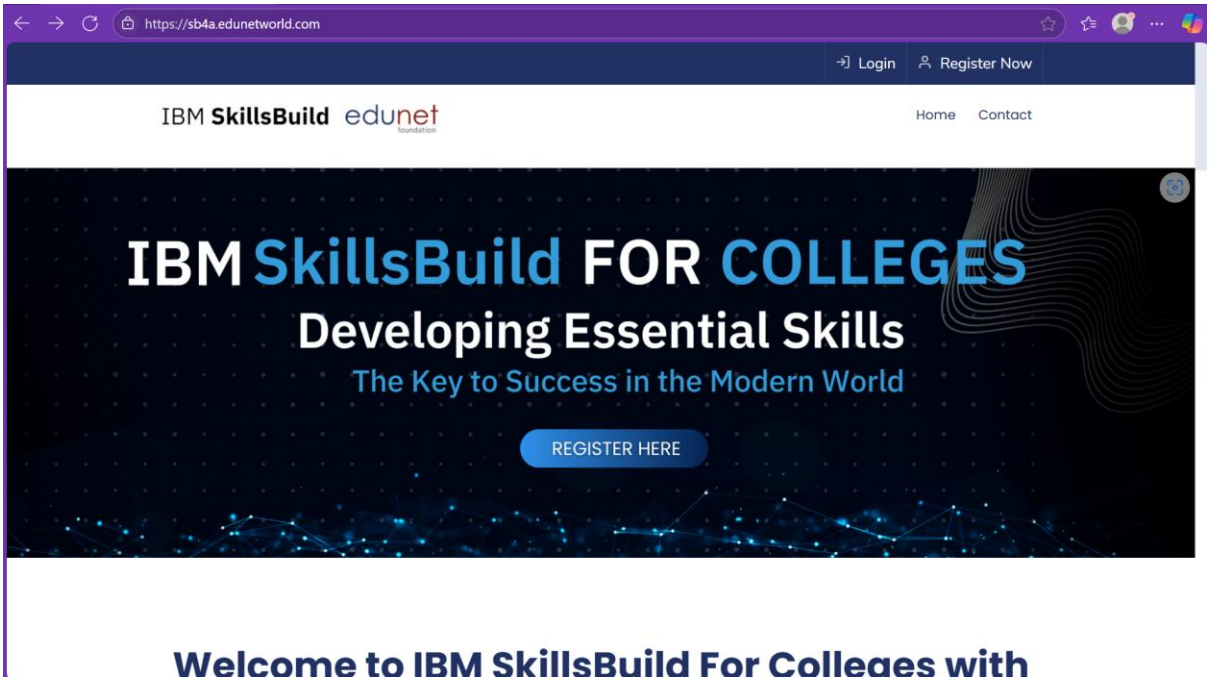
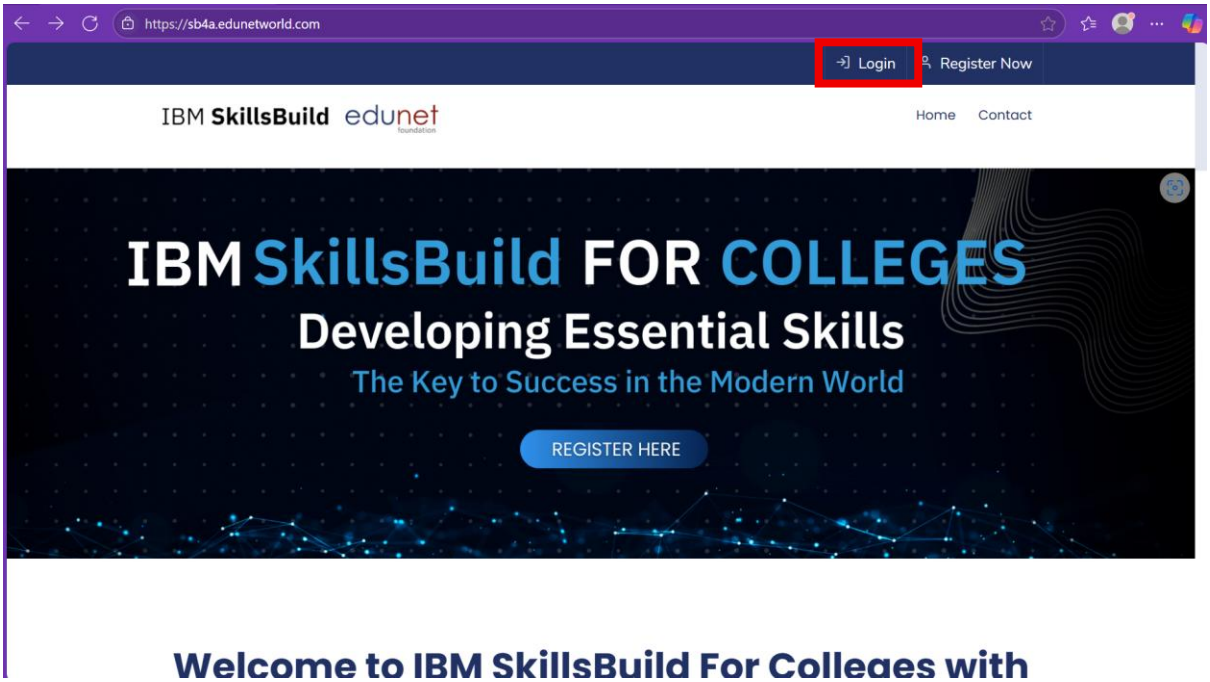
Introduction to RAG Lab

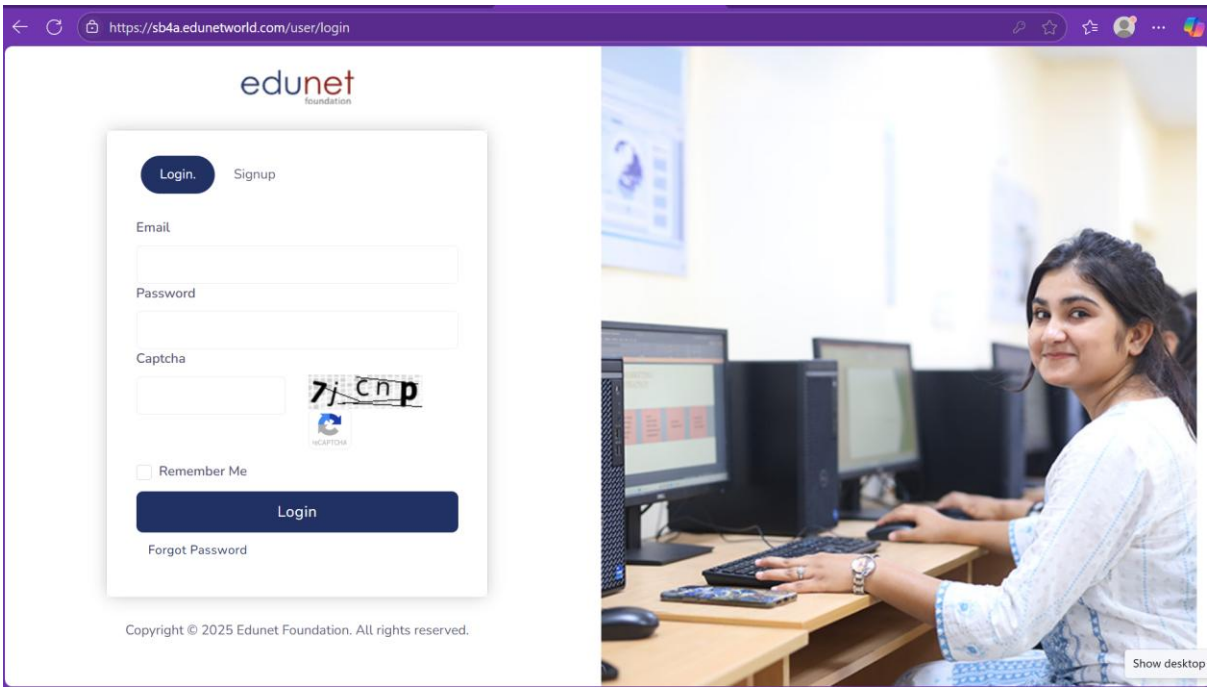
1. Open the browser and type https://sb4a.edunetworld.com/ URL on address bar. Then it directs to the following page.



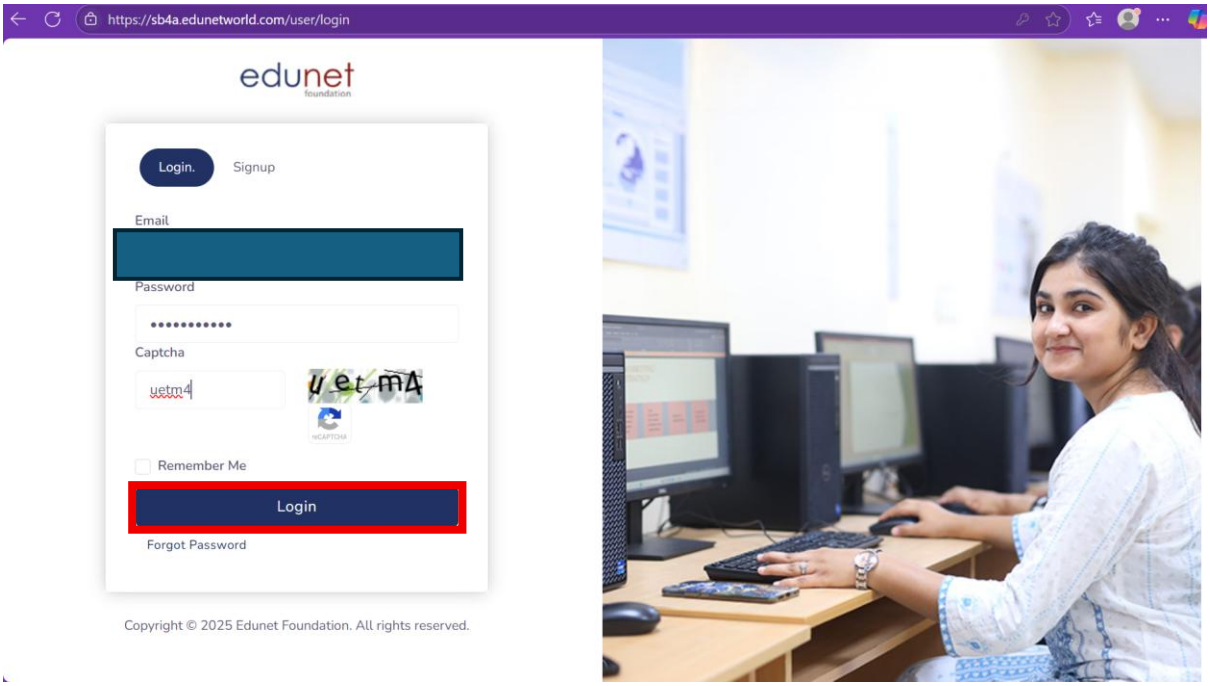
2. Click on Login option from top right corner.



