A REBUTTAL

A.1 Re: Supervisor - Weaknesses: Difference in Inference Times

Experiment	Model	Average Response Length	Average Inference Time per Document (ms)
GEITje	IC - FewShot	17.0	41.0
	IC - ZeroShot	16.1	33.0
	Fine-Tuning	9.5	24.0
Llama	IC - FewShot	160.8	228.0
	IC - ZeroShot	90.2	120.0
	Fine-Tuning	10.5	20.0
Mistral	IC - FewShot	36.0	68.0
	IC - ZeroShot	37.7	60.0
	Fine-Tuning	9.5	22.0

Table 7: Comparison of Average Response Length to Inference Time Across Different Models and Experiments

A.2 QUESTION 2

Table 8: Token Distribution using Llama Tokenizer

	Tokens
Count	20818
Mean	4340
Std	15456
Min	74
25%	612
50%	1031
75%	2378
Max	618067

Table 9: Token distribution per class using Llama's Tokenizer

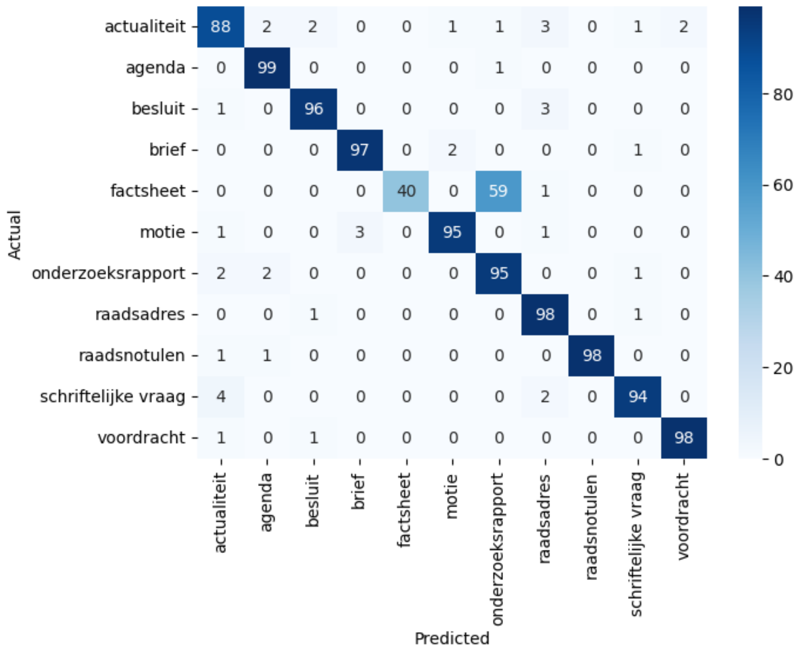
label	count	mean	std	min	25%	50%	75%	max
Actualiteit	800	1630	8662	165	513	733	1305	234499
Agenda	2544	2000	10122	74	517	1054	1918	314952
Besluit	625	2183	3918	227	387	979	2619	56288
Brief	1056	3138	3082	312	1363	2297	3842	56355
Factsheet	**214**	**12337**	**26342**	**261**	**2327**	**5927**	**11868**	**230752**
Motie	7639	901	1680	239	514	634	835	75020
Onderzoeksrapport	**1174**	**31678**	**40611**	**683**	**12640**	**22654**	**38088**	**618067**
Raadsadres	1621	2099	3076	108	737	1261	2256	35101
Raadsnotulen	**231**	**68093**	**22624**	**4606**	**54820**	**71582**	**84302**	**109188**
Schriftelijke Vraag	2932	3614	12015	497	1802	2474	3438	365774
Voordracht	1982	1433	1007	376	918	1133	1535	11156

