

Ontology Population Using LLMs: Which Factors Matter? (Tables and Figures)

List of Tables

1	Zero-Shot prediction results	1
2	1-Shot prediction results	2
3	2-Shot prediction results	3
4	3-Shot prediction results	4
5	4-Shot prediction results	5
6	5-Shot prediction results	6
7	6-Shot prediction results	7
8	7-Shot prediction results	8
9	8-Shot prediction results	9
10	9-Shot prediction results	10
11	10-Shot prediction results	11

List of Figures

1	n-shot mAP@D variation, GPT-4o, generic context	12
2	n-shot mAP@D variation, Llama3-8B, generic context	13
3	n-shot mAP@D variation, GPT-4o, ontology context	14
4	n-shot mAP@D variation, Llama3-8B, ontology context	15
5	n-shot mAP@D variation, GPT-4o, topic context	16
6	n-shot mAP@D variation, Llama3-8B, topic context	17
7	n-shot mAP@D variation, GPT-4o, double context	18
8	n-shot mAP@D variation, Llama3-8B, double context	19
9	Temperature variation Wines ontology, generic context	20
10	Temperature variation CASE ontology, generic context	21
11	Temperature variation Wines ontology, ontology context	22
12	Temperature variation CASE ontology, ontology context	23
13	Temperature variation Wines ontology, topic context	24
14	Temperature variation CASE ontology, topic context	25
15	Temperature variation Wines ontology, double context	26
16	Temperature variation CASE ontology, double context	27

Table 1: Zero-Shot prediction results. Values in **bold** are the highest across all four domain contextualization variations for that ontology and LLM (DeepSeek-R1-Llama-8B abbreviated to DeepSeek-R1*)

Ontology	LLM	Domain Context	mAP@1	mAP@D
Wines	o1-preview	Generic	0.752	0.832
		Ontology	0.826	0.871
		Topic	0.783	0.846
		Ontology and Topic	0.814	0.847
	GPT-4o	Generic	0.764	0.819
		Ontology	0.733	0.785
		Topic	0.708	0.760
		Ontology and Topic	0.745	0.800
	Llama3-8B	Generic	0.329	0.396
		Ontology	0.484	0.571
		Topic	0.367	0.464
		Ontology and Topic	0.528	0.570
	DeepSeek-R1*	Generic	0.460	0.507
		Ontology	0.472	0.542
		Topic	0.416	0.488
		Ontology and Topic	0.398	0.473
CASE	o1-preview	Generic	0.879	0.475
		Ontology	0.878	0.578
		Topic	0.847	0.498
		Ontology and Topic	0.870	0.558
	GPT-4o	Generic	0.763	0.261
		Ontology	0.779	0.316
		Topic	0.725	0.233
		Ontology and Topic	0.641	0.311
	Llama3-8B	Generic	0.603	0.211
		Ontology	0.626	0.270
		Topic	0.695	0.269
		Ontology and Topic	0.641	0.277
	DeepSeek-R1*	Generic	0.550	0.186
		Ontology	0.557	0.196
		Topic	0.367	0.464
		Ontology and Topic	0.527	0.153
Astronomy	o1-preview	Generic	0.971	0.864
		Ontology	0.956	0.870
		Topic	0.971	0.882
		Ontology and Topic	0.971	0.885
	GPT-4o	Generic	0.779	0.670
		Ontology	0.926	0.750
		Topic	0.779	0.605
		Ontology and Topic	0.911	0.729
	Llama3-8B	Generic	0.588	0.416
		Ontology	0.544	0.420
		Topic	0.824	0.549
		Ontology and Topic	0.574	0.452
	DeepSeek-R1*	Generic	0.529	0.330
		Ontology	0.618	0.387
		Topic	0.603	0.378
		Ontology and Topic	0.574	0.313

Table 2: 1-Shot prediction results. Values in **bold** are the highest across all four domain contextualization variations for that ontology and LLM (DeepSeek-R1-Llama-8B abbreviated to DeepSeek-R1*)

Ontology	LLM	Domain Context	mAP@1	mAP@D
Wines	GPT-4o	Generic	0.769	0.843
		Ontology	0.800	0.868
		Topic	0.794	0.853
		Ontology and Topic	0.806	0.866
	Llama3-8B	Generic	0.563	0.628
		Ontology	0.500	0.536
		Topic	0.563	0.594
		Ontology and Topic	0.563	0.594
CASE	GPT-4o	Generic	0.831	0.526
		Ontology	0.823	0.608
		Topic	0.838	0.501
		Ontology and Topic	0.823	0.516
	Llama3-8B	Generic	0.692	0.335
		Ontology	0.700	0.372
		Topic	0.715	0.355
		Ontology and Topic	0.731	0.423
Astronomy	GPT-4o	Generic	0.970	0.794
		Ontology	0.940	0.830
		Topic	0.896	0.785
		Ontology and Topic	0.910	0.804
	Llama3-8B	Generic	0.716	0.567
		Ontology	0.776	0.627
		Topic	0.597	0.539
		Ontology and Topic	0.761	0.637

Table 3: 2-Shot prediction results. Values in **bold** are the highest across all four domain contextualization variations for that ontology and LLM (DeepSeek-R1-Llama-8B abbreviated to DeepSeek-R1*)