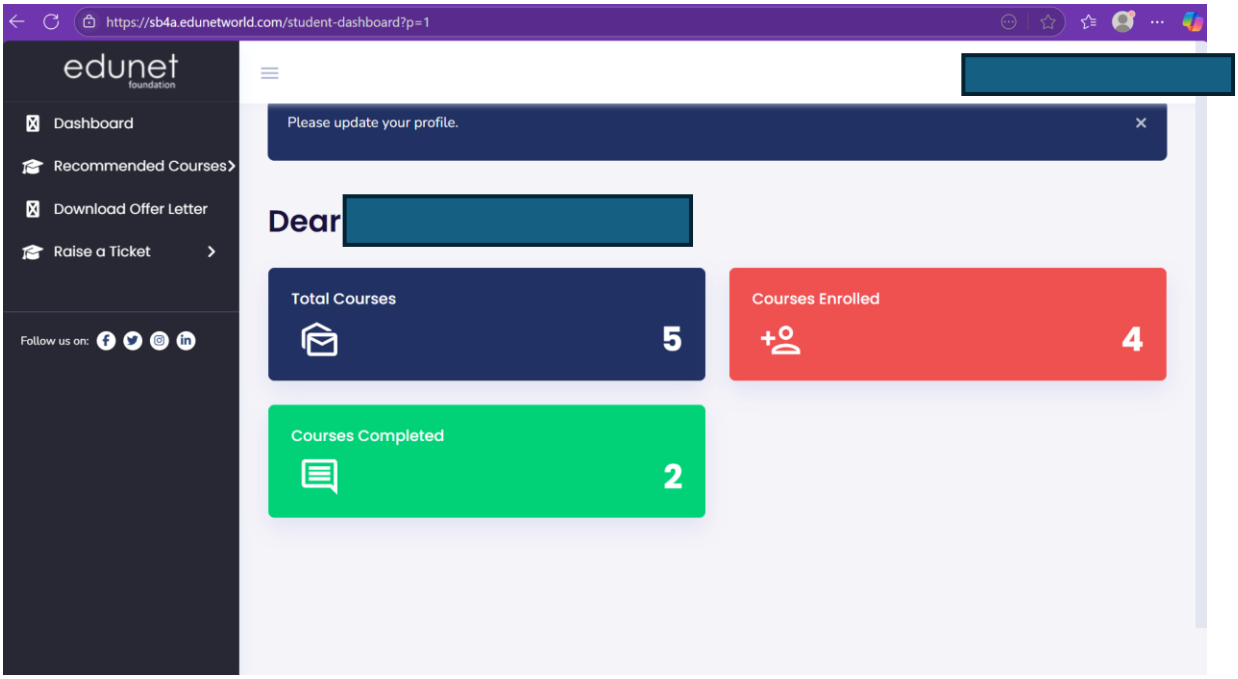
3. Login page will be opened.



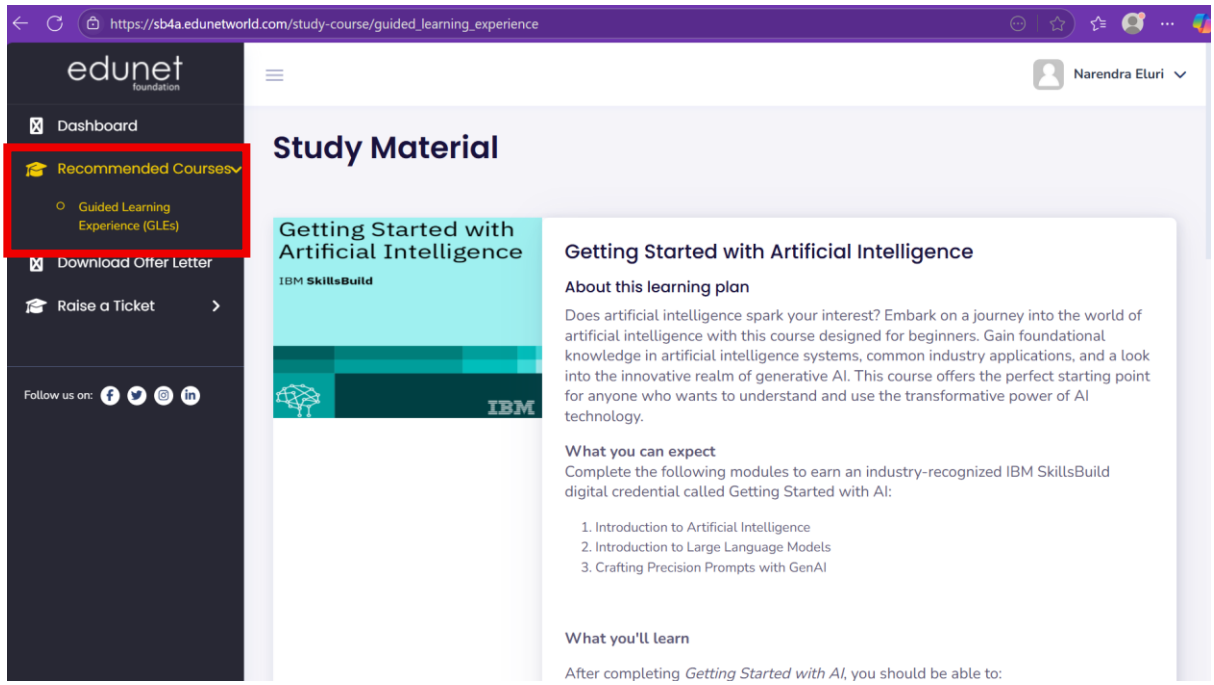
4. Enter your valid Login Credentials, enter captcha and click on login button



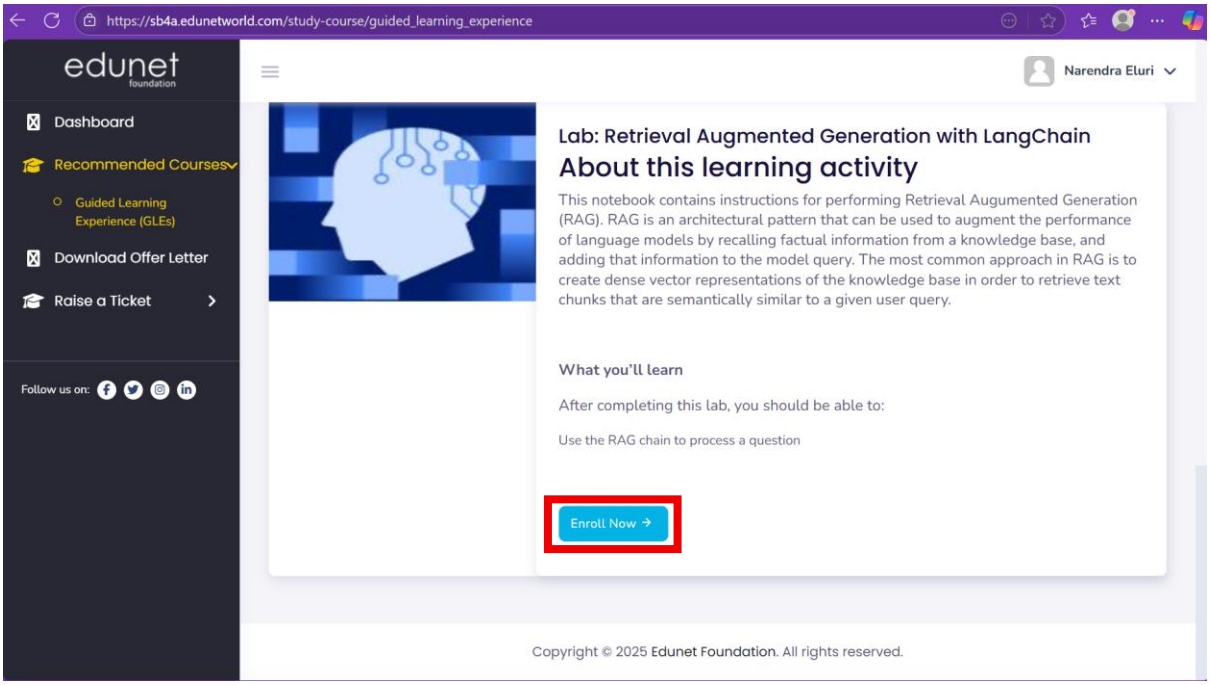
5. You will be logged in successfully. You can see the dashboard like this.