A.3 QUESTION 3

Table 10: Classification Report of GEITje's performance using the Few-Shot Prompt

	precision	**recall**	**f1-score**	**support**
PredictionError	0.00	0.00	0.00	0
actualiteit	0.93	0.71	0.81	100
agenda	0.95	0.95	0.95	100
besluit	0.94	0.72	0.81	100
brief	0.88	0.82	0.85	100
factsheet	0.90	0.37	0.52	100
motie	0.95	0.84	0.89	100
onderzoeksrapport	0.61	0.71	0.66	100
raadsadres	0.84	0.78	0.81	100
raadsnotulen	0.84	0.96	0.90	100
schriftelijke vraag	0.98	0.91	0.94	100
voordracht	0.60	0.99	0.75	100
accuracy			0.80	1100
macro avg	0.79	0.73	0.74	1100
weighted avg	0.86	0.80	0.81	1100

A.4 QUESTION 4



	actualiteit	agenda	besluit	brief	factsheet	motie	onderzoeksrapport	raadsadres	raadsnotulen	schriftelijke vraag	voordracht
actualiteit	88	2	2	0	0	1	1	3	0	1	2
agenda	0	99	0	0	0	0	1	0	0	0	0
besluit	1	0	96	0	0	0	0	3	0	0	0
brief	0	0	0	97	0	2	0	0	0	1	0
factsheet	0	0	0	0	40	0	59	1	0	0	0
motie	1	0	0	3	0	95	0	1	0	0	0
onderzoeksrapport	2	2	0	0	0	0	95	0	0	1	0
raadsadres	0	0	1	0	0	0	0	98	0	1	0
raadsnotulen	1	1	0	0	0	0	0	0	98	0	0
schriftelijke vraag	4	0	0	0	0	0	0	2	0	94	0
voordracht	1	0	1	0	0	0	0	0	0	0	98

Figure 5: Confusion Matrix of Fine-tuned Mistral. Trained for three epochs; first 200 tokens as input.