Ontology	LLM	Domain Context	mAP@1	mAP@D
Wines	GPT-4o	Generic	0.824	0.869
		Ontology	0.843	0.886
		Topic	0.830	0.855
		Ontology and Topic	0.862	0.886
	Llama3-8B	Generic	0.553	0.584
		Ontology	0.591	0.615
		Topic	0.535	0.597
		Ontology and Topic	0.522	0.603
CASE	GPT-4o	Generic	0.814	0.707
		Ontology	0.837	0.761
		Topic	0.822	0.685
		Ontology and Topic	0.837	0.739
	Llama3-8B	Generic	0.667	0.357
		Ontology	0.667	0.603
		Topic	0.705	0.576
		Ontology and Topic	0.705	0.563
Astronomy	GPT-4o	Generic	0.924	0.771
		Ontology	0.864	0.772
		Topic	0.909	0.783
		Ontology and Topic	0.879	0.760
	Llama3-8B	Generic	0.773	0.601
		Ontology	0.712	0.604
		Topic	0.667	0.578
		Ontology and Topic	0.697	0.585

Table 4: 3-Shot prediction results. Values in **bold** are the highest across all four domain contextualization variations for that ontology and LLM (DeepSeek-R1-Llama-8B abbreviated to DeepSeek-R1*)

Ontology	LLM	Domain Context	mAP@1	mAP@D
Wines	GPT-4o	Generic	0.861	0.884
		Ontology	0.873	0.889
		Topic	0.867	0.887
		Ontology and Topic	0.873	0.878
	Llama3-8B	Generic	0.563	0.613
		Ontology	0.690	0.720
		Topic	0.658	0.707
		Ontology and Topic	0.551	0.605
CASE	GPT-4o	Generic	0.852	0.737
		Ontology	0.836	0.772
		Topic	0.828	0.734
		Ontology and Topic	0.844	0.768
	Llama3-8B	Generic	0.727	0.375
		Ontology	0.711	0.552
		Topic	0.711	0.486
		Ontology and Topic	0.703	0.590
Astronomy	GPT-4o	Generic	0.969	0.827
		Ontology	0.954	0.807
		Topic	0.954	0.840
		Ontology and Topic	0.923	0.827
	Llama3-8B	Generic	0.785	0.636
		Ontology	0.769	0.648
		Topic	0.708	0.594
		Ontology and Topic	0.846	0.623

Table 5: 4-Shot prediction results. Values in **bold** are the highest across all four domain contextualization variations for that ontology and LLM (DeepSeek-R1-Llama-8B abbreviated to DeepSeek-R1*)

Ontology	LLM	Domain Context	mAP@1	mAP@D
Wines	GPT-4o	Generic	0.860	0.887
		Ontology	0.854	0.882
		Topic	0.885	0.894
		Ontology and Topic	0.854	0.885
	Llama3-8B	Generic	0.631	0.674
		Ontology	0.554	0.597
		Topic	0.669	0.728
		Ontology and Topic	0.618	0.638
CASE	GPT-4o	Generic	0.819	0.743
		Ontology	0.819	0.750
		Topic	0.858	0.735
		Ontology and Topic	0.843	0.752
	Llama3-8B	Generic	0.756	0.573
		Ontology	0.732	0.575
		Topic	0.724	0.586
		Ontology and Topic	0.661	0.608
Astronomy	GPT-4o	Generic	0.906	0.820
		Ontology	0.906	0.786
		Topic	0.969	0.825
		Ontology and Topic	0.844	0.769
	Llama3-8B	Generic	0.734	0.613
		Ontology	0.766	0.624
		Topic	0.828	0.671
		Ontology and Topic	0.734	0.581

Table 6: 5-Shot prediction results. Values in **bold** are the highest across all four domain contextualization variations for that ontology and LLM (DeepSeek-R1-Llama-8B abbreviated to DeepSeek-R1*)

Ontology	LLM	Domain Context	mAP@1	mAP@D
Wines	GPT-4o	Generic	0.885	0.907
		Ontology	0.827	0.859
		Topic	0.859	0.880
		Ontology and Topic	0.776	0.828
	Llama3-8B	Generic	0.635	0.664
		Ontology	0.583	0.635
		Topic	0.667	0.707
		Ontology and Topic	0.622	0.662
CASE	GPT-4o	Generic	0.841	0.765
		Ontology	0.849	0.791
		Topic	0.833	0.714
		Ontology and Topic	0.849	0.797
	Llama3-8B	Generic	0.714	0.557
		Ontology	0.667	0.622
		Topic	0.667	0.622
		Ontology and Topic	0.698	0.629
Astronomy	GPT-4o	Generic	0.937	0.817
		Ontology	0.952	0.835
		Topic	0.952	0.788
		Ontology and Topic	0.921	0.821
	Llama3-8B	Generic	0.810	0.617
		Ontology	0.714	0.621
		Topic	0.667	0.604
		Ontology and Topic	0.698	0.619

Table 7: 6-Shot prediction results. Values in **bold** are the highest across all four domain contextualization variations for that ontology and LLM (DeepSeek-R1-Llama-8B abbreviated to DeepSeek-R1*)