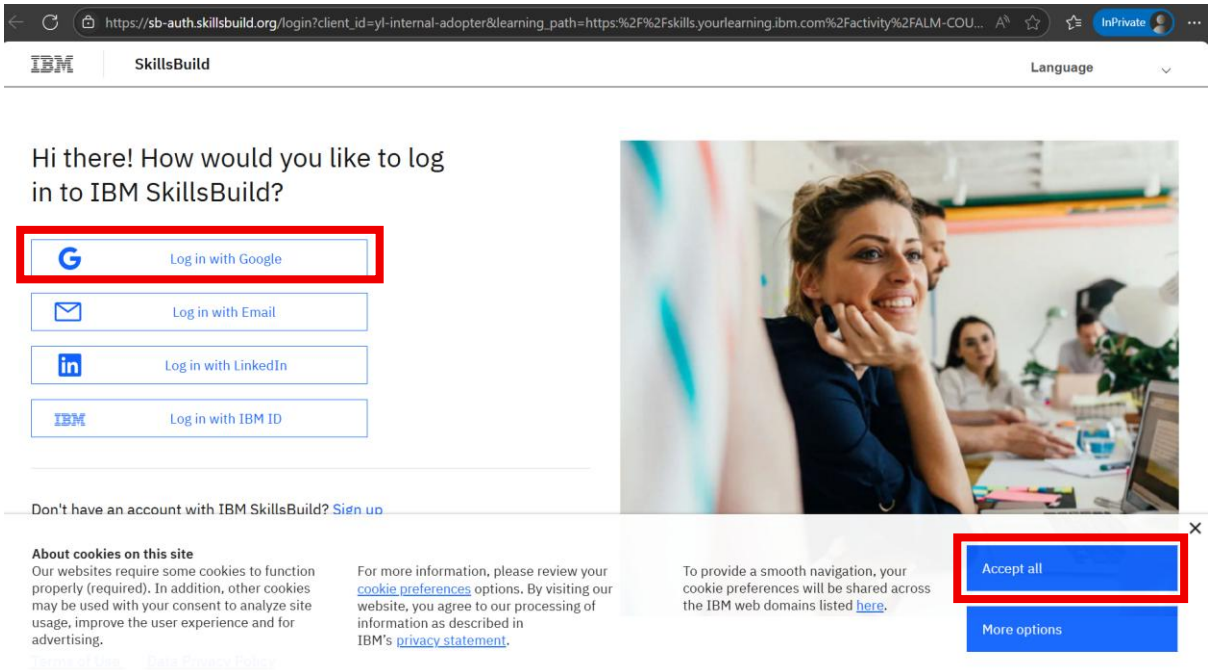
6. Click on Recommended Courses from left menu and then select Guided Learning Experience (GLE)



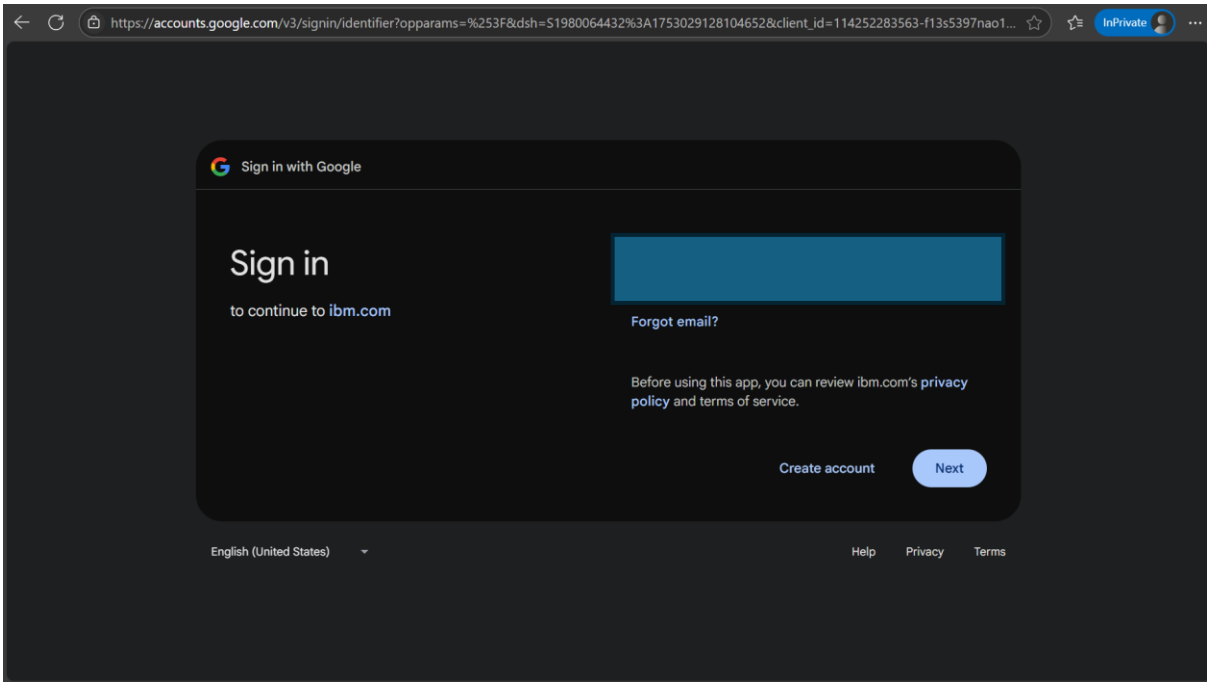
7. Scroll down and find Lab on Retrieval Augmented Generation with Langchain and click on **Enrol Now button**



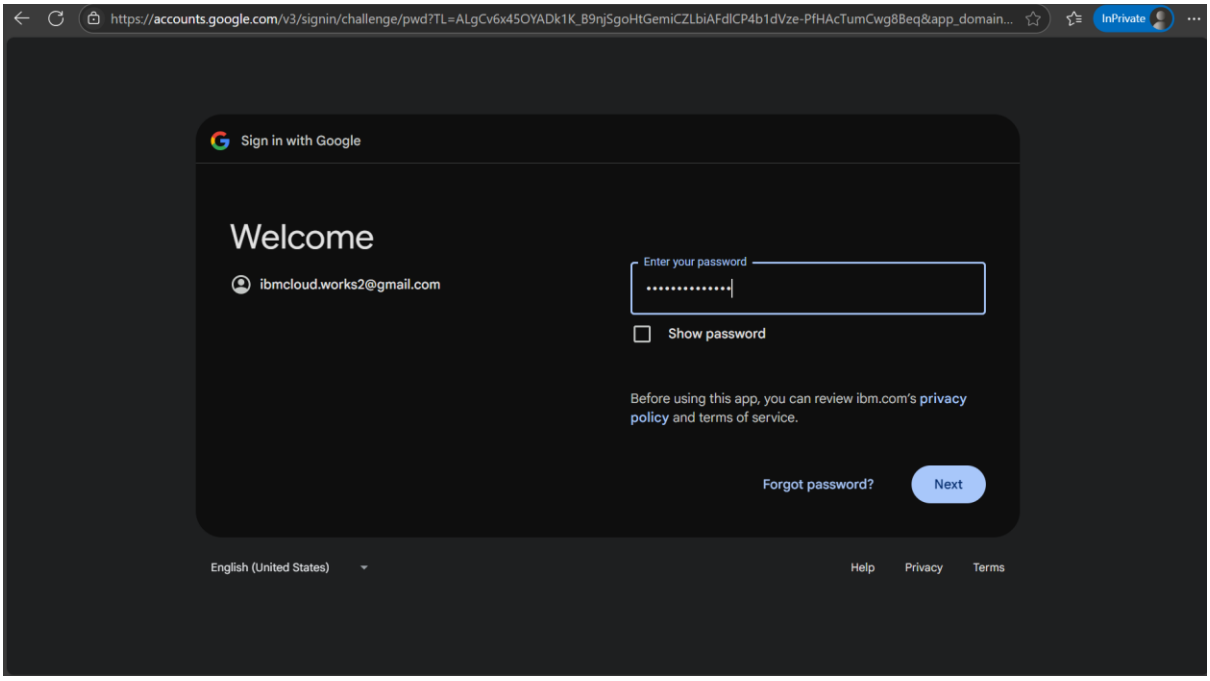
8. It takes you to IBM SkillsBuild Login page. Click on **Accept all** button to allow all cookies and click on **Log in with Google** button.



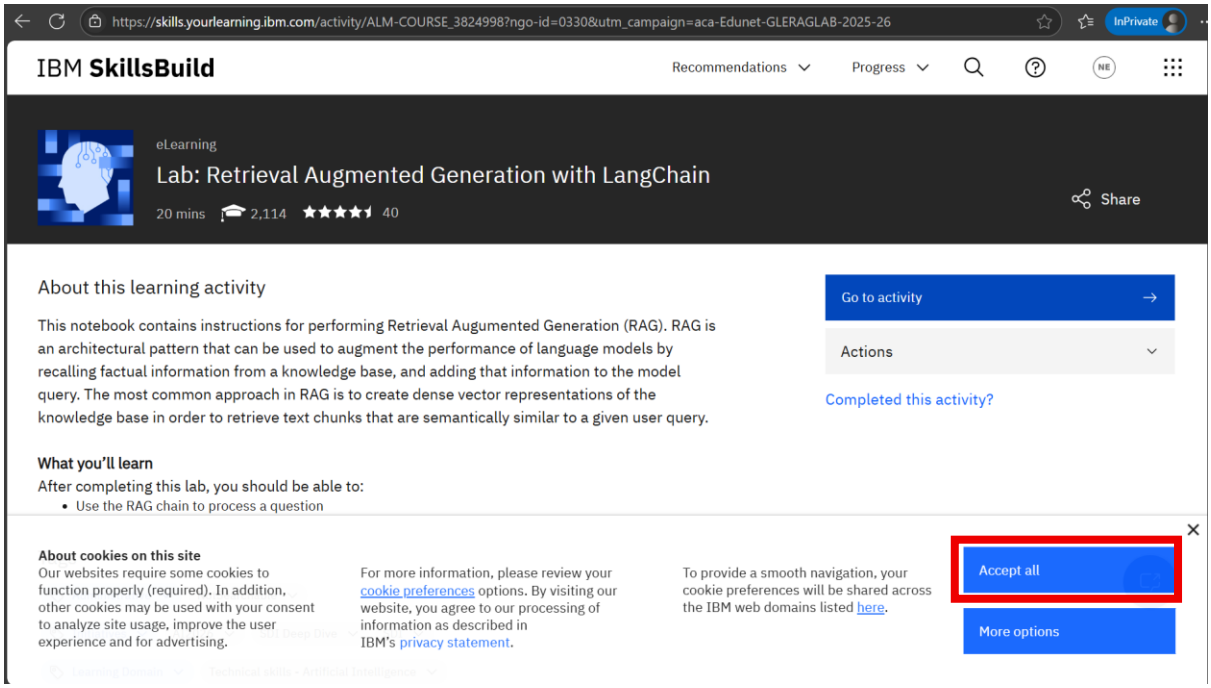
9. It asks you to enter your LMS Gmail id to login (or it prompts you to select your LMS Gmail id from the list) and click on Next



