

Debian Server with Epyc 9334 32 x 2.7 GHz 32-Core Prozessor
 Settings: Model tiny, device "cpu", compute_type "int8"
 Model is instantiated before the task is multithreaded (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 5

"CPU % in use" ist eine geschätzte Angabe anhand der Betrachtung der Aktivitätsanzeige.
 Epyc 9334 hat 128 cores

cpu_threads	num_workers	CPU % in use	Zeit File 1 in s	Zeit File 2 in s
1	1	0,8	26,6	27,7
2	1	1,6	14,3	17,4
4	1	3,1	17,9	18,8
8	1	6,20	15,7	16,5
16	1	11,1	15,5	16,4
32	1	24	16,2	16,9
64	1	49	19,7	20,8
128	1	1,6 > 1min	> 1min	
2	2	3,1	13,8	14,1
4	2	6,2	10,2	10,9
8	2	12,30	8,8	9,3
16	2	24,5	8,4	8,6
32	2	48	9,8	10,2
2	4	3,1	13,1	13,9
4	4	6,2	9,9	10,8
8	4	12,40	8,4	8,8
16	4	24,9	8,2	8,5
32	4	3 > 1min	> 1min	
2	8	3,1	12,7	13
4	8	6,3	10,7	10,7
8	8	12,40	8,6	9
16	8	3,1	32,9	33