treat what comes out of the search engine as unmediated truth.

The way digitised language is structured is also dependant on the monetary value of words in the online advertising industry. Google is one of the main players in this marketplace, and its commodification and exploitation of language has been described as a form of linguistic capitalism. Google has around 87% of the UK’s search searches in the UK and its advertising platforms AdWords and AdSense have an ever increasingly significant impact on how all kinds of information circulates on the Web.

AdWords is the system by which advertisers bid for keywords and phrases in order to secure the top spots on Google’s search engine results page. Each time somebody searches for a word on Google, a mini-auction takes place, and the advertiser with the highest bid for a particular word at that time wins, and as long as their advert is considered worthy by the algorithmic ranking system, their advert will appear at the top of the search results page. So much of our online experience is now defined by the top search results, and at the top of these results is often a paid advert.

The process of curation and control here has the frightening power over the generation and dissemination of information. As a result, we need to be asking what narratives are we creating when our online discourse is optimised for the spread of capital rather than the narrative communication? What does it mean that every time we make a search engine is influenced by often opaque algorithmic ‘market forces’, or that YouTube videos aimed at children contain sexual or violent material to encourage more views and therefore more advertising revenue? As we have seen in the revelations about Cambridge Analytics, the spread of fake news through digital advertising is perhaps the tip of the iceberg.

The systemic manipulation and monetisation of digitised language is a threat to the security and stability of modern society. The very words we use to communicate, learn, debate, and critique have become compromised by opaque algorithmic organisation and optimisation, and the market-driven profits of private companies such as Google. We might therefore ask ourselves, just how resilient and secure is language in the digital age? Indeed, how can we even talk about security when we cannot talk securely?

Pip Thornton is a Post-Doctoral Research Associate in Creative Informatics at the University of Edinburgh. The material in this essay is based on her recently published article ‘A Critique of Linguistic Capitalism: Provocation/Intervention’ (2019) and ‘Geographies of Context: Language and Structure in a Digital Age’ (2017), and on her research blog www.linguisticgeographies.com