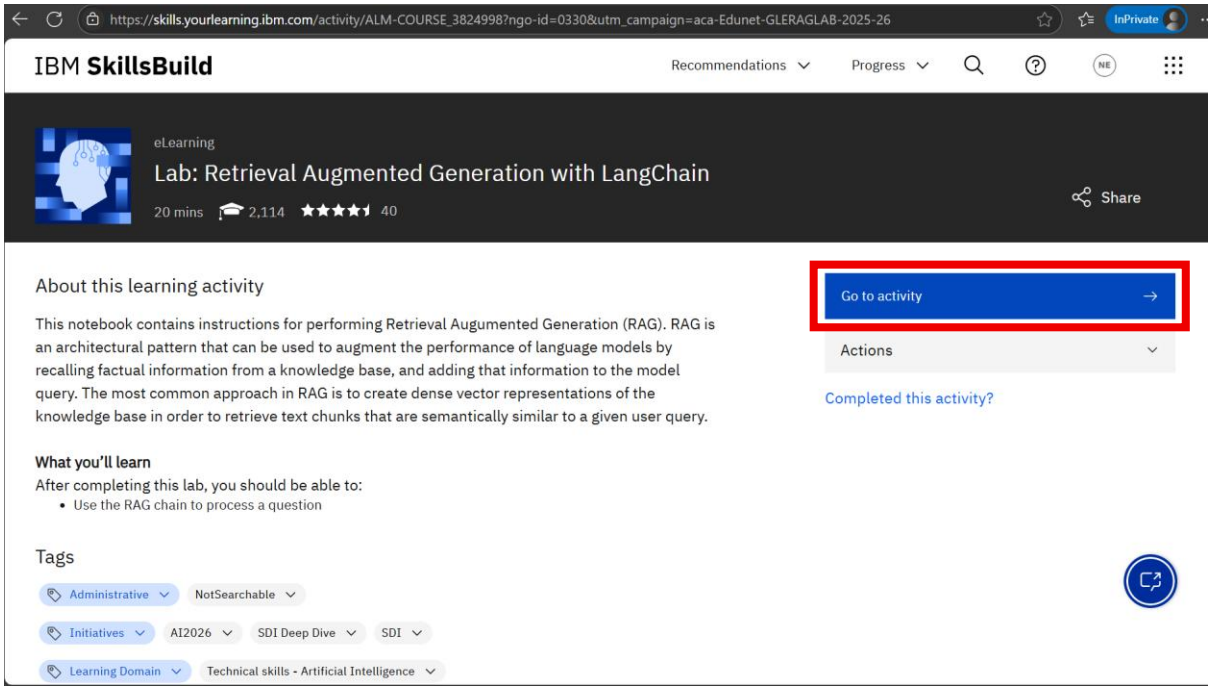
10. Enter Password of your Gmail and click on Next.



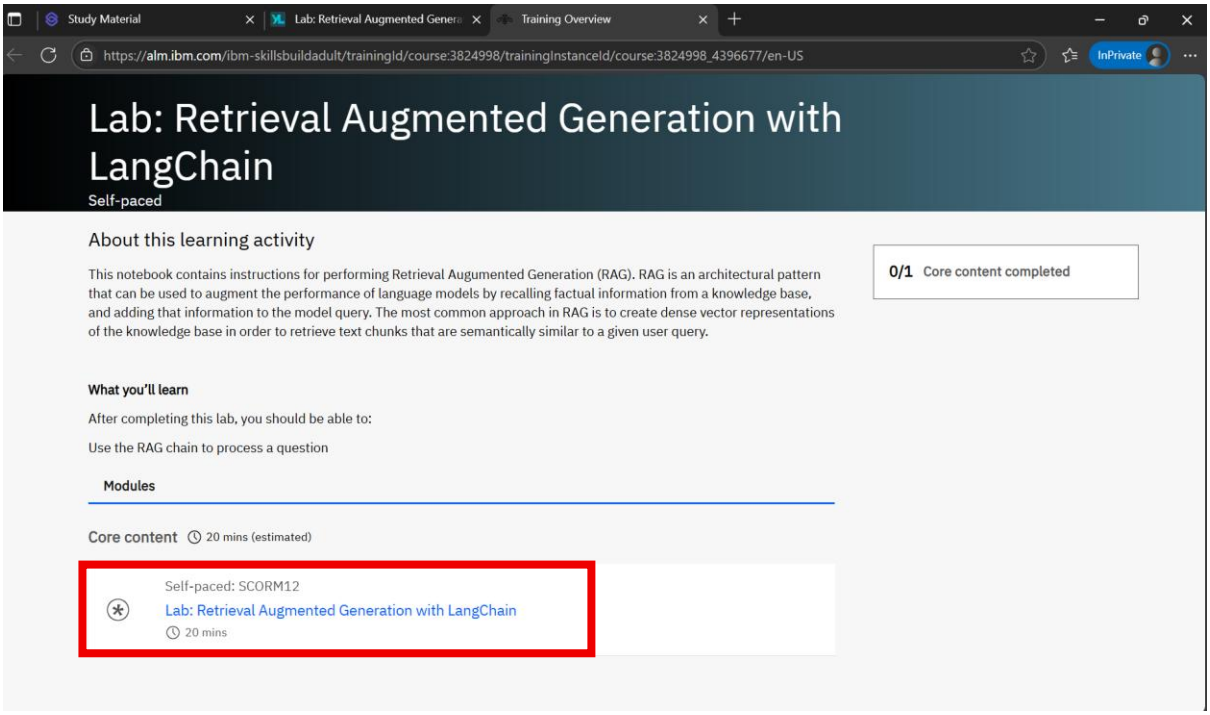
11. Now you are logged in to IBM SkillsBuild platform for the Lab. Click on Accept all button.



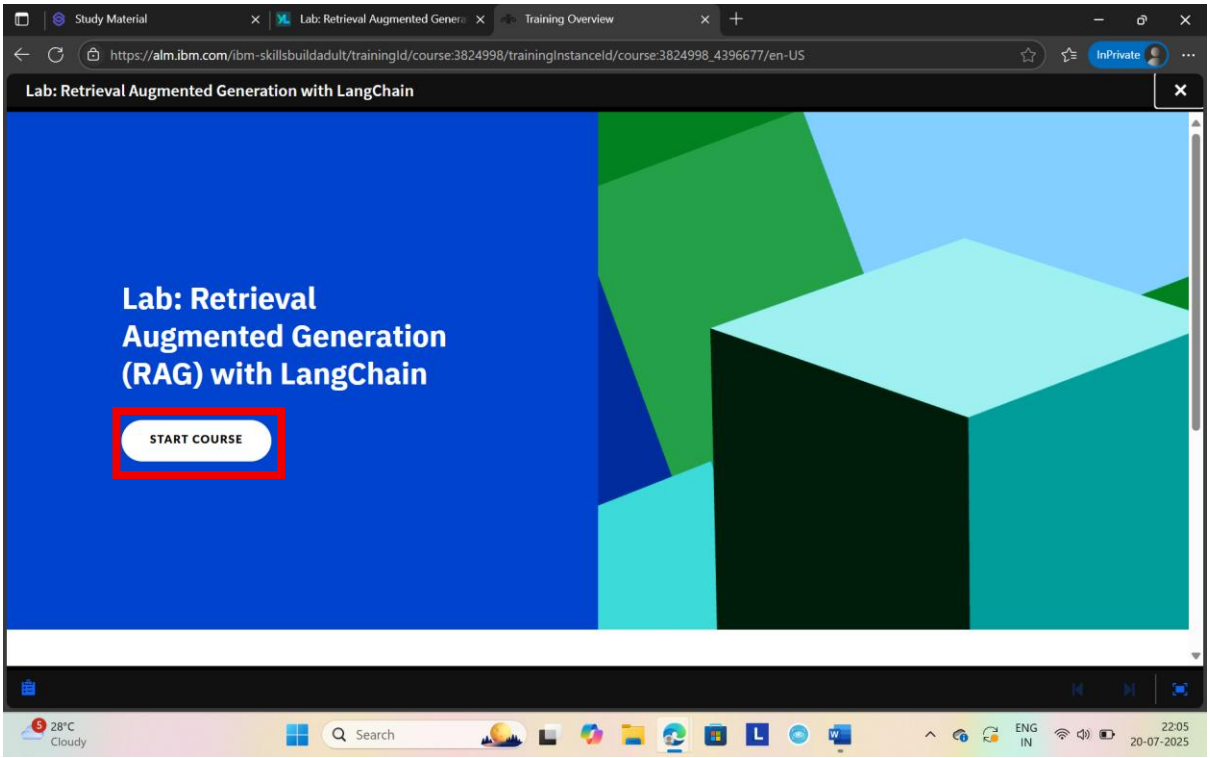
12. Click on **Go to Activity** button



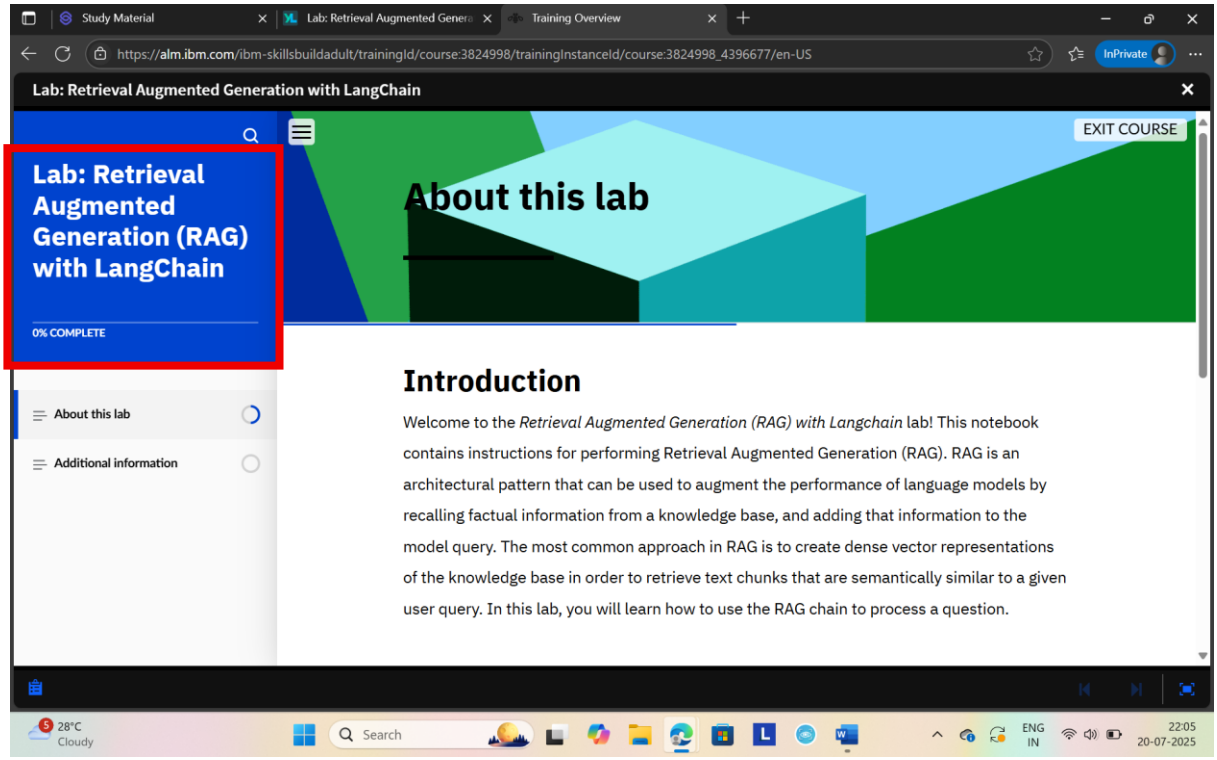
13. In the next page, scroll down and click on Lab: Retrieval Augmented Generation