Ontology	LLM	Domain Context	mAP@1	mAP@D
Wines	GPT-4o	Generic	0.858	0.883
		Ontology	0.813	0.855
		Topic	0.845	0.881
		Ontology and Topic	0.858	0.893
	Llama3-8B	Generic	0.568	0.628
		Ontology	0.600	0.622
		Topic	0.626	0.675
		Ontology and Topic	0.587	0.644
CASE	GPT-4o	Generic	0.848	0.786
		Ontology	0.840	0.779
		Topic	0.84	0.775
		Ontology and Topic	0.848	0.791
	Llama3-8B	Generic	0.688	0.610
		Ontology	0.760	0.654
		Topic	0.696	0.642
		Ontology and Topic	0.664	0.615
Astronomy	GPT-4o	Generic	0.968	0.824
		Ontology	0.919	0.835
		Topic	0.952	0.809
		Ontology and Topic	0.887	0.813
	Llama3-8B	Generic	0.726	0.593
		Ontology	0.677	0.581
		Topic	0.758	0.591
		Ontology and Topic	0.839	0.692

Table 8: 7-Shot prediction results. Values in **bold** are the highest across all four domain contextualization variations for that ontology and LLM (DeepSeek-R1-Llama-8B abbreviated to DeepSeek-R1*)

Ontology	LLM	Domain Context	mAP@1	mAP@D
Wines	GPT-4o	Generic	0.890	0.900
		Ontology	0.916	0.912
		Topic	0.896	0.907
		Ontology and Topic	0.877	0.904
	Llama3-8B	Generic	0.597	0.646
		Ontology	0.578	0.620
		Topic	0.623	0.660
		Ontology and Topic	0.584	0.617
CASE	GPT-4o	Generic	0.847	0.792
		Ontology	0.863	0.796
		Topic	0.839	0.781
		Ontology and Topic	0.863	0.779
	Llama3-8B	Generic	0.702	0.586
		Ontology	0.645	0.634
		Topic	0.621	0.614
		Ontology and Topic	0.710	0.690
Astronomy	GPT-4o	Generic	0.934	0.811
		Ontology	0.902	0.794
		Topic	0.934	0.836
		Ontology and Topic	0.934	0.822
	Llama3-8B	Generic	0.705	0.596
		Ontology	0.787	0.621
		Topic	0.738	0.659
		Ontology and Topic	0.721	0.615

Table 9: 8-Shot prediction results. Values in **bold** are the highest across all four domain contextualization variations for that ontology and LLM (DeepSeek-R1-Llama-8B abbreviated to DeepSeek-R1*)

Ontology	LLM	Domain Context	mAP@1	mAP@D
Wines	GPT-4o	Generic	0.843	0.874
		Ontology	0.908	0.900
		Topic	0.869	0.897
		Ontology and Topic	0.850	0.887
	Llama3-8B	Generic	0.627	0.654
		Ontology	0.634	0.653
		Topic	0.712	0.736
		Ontology and Topic	0.673	0.691
CASE	GPT-4o	Generic	0.846	0.789
		Ontology	0.854	0.789
		Topic	0.837	0.744
		Ontology and Topic	0.870	0.806
	Llama3-8B	Generic	0.724	0.619
		Ontology	0.699	0.658
		Topic	0.707	0.595
		Ontology and Topic	0.724	0.648
Astronomy	GPT-4o	Generic	0.917	0.784
		Ontology	0.967	0.847
		Topic	0.917	0.782
		Ontology and Topic	0.917	0.810
	Llama3-8B	Generic	0.683	0.589
		Ontology	0.700	0.596
		Topic	0.717	0.614
		Ontology and Topic	0.85	0.638

Table 10: 9-Shot prediction results. Values in **bold** are the highest across all four domain contextualization variations for that ontology and LLM (DeepSeek-R1-Llama-8B abbreviated to DeepSeek-R1*)

Ontology	LLM	Domain Context	mAP@1	mAP@D
Wines	GPT-4o	Generic	0.875	0.900
		Ontology	0.882	0.908
		Topic	0.888	0.909
		Ontology and Topic	0.888	0.902
	Llama3-8B	Generic	0.632	0.672
		Ontology	0.691	0.717
		Topic	0.618	0.667
		Ontology and Topic	0.638	0.699
CASE	GPT-4o	Generic	0.844	0.805
		Ontology	0.852	0.813
		Topic	0.852	0.789
		Ontology and Topic	0.852	0.799
	Llama3-8B	Generic	0.697	0.595
		Ontology	0.770	0.663
		Topic	0.770	0.623
		Ontology and Topic	0.730	0.688
Astronomy	GPT-4o	Generic	0.932	0.810
		Ontology	0.932	0.784
		Topic	0.932	0.818
		Ontology and Topic	0.881	0.824
	Llama3-8B	Generic	0.695	0.600
		Ontology	0.712	0.646
		Topic	0.695	0.649
		Ontology and Topic	0.695	0.630

Table 11: 10-Shot prediction results. Values in **bold** are the highest across all four domain contextualization variations for that ontology and LLM (DeepSeek-R1-Llama-8B abbreviated to DeepSeek-R1*)

