

Additional file 6

Table S6: Performance in the *chemical-fungus relation determination via entity generation task*

Model	No of Prompts	Tested prompts	Prompts with fungus name in the answer	Prompts with <i>Aspergillus</i> name in the answer	Prompts with correct <i>Aspergillus</i> name in the answer	Prompts with <i>Aspergillus</i> name in the answer/ Prompts with fungus name in the answer (%)	Prompts with correct <i>Aspergillus</i> name in the answer/ Prompts with fungus name in the answer (%)	Prompts with correct <i>Aspergillus</i> name in the answer/ Prompts with <i>Aspergillus</i> name in the answer (%)
Galactica	1	10	8	7	5	87.5	62.5	71.4
BioGPT	11	110	57	45	13	78.9	22.8	28.9
GPT-2	11	110	0	0	0	0.0	0.0	0.0
GPT-neo	11	110	28	11	4	39.3	14.3	36.4
Bloom	11	110	12	3	2	25.0	16.7	66.7
GPT-3	11	110	56	21	14	37.5	25.0	66.7
BioGPT-large	11	110	74	20	13	27.0	17.6	65.0
ChatGPT	11	110	101	21	12	20.8	11.9	57.1
GPT-4	11	110	77	21	15	27.3	19.5	71.4

Evaluation according to the framework (see Fig. 3). 10 chemicals, 11 prompts (P8-P18) (for Galactica only P19 prompt), 110 outputs per model in total (for Galactica 10 outputs). Text generation task for Occurrence of the most common fungus in chemical compound-fungus relation in tested prompts.