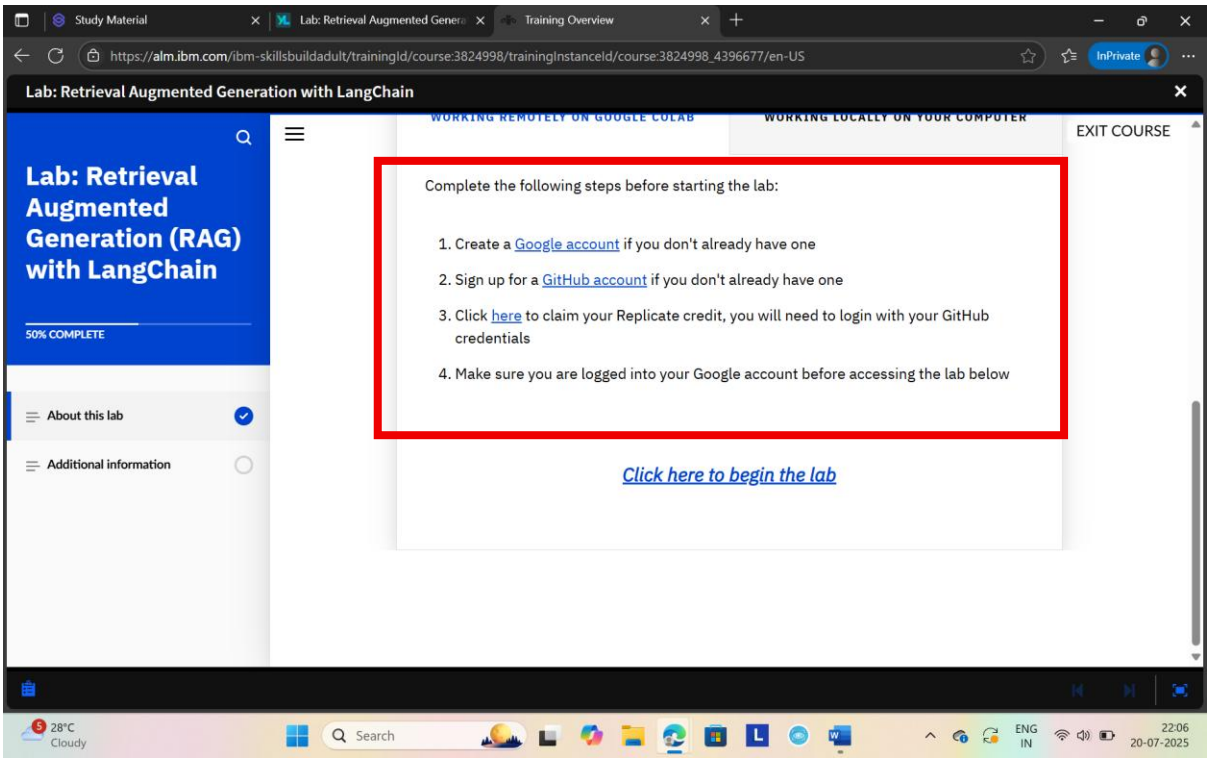
14. It will show start **Course button**. Click on it to start the Lab.



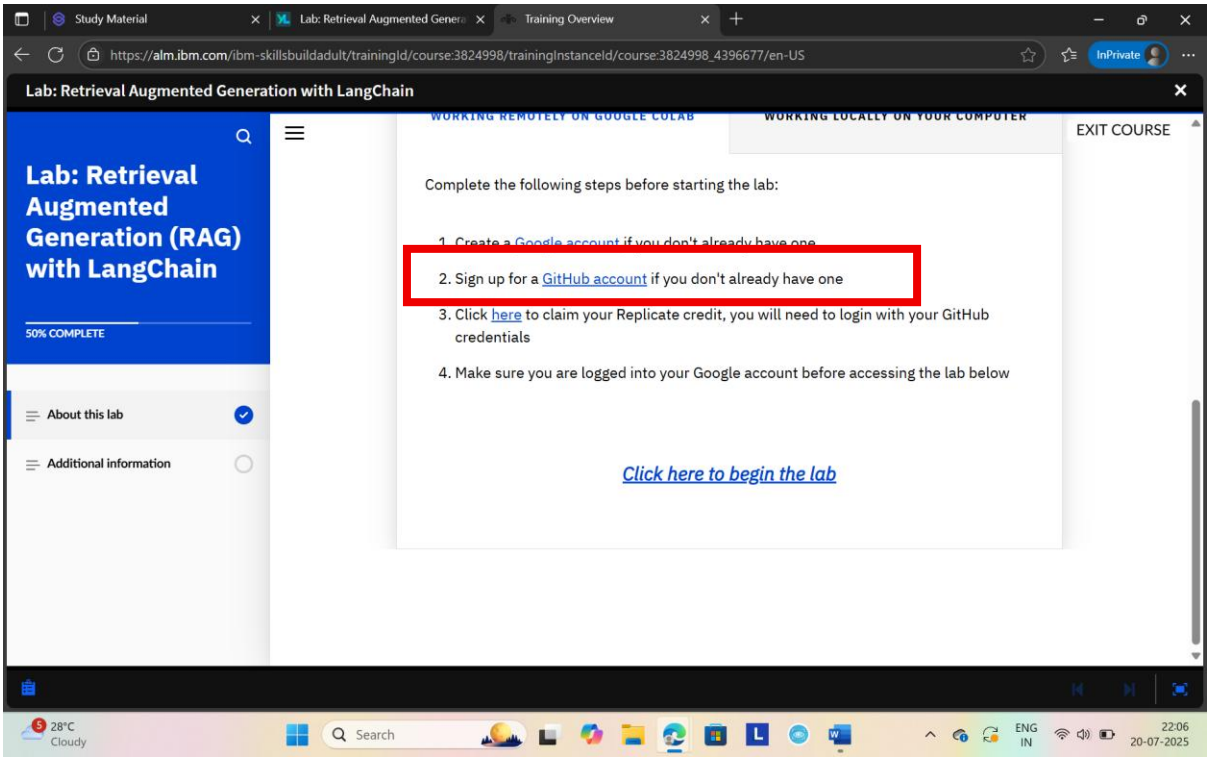
15. The Lab has started. This is the first interface for the lab. Scroll down to find the requirements to complete this lab.



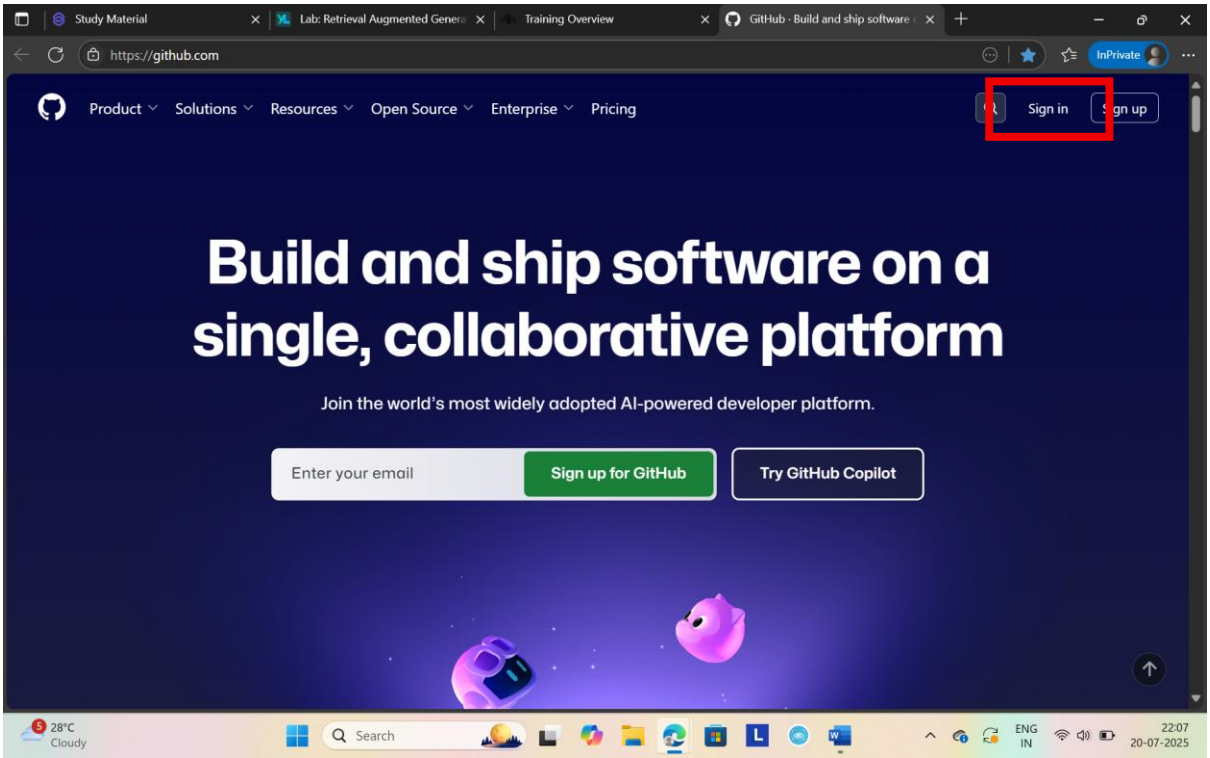
16. The requirements are shown at the bottom of the page like this.



