**Supplementary figure 1: illustration of the truncation/capitalisation issue (right truncation)**

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Example generated 18 October 2021. First shot, Cost is capitalised and in the second shot, it is not.

**Supplementary figure 2: illustration of the truncation/capitalisation issue (left truncation)**

**Graphical user interface

Description automatically generated with medium confidence**

Graphical user interface, text, application

Description automatically generated

Example generated 21 October 2021. First shot, Diabetes is capitalised and in the second shot, it is not.

**Supplementary figure 3: illustration of the straight-sided quotation marks issues**

Graphical user interface, application

Description automatically generated

Graphical user interface, application, website

Description automatically generated

Example generated 15 October 2021. In the first shot, brackets are curly and in the second they are not.