Apple Macbook mit M2 Chip und 16GB Ram
Settings: Model tiny, device "cpu", compute_type "int8"
Model is instantiated before the task is multithreded (single model loaded)

File ist eine 2min langer Ausschnitt aus einer Kanzler Scholz Rede Transkription von 2 Files parallel per Python Multithreading, Zeit File 1 und File 2 sind die besten aus!

"CPU Cores in use" ist eine geschätze Angabe anhand der Betrachtung der Aktivitätsanzeige. M2 hat 4 Energy efficient und 4 Performace Kerne, diese werden hier vereinfacht zusammengezählt Bevorzugt verwendet der Mac immer die 4 Performance Kerne, danach erst die Energy Kerne.

cpu_threads	num_workers	CPL	J Cores in use	Zeit File 1 in s	Zeit File 2 in s
	1	1	1	. 22,5	27,5
	1	2	2	14,3	17,4
	1	4	2	16,9	17,5
	2	1	2,50	22,5	25,1
	2	2	4	14,5	18,1
	2	4	4	. 17,1	18,6
	4	1	5	30	31,7
	4	2	8	27,5	27,9
	4	4	8	22	25,8
-	-	-		-	-
	2	8	4	. 12,2	15,9
	8	2	8	97,1	97,4
	8	8	8	94,7	96,2