Ontology	LLM	Domain Context	mAP@1	mAP@D
Wines	GPT-4o	Generic	0.881	0.886
		Ontology	0.894	0.901
		Topic	0.874	0.894
		Ontology and Topic	0.874	0.900
	Llama3-8B	Generic	0.689	0.692
		Ontology	0.603	0.630
		Topic	0.623	0.655
		Ontology and Topic	0.682	0.708
CASE	GPT-4o	Generic	0.860	0.793
		Ontology	0.851	0.829
		Topic	0.860	0.777
		Ontology and Topic	0.860	0.790
	Llama3-8B	Generic	0.711	0.614
		Ontology	0.785	0.673
		Topic	0.711	0.677
		Ontology and Topic	0.744	0.674
Astronomy	GPT-4o	Generic	0.914	0.799
		Ontology	0.914	0.778
		Topic	0.339	0.316
		Ontology and Topic	0.897	0.812
	Llama3-8B	Generic	0.776	0.667
		Ontology	0.810	0.670
		Topic	0.810	0.627
		Ontology and Topic	0.845	0.691

Figure 1: mAP@D variation with number of examples for GPT-4o using generic domain contextualization

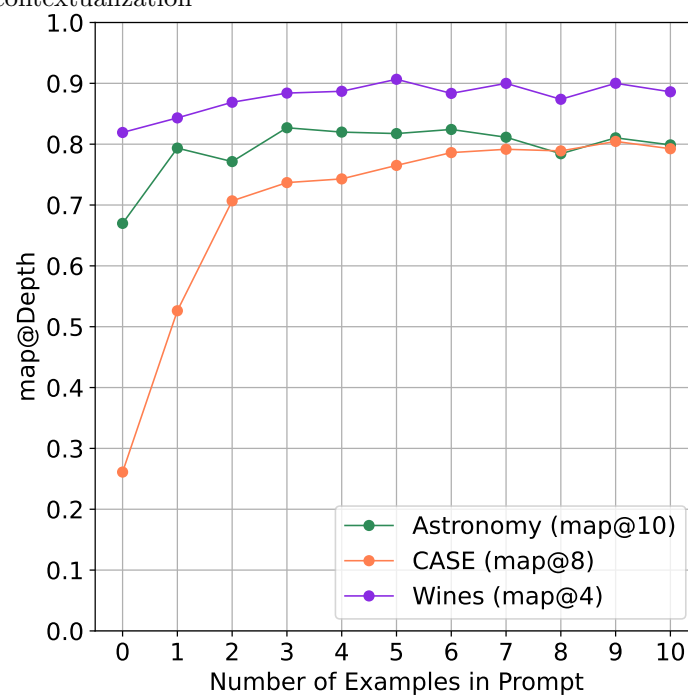


Figure 2: mAP@D variation with number of examples for Llama3-8B using generic domain contextualization

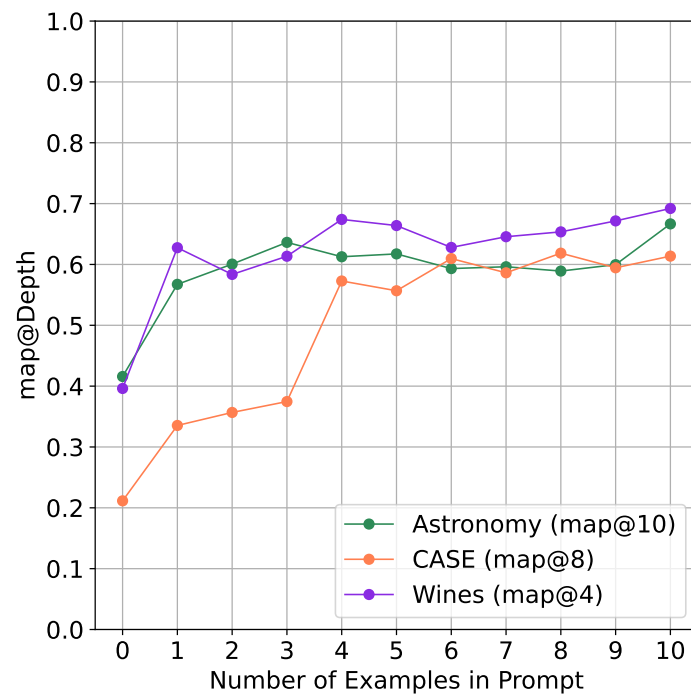


Figure 3: mAP@D variation with number of examples for GPT-4o using ontology domain contextualization

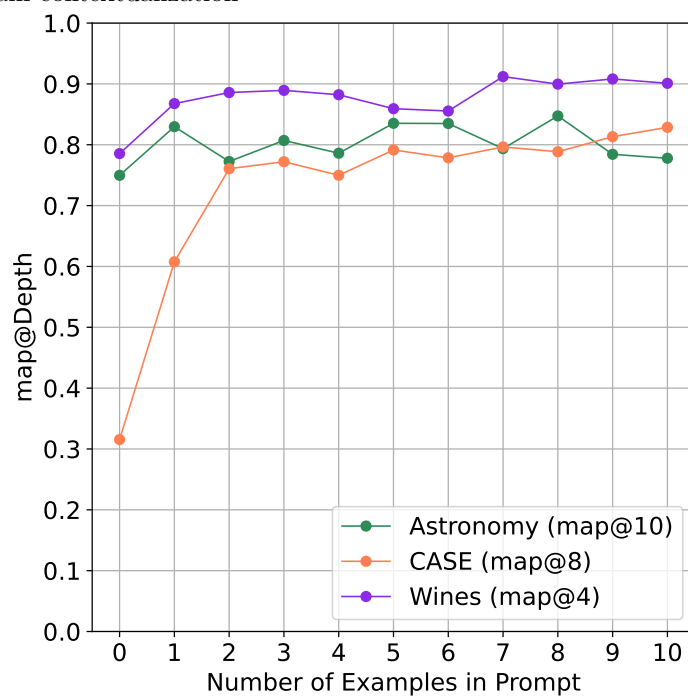


Figure 4: mAP@D variation with number of examples for Llama3-8B using ontology domain contextualization

