Checkpoint 3 Write-Up

Section 1: Thoughtful Use of AI

If we didn't know how to code, we would use three existing LLMs to accomplish our task: Gemini Pro, Claude Sonnet 4.6, and ChatGPT. We would present each phrase in our own test set, and for each phrase, we would ask each of the models to make their top three predictions for the next character. We would count the total occurrences of each character across the three models' predictions, and we'd select the top three characters. If any characters are tied in occurrences, we will break ties by prioritizing characters predicted by Gemini Pro and then Claude Sonnet 4.6, based on the https://artificialanalysis.ai/leaderboards/models rankings of model performance.

Consistent Prompt for Each Model:

Given the following strings, go line-by-line for each string, and provide exactly your top three most probable characters for the most likely next character, in descending order of probability. Return ONLY the characters, with each group of top three characters separated by newlines. Do not provide any additional text, explanation, or formatting.

Section 2: Benchmarking

With our weighted LLM strategy, we had a **100% accuracy rate** on our new test set (all of the predictions were correct). We believe that using multiple models and counting the overall occurrences of predicted characters helped us balance against the strengths and weaknesses of each individual model. Gemini Pro and Claude seemed to interpret the prompt instructions correctly. Those two models outputted the three characters that they interpreted as the most likely "next" character.

However, ChatGPT seemed to misinterpret the prompt instructions. ChatGPT appeared like it attempted to complete each phrase with three characters, rather than outputting the three characters it interpreted as the most likely "next" character. However, ChatGPT often correctly predicted the next character with its first character prediction. By averaging the predictions of the three models, we always predicted the correct character within the top 3 possible characters. Our evidence is below.

Input	Gemini Pro Prediction	Claude Sonnet 4.6 Prediction	ChatGPT Prediction	**True Answer**	**Weighted (Final) Prediction** of the top 3 possible characters	**Is the Prediction Correct**? (contains correct character?)
مرح	ب ا ی	بپت	بای	ب	**ب, ا, ی**	**Yes**
Здрав	с т у	сти	оес	е	**с, т, е**	**Yes**
Guten Mor	g e n	нgG	gn	g	**g, n, H**	**Yes**
Καλημέ	ρ α τ	ραν	ρν!	ρ	**ρ, α, ν**	**Yes**
The quick brown fo	x r d	xXw	xor	x	**x, r, d**	**Yes**
Buenos dí	a s o	aáa	as!	a	**a, s, o**	**Yes**
Comment ça v	a i o	aev	a?e	a	**a, e, i**	**Yes**
मे	र स ल	रर्	ंरह	र	**र, स, ल**	**Yes**
Buongiorn o a tutt	i o a	ioI	io!	i	**i, o, a**	**Yes**

こんにち	**は わ け**	はワは	はわ!	は	**は, わ, け**	**Yes**
Goedemor	**g e n**	gGn	gn!	g	**g, n, e**	**Yes**
Dzień dob	**r y e**	rRr	r!	r	**r, y, e**	**Yes**
Muito obrigad	**o a e**	oOa	oa!	o	**o, a, e**	**Yes**
Доброе утр	**o a e**	oOy	oa!	o	**o, a, e**	**Yes**
Habari yak	**o a i**	eaE	oe?	o	**o, a, e**	**Yes**
สวัสดีครั	**บ ญ า**	บปส	บบบ	บ	**บ, ญ, า**	**Yes**
Teşekkü	**r l m**	rlR	**rnl**	r	**r, l, m**	**Yes**
شکری	**ه ا ي**	اعو	**ها!**	**ه**	**ه, ا, ي**	**Yes**
Xin chà	**o n m**	oOò	**on!**	o	**o, n, m**	**Yes**
很高兴认识	**你 了 的**	你您我	你啊**!**	你	你**,** 了**,** 的	**Yes**

How we used Generative AI: We used Generative AI (Google Gemini) for code completion within the course guidelines and ideas for the test set in languages we do not know. We used Generative AI as required for the Benchmarking section.