Reading simCataloguer

A worksheet produced in August 2021 as part of the project 'Legacies of Catalogue Descriptions and Curatorial Voice: Opportunities for Digital Scholarship' (AHRC, Project Reference AH/T013036/1)

A simCataloguer is created by taking text written by a cataloguer or group of cataloguers and using that text to inflect a generalisable language model. The outputs of this process are catalogue descriptions based on the input text, or – as we like to call them – imaginary descriptions of imaginary objects.

The *Legacies* team did this in the case of simGeorge, which uses OpenAI's GPT-2 release (a Recurrent Neural Network – or RNN – trained on 8-million upvoted onward links from the news aggregator site Reddit) inflected with text written by the cataloguer Mary Dorothy George between 1930 and 1954 for the multi-volume *Catalogue of Political and Personal Satires Preserved in the Department of Prints and Drawings in the British Museum*.

The process of making simGeorge is described in James's blogpost. In this worksheet, we present you with some example outputs 'written by' simGeorge and pose a series of questions for you to consider. To get the most out of the worksheet, we encourage you to work in small groups. We anticipate that completing the worksheet will take no longer than one hour.

CW: Dorothy George worked as a cataloguer between 1930 and 1954. Her descriptions are structurally raced, sexed, classed, and gendered, in addition to including offensive and insensitive terms. The amplification of these historically-specific choices by simGeorge is an ethical challenge of which we are aware, and is the reason why we do not share the outputs of simGeorge as a dataset – see 'The problem with simGeorge'. For the purposes of this exercise we have redacted a racial epithet in one of these outputs.

Note: we use *TRANSCRIBED* and *BRACKETED* as placeholders for different types of text in the description that were not written by George. simGeorge learnt the format and placement of these, and incorporated them into their outputs.

1. Can you imagine this "print" in your mind's eye?

> *Wellington, wearing the cocked hat and star, marches beside his horse. Behind *BRACKETED* are the houses of the *TRANSCRIBED* . Sheridan *BRACKETED*, with a black patch over his eye, marches beside Wellington, looking down at Blomfield. Behind stands a naval officer, probably the Duke of Richmond, and on the extreme left walks George Germain *BRACKETED* holding a letter addressed *TRANSCRIBED*. He says: *TRANSCRIBED*.*

2. What normative assumptions underpin these two descriptions? (e.g. what can we assume is unsaid about the women?)

> *A foppish farmer sits in a well-furnished room with a quizzing-table. Beside him stands his wife, turning her head in profile to the right towards their two children, who eagerly take the bread from the table. One wears a swathed neck-cloth, the other a flowered double-breasted state. The farmer says: *TRANSCRIBED*. A boy with a hoop, a toy horse in his cap, and a bunch of big loaves and fishes hangs from a stool. The table is covered with stoppered bottles. Through a doorway *BRACKETED* is seen a crowd of elderly citizens, most of whom are in the middle of a hall watching the family.*

> *A fat elderly woman sits in an armchair reading a short speech while an elderly man *BRACKETED* stands beside her holding a long poker. She puts her head on her hand and looks down in amusement. Her dress is old-fashioned.*

3. How do we balance a desire to create open catalogue data with the risks of recurrent neural networks projecting these kinds of racial distortions into the future?

> *John Bull, in profile to the right, stands, with arms extended, turning his head in profile to address a *TRANSCRIBED*, who stands on his right, glaring angrily at him. He says: *TRANSCRIBED*. His left hand is on his hip, and he says: *TRANSCRIBED*. Behind him an elderly n***o, dressed as a coachman, stands full face, saying with a sly smile: *TRANSCRIBED*.*