Clama 3.1 Paper Pretraining - O filtering of large scale training larges @ Devlopment of model Architecture a sealing laws for model line 3) efficient large scale purboining dechinque development pretraining veeige development Removed domains that has personal into of People Removed and adult content; meate content end of 2023 used Hour for high quality diverse tent. built cuton paren, that entrack Himi boiler plate summeral, content recons Usel HML - tent entraeber like. found - Marddown is harmful to that model performance. is trained on web date than plane tant

Serveral rounds of de duplication

Heuristic filtuing: - s to summer low zuit zualty downent followed Similar pipeline pursonal dots Multilinguel Data : - language identification very fast tent -> downment & line level dedup beating - Applied language level heristics of humere model hand filterry I low quality Ly Categoryse type of info Ly down sample data categories that are over represented Dorta Mix Knowledge classification Scaling laws:

	· · · · · · · · · · · · · · · · · · ·
to londan	
Mann 34 paper (continuation)	Tracks that will
deallocali th	1 h. consulation
llam 34 paper [continuation] I seedine memory west -> deallocate the not be used How to figure out which including input feature are not being west? each pipeline	for further longer is
A which muluding input	a compart of a
flow to by give out herry want? . Barla Dipeline	
Jewers are the	1 performan
ton estimator	end perf
Developed menery conting on love	parallel un anglig
Jewes are not many convery how estimator Developed menuny convery how extinuator Pa 12 Projection tool — In employe and overall and identife	having performan
(Pa 12) projection and overes	menung gay
Training: - 1 initial prefraining - 800 B) Annealing	F 1 1 5
Adam Adam	1 2 = 0 × 10
raining :- 1 mitral protraining - 800	B tokens , "
(i) the carrier	
3 Annealing	

Man 14 poper (continuation) It seedene number which the best of further computation in the bigue out which mulustry input a output fencer is feature are not buy it is care pipeline) Developed menusy convery how estimator and performan Pa12 Brojection tool - to emplore parallelien config ment more of training ment and overall training performance and identify menuny gar Training: - 1) initial pretraining - Adam, & = 1×105 1) long content pretraining - 800 B tokens, 120 k content

B) Annealing