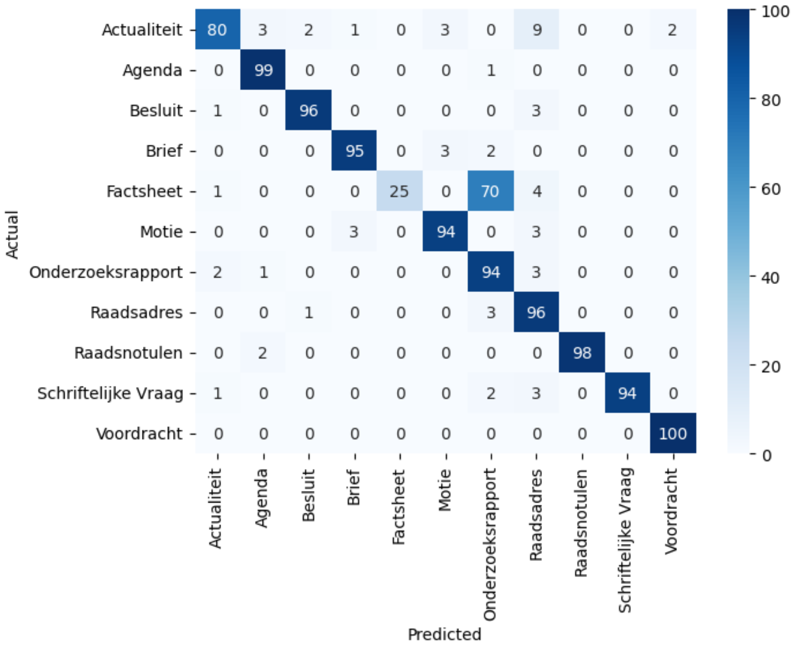


Figure 6: Confusion Matrix of Linear SVM Using First 200 Tokens as Input

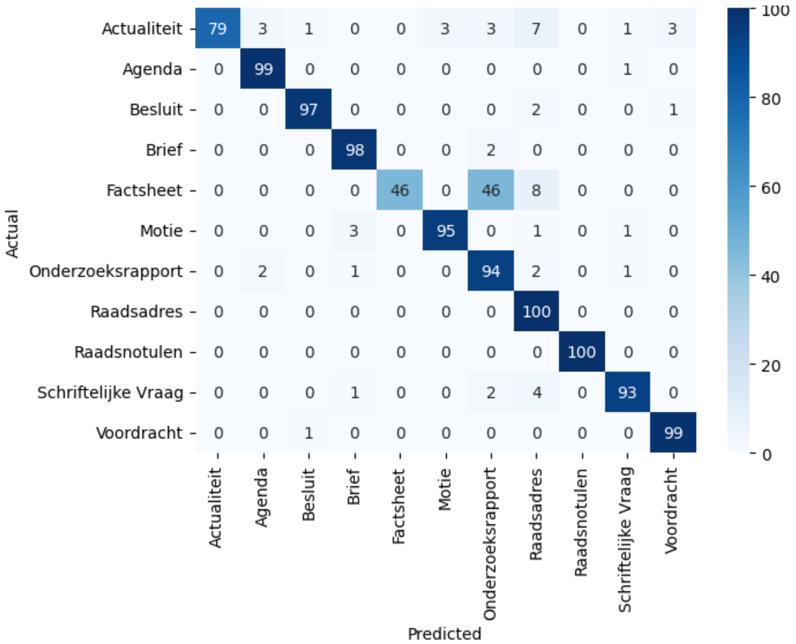


Figure 7: Confusion Matrix of Linear SVM Using Full Text as Input.

Table 11: Classification Report of Fine-Tuned Mistral's Performance

	precision	recall	f1-score	support
actualiteit	0.90	0.88	0.89	100
agenda	0.95	0.99	0.97	100
besluit	0.96	0.96	0.96	100
brief	0.97	0.97	0.97	100
factsheet	1.00	0.40	0.57	100
motie	0.97	0.95	0.96	100
onderzoeksrapport	0.61	0.95	0.74	100
raadsadres	0.91	0.98	0.94	100
raadsnotulen	1.00	0.98	0.99	100
schriftelijke vraag	0.96	0.94	0.95	100
voordracht	0.98	0.98	0.98	100
accuracy			0.91	1100
macro avg	0.93	0.91	0.90	1100
weighted avg	0.93	0.91	0.90	1100

Table 12: Classification Report of Linear SVM with First 200 Tokens as Input

	precision	recall	f1-score	support
actualiteit	0.94	0.80	0.86	100
agenda	0.94	0.99	0.97	100
besluit	0.97	0.96	0.96	100
brief	0.96	0.95	0.95	100
factsheet	1.00	0.25	0.40	100
motie	0.94	0.94	0.94	100
onderzoeksrapport	0.55	0.94	0.69	100
raadsadres	0.79	0.96	0.87	100
raadsnotulen	1.00	0.98	0.99	100
schriftelijke vraag	1.00	0.94	0.97	100
voordracht	0.98	1.00	0.99	100
accuracy			0.88	1100
macro avg	0.92	0.88	0.87	1100
weighted avg	0.92	0.88	0.87	1100

Table 13: Classification Report of Linear SVM with Full Text as Input

	precision	recall	f1-score	support
actualiteit	1.00	0.79	0.88	100
agenda	0.95	0.99	0.97	100
besluit	0.98	0.97	0.97	100
brief	0.95	0.98	0.97	100
factsheet	1.00	0.46	0.63	100
motie	0.97	0.95	0.96	100
onderzoeksrapport	0.64	0.94	0.76	100
raadsadres	0.81	1.00	0.89	100
raadsnotulen	1.00	1.00	1.00	100
schriftelijke vraag	0.96	0.93	0.94	100
voordracht	0.96	0.99	0.98	100
accuracy			0.91	1100
macro avg	0.93	0.91	0.91	1100
weighted avg	0.93	0.91	0.91	1100