Bewertungskriterien für den fünften Anwendungsfall

Bewertung

Tests:

	Qwen2.5-72B	**CodeLlama-34B**	**Llama-3.1-70B**	**Mistral-7B-Instruct-v0.2**	**DeepSeek-V2-Lite**	**Summe:**
Klassifizierte Logs	**99**	**40**	**79**	**99**	**99**	
Richtig „Ok“ klassifizierte Logs (TP)	**44**	**8**	**36**	**59**	**40**	
Richtig „Fehlerhaft“ klassifizierte Logs (TN)	**34**	**7**	**30**	**31**	**34**	
Falsch „Ok“ klassifizierte Logs (FP)	**1**	**0**	**1**	**4**	**1**	
Falsch „Fehlerhaft“ klassifizierte Logs (FN)	**20**	**25**	**12**	**5**	**24**	
Precision (P)	**0,9777777…**	**1,00**	**0,973**	**0,937**	**0,976**	
Recall (R)	**0,688**	**0,242**	**0,750**	**0,922**	**0,625**	
Insgesamt Falsch	**21**	**25**	**13**	**9**	**25**	
Insgesamt Richtig	**78**	**15**	**66**	**90**	**74**	
F1-Score	**0,807**	**0,390**	**0,847**	**0,929**	**0,762**	

Anmerkungen:

Wenn Logs falsch klassifiziert wurden dann meist als falsch „Fehlerhaft“. Aus den jeweiligen Begründungen für diese Klassifizierung geht hervor, dass meist Warnungen oder fehlende Rechte als fälschlicher weise als fataler Fehler interpretiert wurde. Demzufolge wurden dann Logs der Klasse „Ok“ als „Fehlerhaft“ klassifiziert.

Eine weitere Anmerkung ist, dass die meisten OPSLLM die Logs erst dann verarbeiten konnten, wenn ein Preprocessing stattgefunden hat. Es mussten alle Zeilenumbrüche, Sonderzeichen und Links entfernt werden, um die Testläufe erfolgreich durchführen zu können.