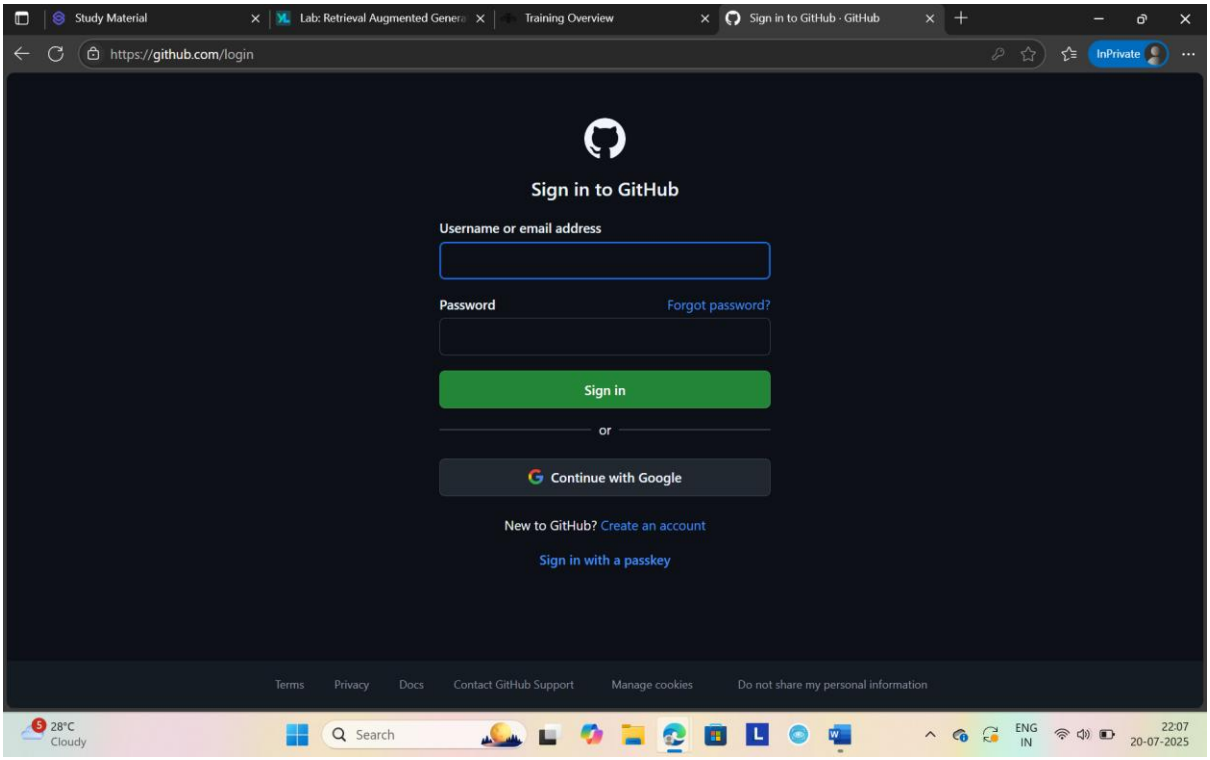
17. Since Gmail already created for you, click on Github account link.



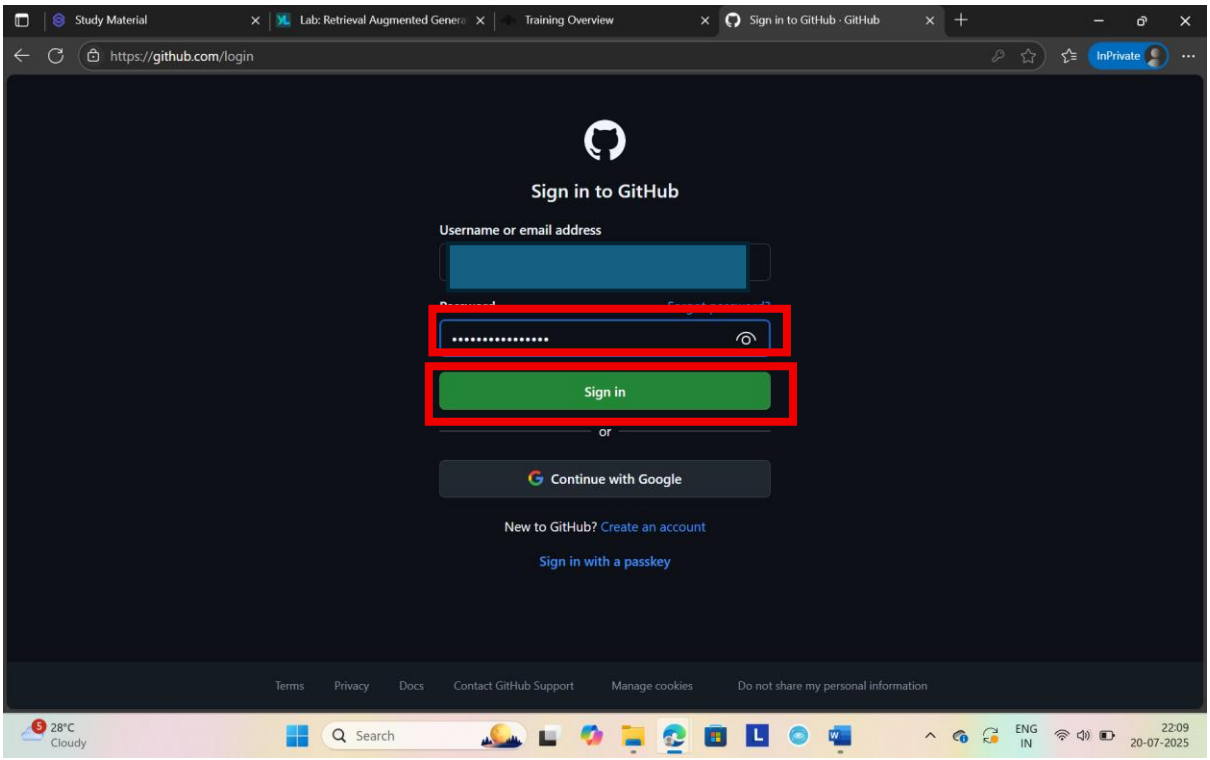
18. Sign in to Github account, if you have. Otherwise create a Github account on the same Gmail id and sign in. You can find the procedure for creating Github account with Gmail document, if you not yet created your Github account.



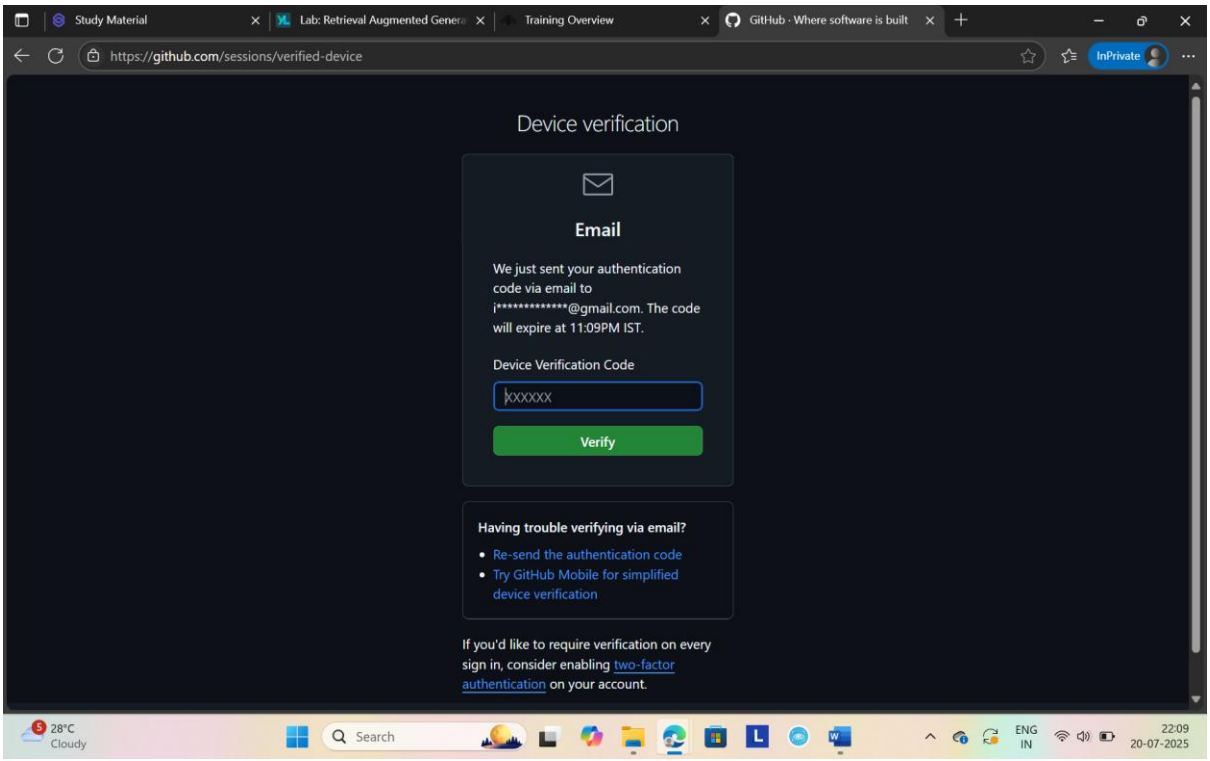
19. Provide your username for the Github account to sign in and hit enter.