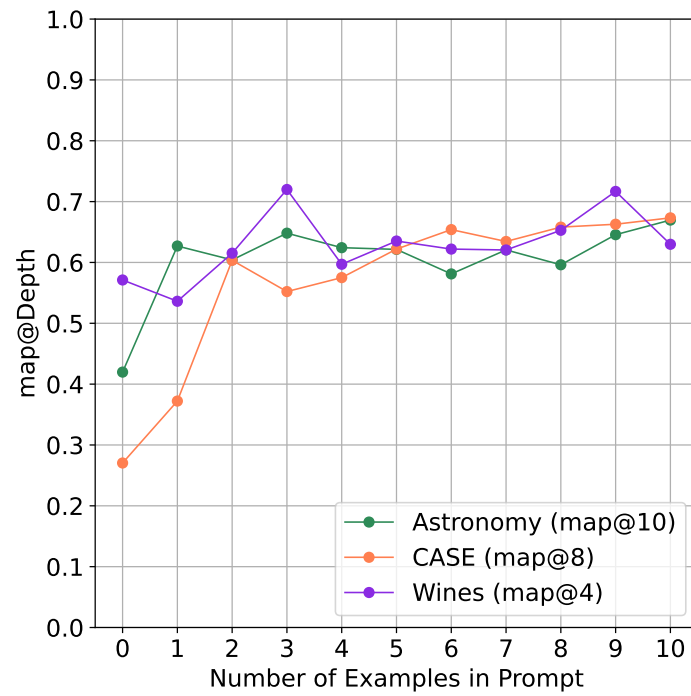


Figure 5: mAP@D variation with number of examples for GPT-4o using topic domain contextualization

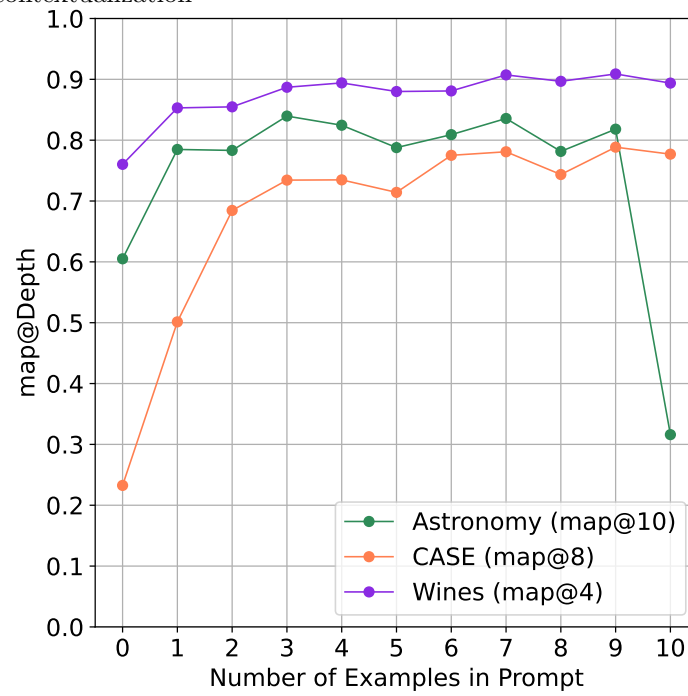


Figure 6: mAP@D variation with number of examples for Llama3-8B using topic domain contextualization

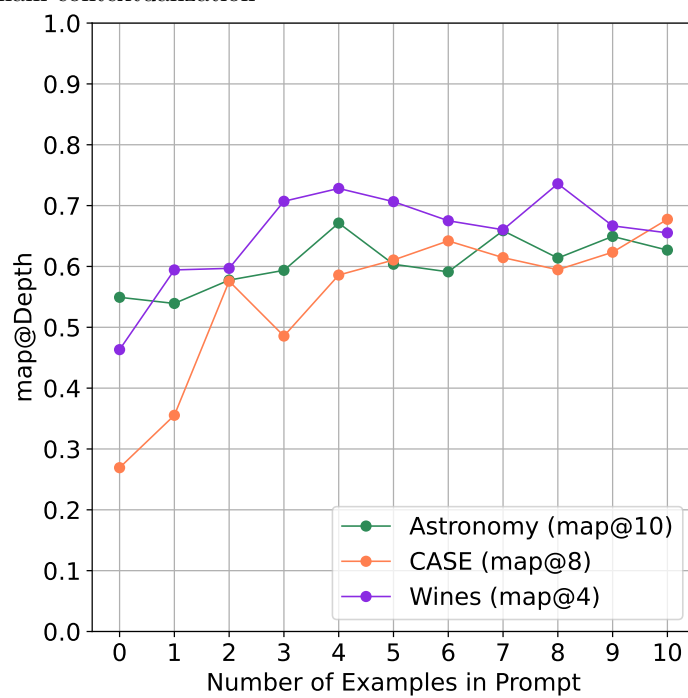


Figure 7: mAP@D variation with number of examples for GPT-4o using ontology and topic domain contextualization

