03:59:39 summary, phew, build-nanogpt github repo

okay so at this point I should probably start wrapping up the video because I think it's getting way longer than I anticipated uh but we did Cover a lot of ground and we built everything from scratch so as a brief summary we were looking at the gpt2 and GPT 3 papers we were looking at how you set up these training runs uh and all the considerations involved we wrote everything from scratch and then we saw that over the duration of either a 2-hour training run or an overnight run we can actually match the 124 million parameter checkpoints of gbt2 and gpt3 uh to a very large extent um in principle the code that we wrote would be able to train even bigger models if you have the patients or the Computing resources uh and so you could potentially think about training some of the bigger checkpoints as well um there are a few remaining issues to

address what's happening with the loss here which I suspect has to do with the fine web edu data sampling uh why can't we turn on Torch compile uh it currently breaks generation and H swag what's up with that in the data loader we should probably be permuting our data when we reach boundaries so there's a few more issues like that and I expect to be documenting some of those over time in the uh build n GPT repository here which I'm going to be releasing with this video if you have any questions or like to talk about anything that we covered please go to discussions tab uh so we can talk here uh or please go to issues or pull request pull requests um depending on what you'd like to contribute or also have a look at the uh Zero to Hero Discord and uh I'm going to be hanging out here on N GPT um otherwise for now I'm pretty happy about where we got um and I hope you enjoyed the video and I will see

you later