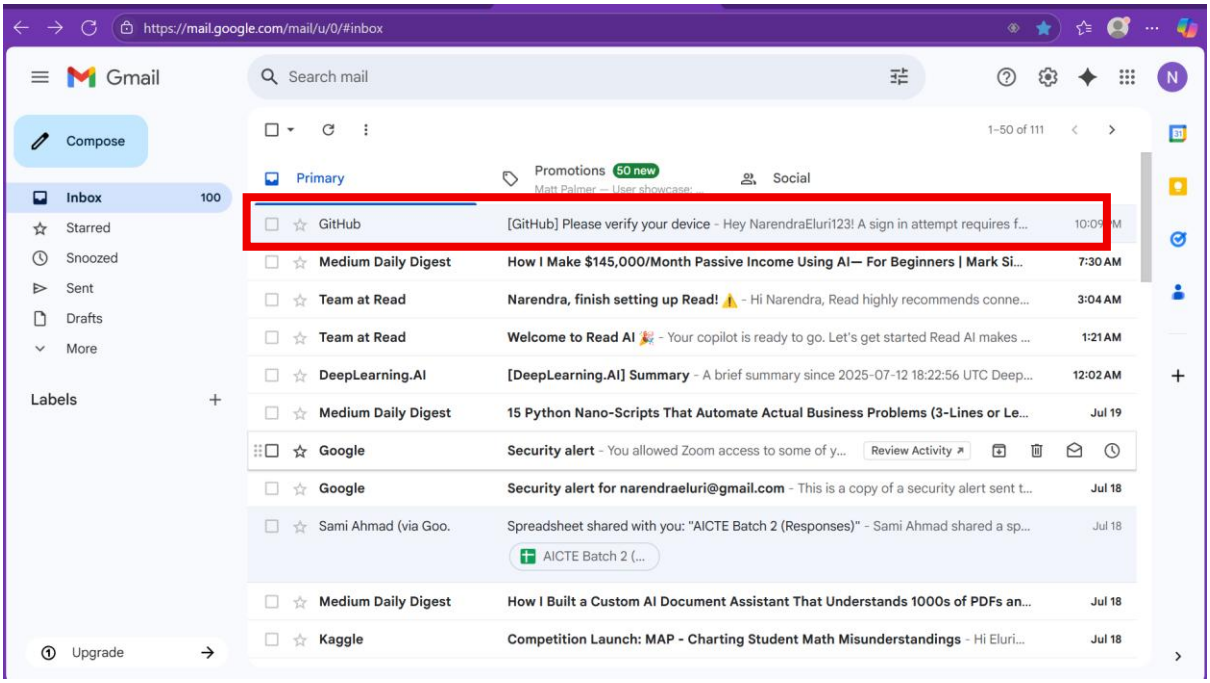
20. Provide password for the Github account and click on Sign in button.



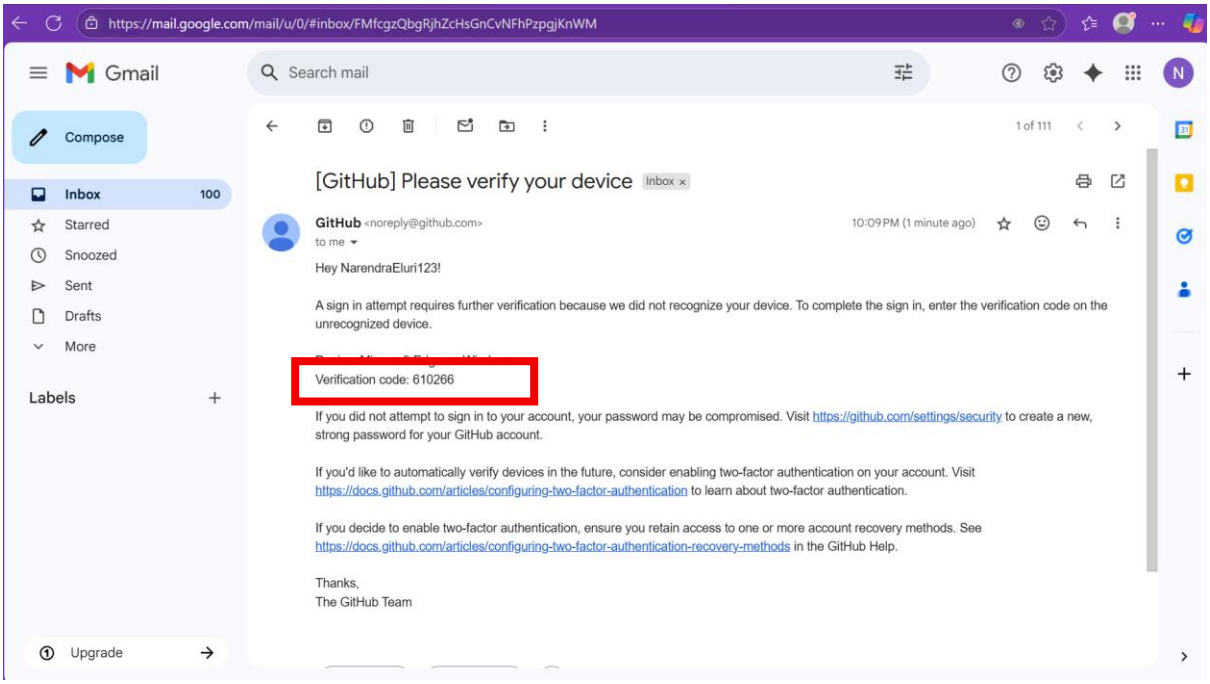
21. It sends verification code to your Gmail.



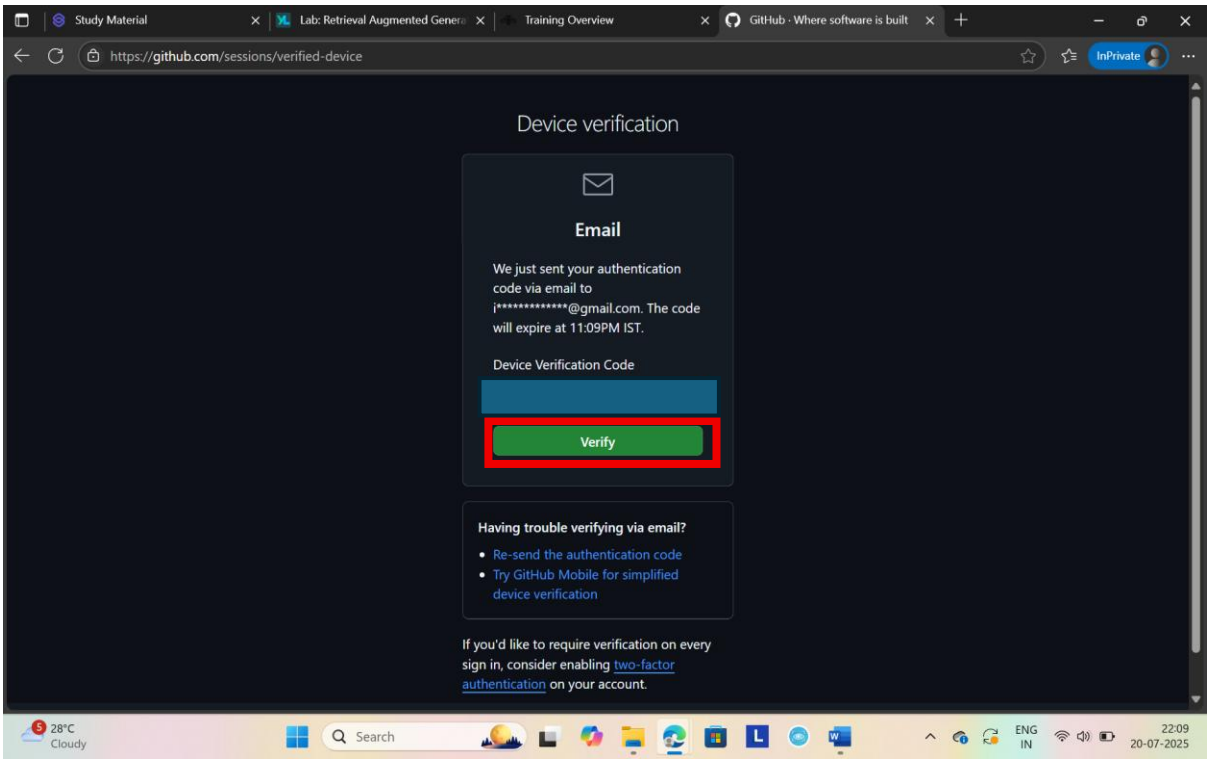
22. Go to Your Gmail inbox and find a mail from Github.



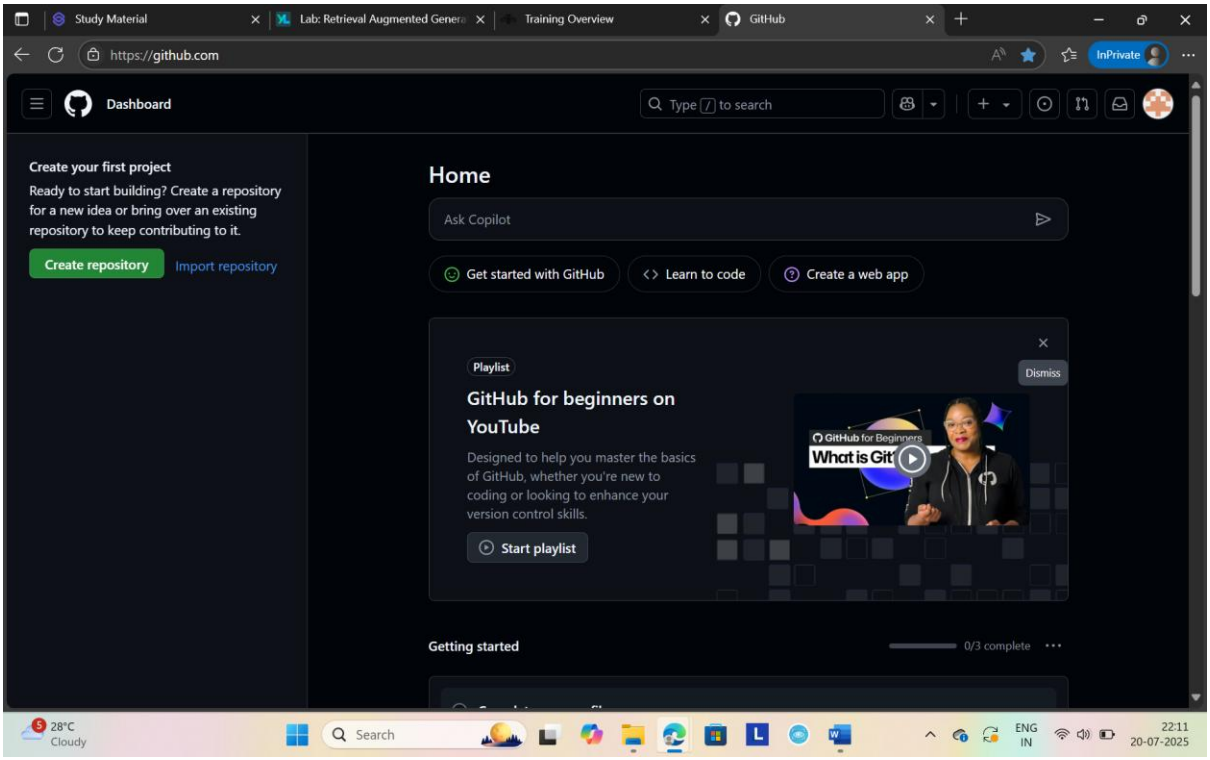
23. Open the mail received from Github and find the verification code. Copy the code.