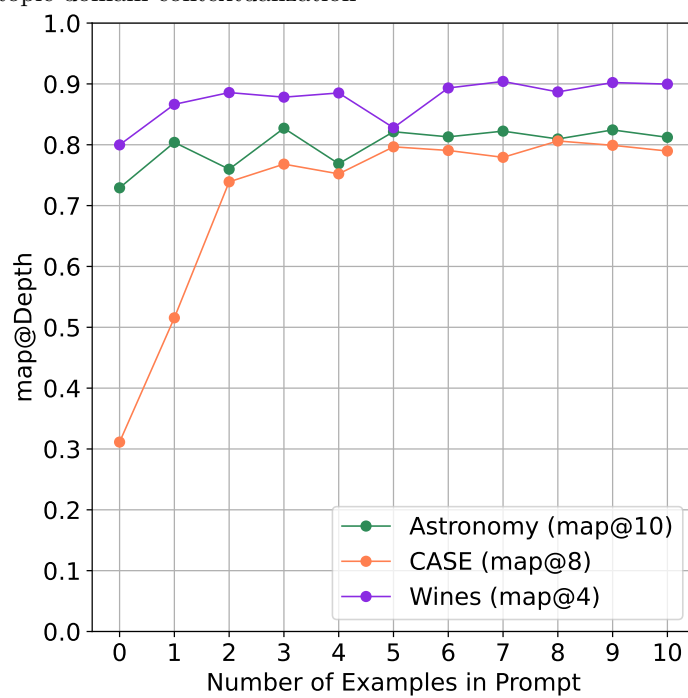


Figure 8: mAP@D variation with number of examples for Llama3-8B using ontology and topic domain contextualization

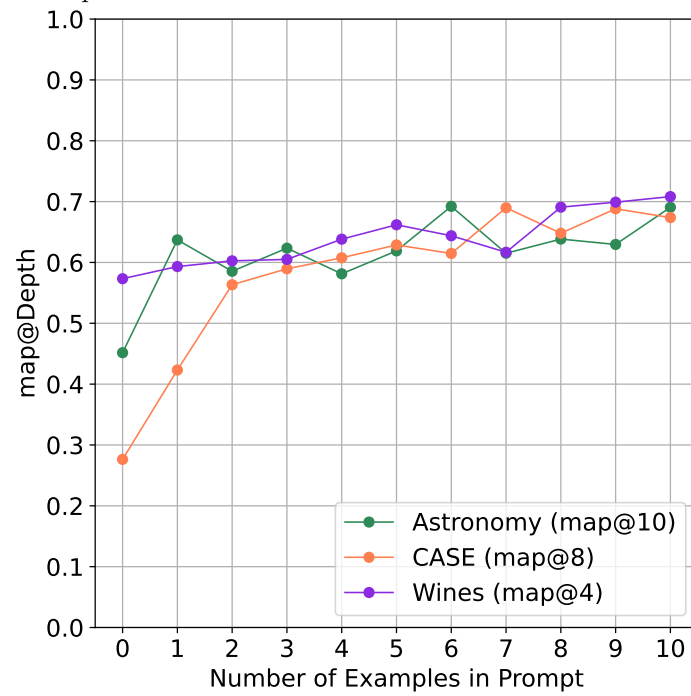


Figure 9: Temperature variation for Wines ontology for GPT-4o and Llama3-8B with 3-shot, generic domain contextualization. Central points are the average value over 10 runs. Lighter shaded regions are approximated densities over the range of values

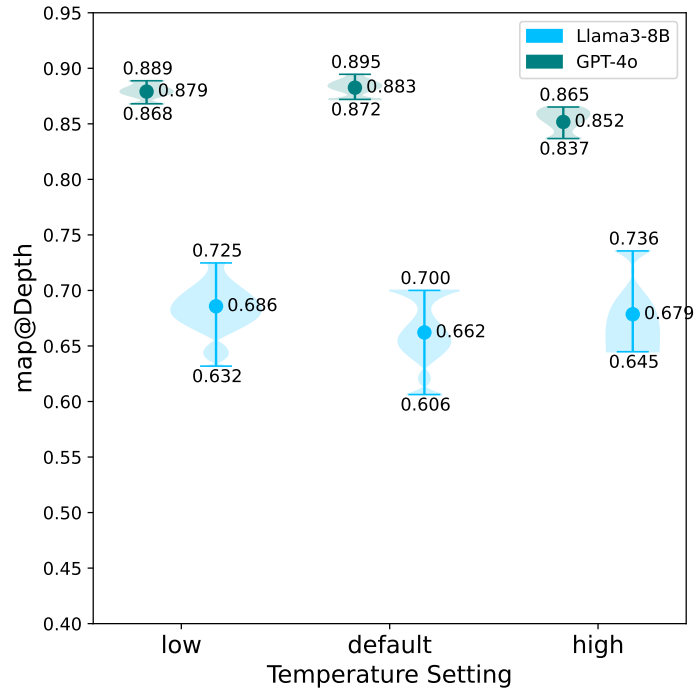


Figure 10: Temperature variation for CASE ontology for GPT-4o and Llama3-8B with 3-shot, generic domain contextualization. Central points are the average value over 10 runs. Lighter shaded regions are approximated densities over the range of values

