

Contributions

- What languages does **ChatGPT** know?
- Proposal of Alias-Dialect-Accepting evaluation method for LID task.
- Wide-range analyses: geographical, dialectal, and script-wise.

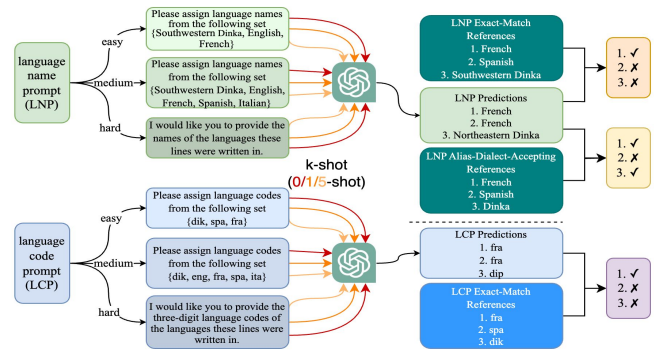
Dataset

670 languages, 24 families, 30 scripts

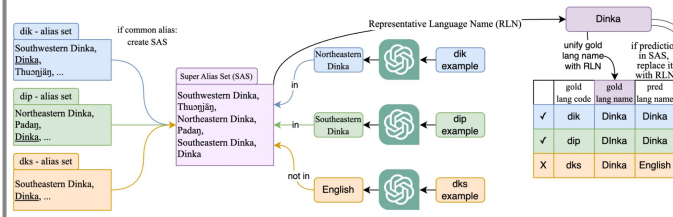
Family	#	%	Family	#	%
Afro-Asiatic	72	10.75	Koreanic	1	0.15
Austro-Asiatic	3	0.45	Kra-Dai	3	0.45
Austronesian	21	3.13	Language Isolate	1	0.15
Aymaran	1	0.15	Mongolic	1	0.15
Chibchan	1	0.15	Niger Congo	386	57.61
Creole	12	1.79	Nilo-Saharan	57	8.51
Constructed Language	1	0.15	Quechuan	1	0.15
Dravidian	4	0.60	Sino-Tibetan	9	1.34
Indo-European	74	11.04	Tucanoan	2	0.30
Japonic	1	0.15	Tupian	1	0.15
Kartvelian	1	0.15	Turkic	11	1.64
Khoe-Kwadi	3	0.45	Uralic	3	0.45

Script	#	%	Script	#	%
Arabic	23	3.43	Hebrew	2	0.30
Armenian	1	0.15	Japanese	1	0.15
Bengali	3	0.45	Kannada	1	0.15
Burmese	2	0.30	Khmer	1	0.15
Coptic	1	0.15	Laoo	1	0.15
Cyrillic	12	1.79	Latin	589	87.91
Devanagari	10	1.49	Malayalam	1	0.15
Ethiopic	6	0.9	Odia	1	0.15
Georgian	1	0.15	Ol Chiki	1	0.15
Greek	1	0.15	Sinhala	1	0.15
Gujarati	1	0.15	Tamil	1	0.15
Gurmukhi	1	0.15	Telugu	1	0.15
Hangle	1	0.15	Thai	1	0.15
Hans	1	0.15	Tibetan	2	0.30
Hant	1	0.15	Vai	1	0.15

Methodology



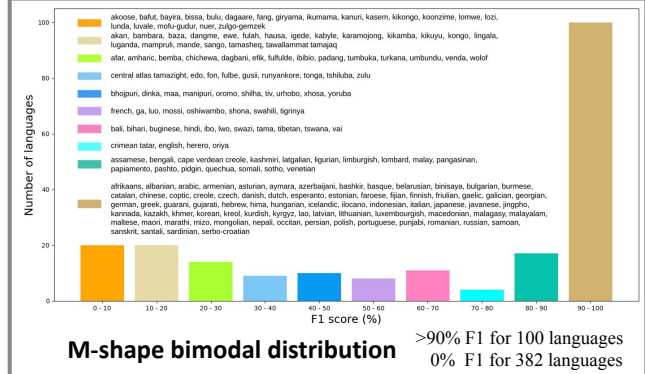
Alias-Dialect-Accepting Evaluation



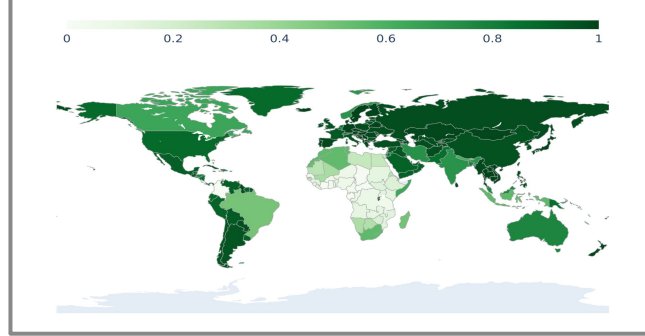
Experimental Results

Level	#	GPT-3.5-turbo-0613												GPT-4-0613											
		LNP (exact)				LNP (ADA)				LCP (exact)				LNP (exact)				LNP (ADA)				LCP (exact)			
		ACC	F ₁	ACC	F ₁	ACC	F ₁	ACC	F ₁	ACC	F ₁	ACC	F ₁	ACC	F ₁	ACC	F ₁	ACC	F ₁	ACC	F ₁				
easy	0	24.86	28.28	24.90	28.14	14.63	17.63	65.82	65.36	65.97	65.92	47.16	46.29												
	1	28.13	31.18	28.23	31.12	27.32	27.51	65.19	65.70	65.29	66.16	45.88	45.96												
	5	33.82	34.46	33.96	34.97	27.37	27.17	68.29	68.05	68.42	68.69	47.95	47.31												
medium	0	19.81	21.09	20.26	21.62	10.27	12.30	48.66	45.39	49.08	46.28	37.79	34.24												
	1	25.30	22.63	26.32	23.58	22.88	19.60	48.02	45.57	48.48	46.22	38.19	34.64												
	5	26.22	23.02	26.94	24.08	23.76	19.64	50.49	47.63	50.89	48.32	39.39	35.62												
hard	0	12.70	12.17	17.39	16.36	2.47	3.82	20.02	17.80	28.32	24.16	21.47	18.93												
	1	16.05	13.81	23.82	19.36	16.66	14.34	20.48	18.26	28.58	24.76	22.10	19.72												
	5	17.12	14.65	25.36	20.25	17.71	15.09	20.94	18.97	28.79	25.31	22.52	20.26												

Visualization



Geographical Analysis



Acknowledgement

NSERC, SSHRC, Digital Research Alliance of Canada, CFL, UBC ARC-Sockeye