Split	Language	Source	Target		Total
			Human	LLM	
Train	C++	LeetCode	2,242	46,888	**49,130**
		CodeForces	33,005	9,766	**42,771**
		GitHub	49,000	19,885	**68,885**
	Python	LeetCode	6,397	44,164	**50,561**
		CodeForces	25,569	9,646	**35,215**
		GitHub	12,442	8,434	**20,876**
	Java	LeetCode	2,283	46,988	**49,271**
		CodeForces	24,121	3,853	**27,974**
		GitHub	48,998	11,874	**60,872**
Validation	C++	LeetCode	282	4,962	**5,244**
		CodeForces	4,194	1,221	**5,415**
		GitHub	1,562	1,056	**2,618**
	Python	LeetCode	738	4,640	**5,378**
		CodeForces	3,285	482	**3,767**
		GitHub	5,500	2,488	**7,988**
	Java	LeetCode	287	4,929	**5,216**
		CodeForces	3,060	1,207	**4,267**
		GitHub	5,500	1,483	**6,983**
Test	C++	LeetCode	283	4,978	**5,261**
		CodeForces	4,203	1,221	**5,424**
		GitHub	1,564	1,056	**2,620**
	Python	LeetCode	728	4,722	**5,450**
		CodeForces	3,291	482	**3,773**
		GitHub	5,500	2,491	**7,991**
	Java	LeetCode	288	4,972	**5,260**
		CodeForces	3,064	1,206	**4,270**
		GitHub	5,500	1,487	**6,987**
Total			252,886	246,581	**499,467**