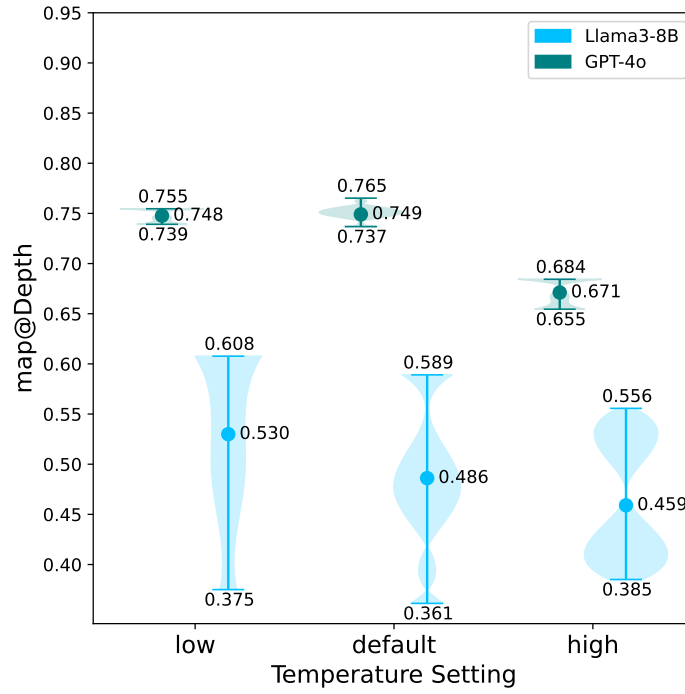


Figure 11: Temperature variation for Wines ontology for GPT-4o and Llama3-8B with 3-shot, ontology domain contextualization. Central points are the average value over 10 runs. Lighter shaded regions are approximated densities over the range of values

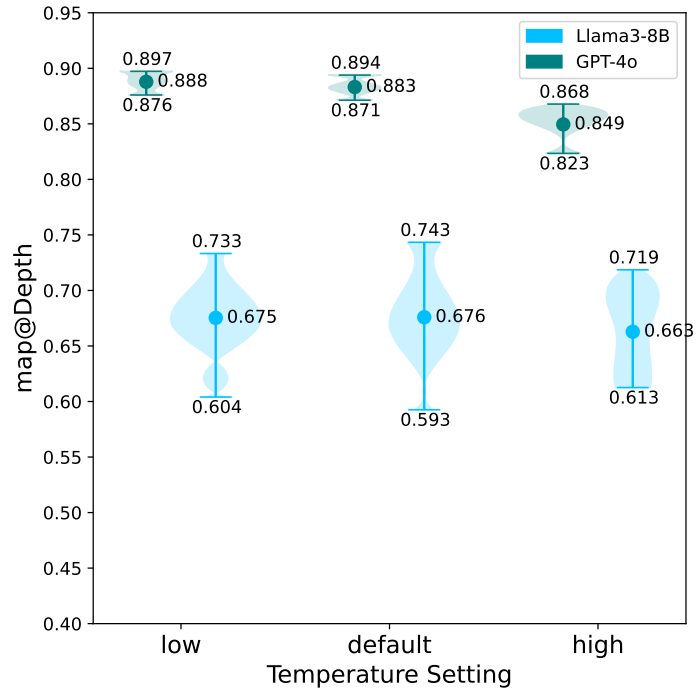


Figure 12: Temperature variation for CASE ontology for GPT-4o and Llama3-8B with 3-shot, ontology domain contextualization. Central points are the average value over 10 runs. Lighter shaded regions are approximated densities over the range of values

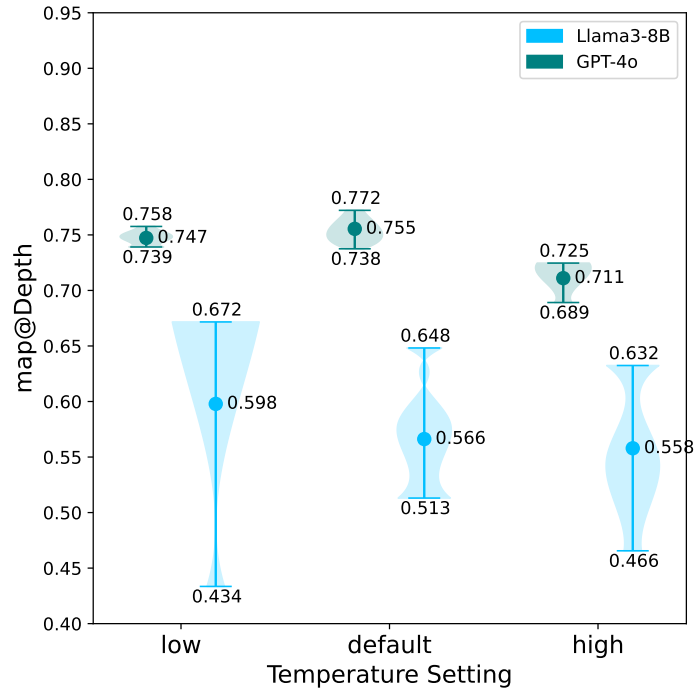


Figure 13: Temperature variation for Wines ontology for GPT-4o and Llama3-8B with 3-shot, topic domain contextualization. Central points are the average value over 10 runs. Lighter shaded regions are approximated densities over the range of values

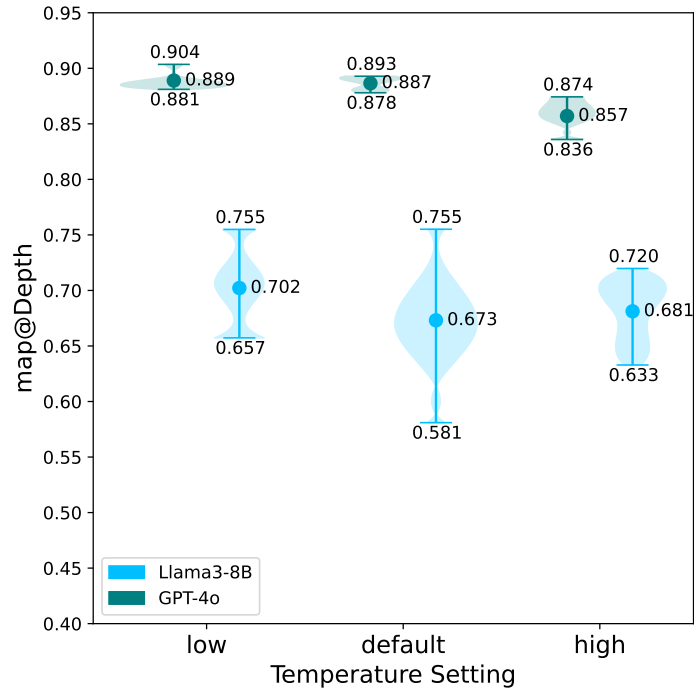


Figure 14: Temperature variation for CASE ontology for GPT-4o and Llama3-8B with 3-shot, topic domain contextualization. Central points are the average value over 10 runs. Lighter shaded regions are approximated densities over the range of values

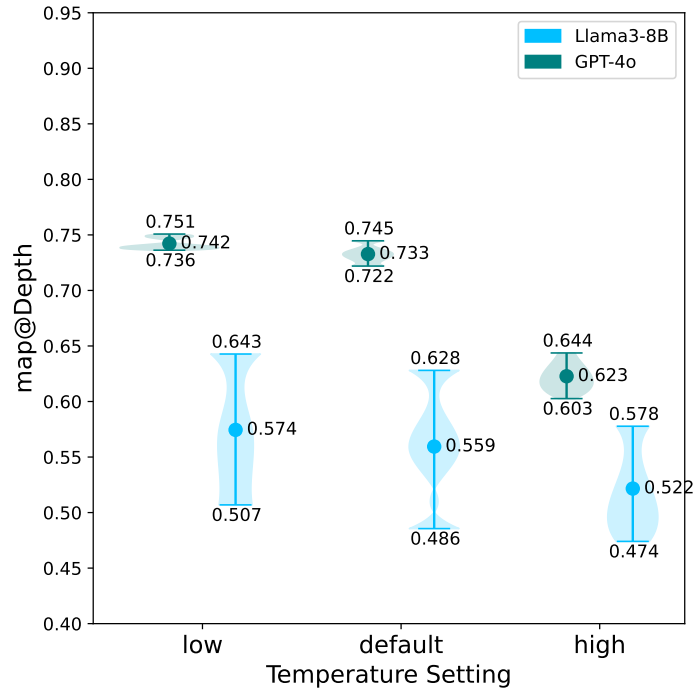


Figure 15: Temperature variation for Wines ontology for GPT-4o and Llama3-8B with 3-shot, ontology and topic domain contextualization. Central points are the average value over 10 runs. Lighter shaded regions are approximated densities over the range of values

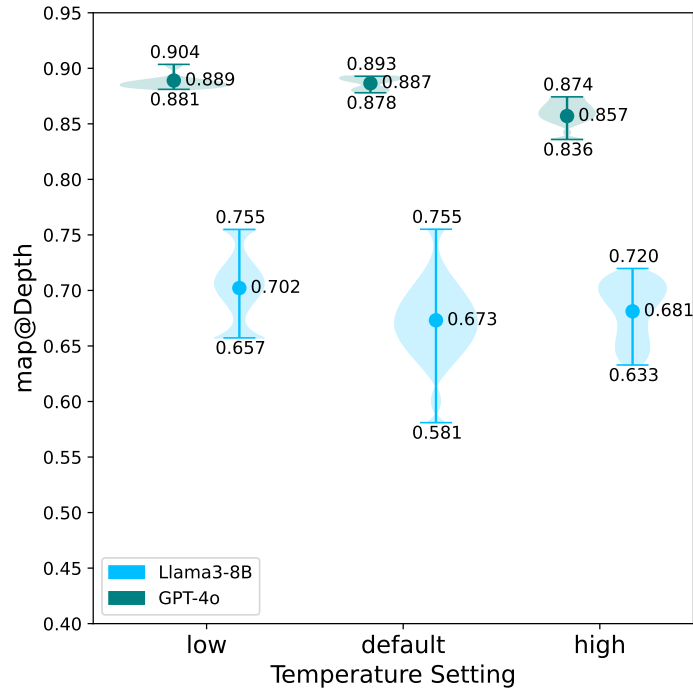


Figure 16: Temperature variation for CASE ontology for GPT-4o and Llama3-8B with 3-shot, ontology and topic domain contextualization. Central points are the average value over 10 runs. Lighter shaded regions are approximated densities over the range of values