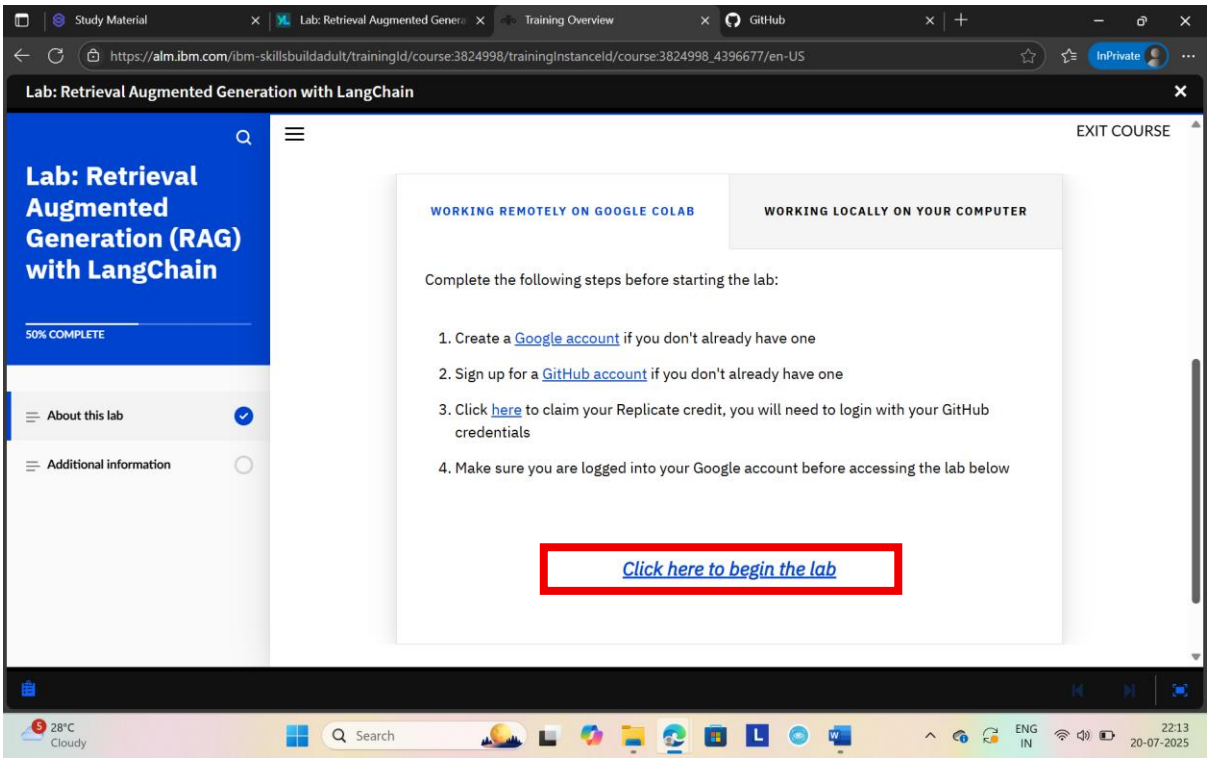
24. Return to the Github login page and paste the verification code in the text box and click on **Verfity** button.



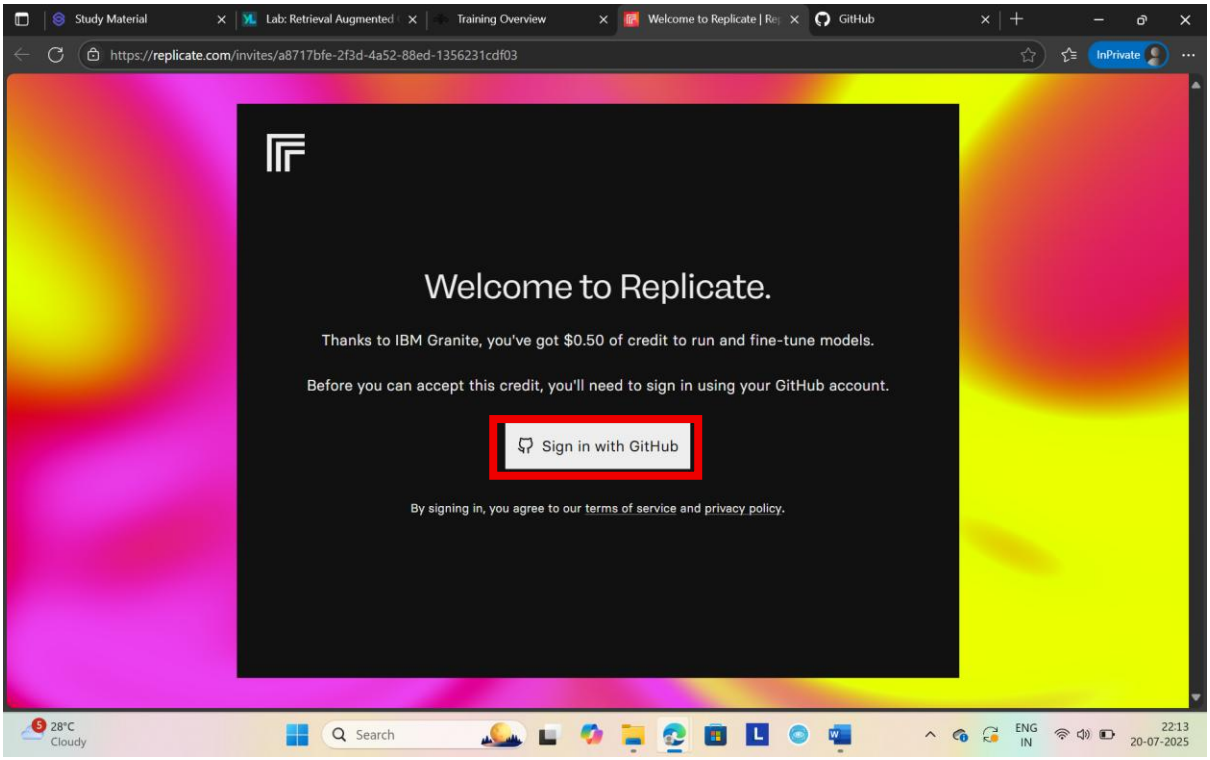
25. After successful verification of the code, you will be logged into your Github.



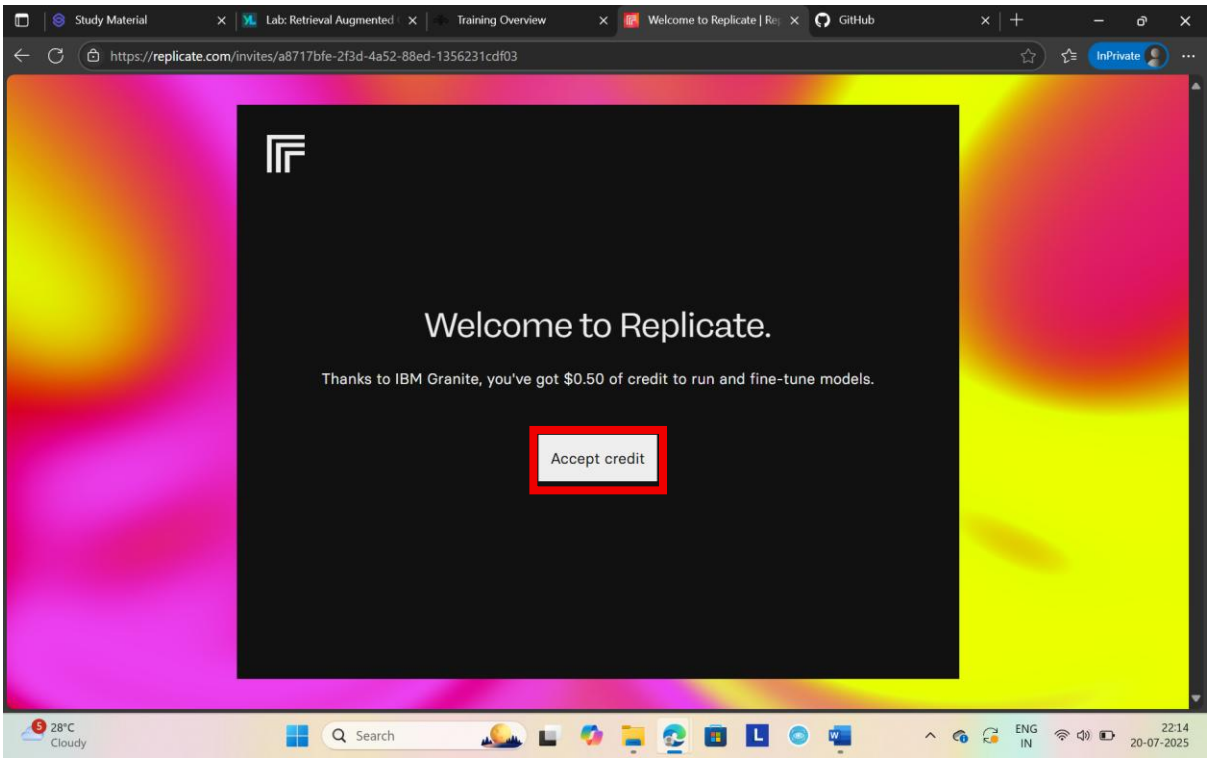
26. Keep your Github account open, go back to the Lab. Click on **Click here to begin the lab.**



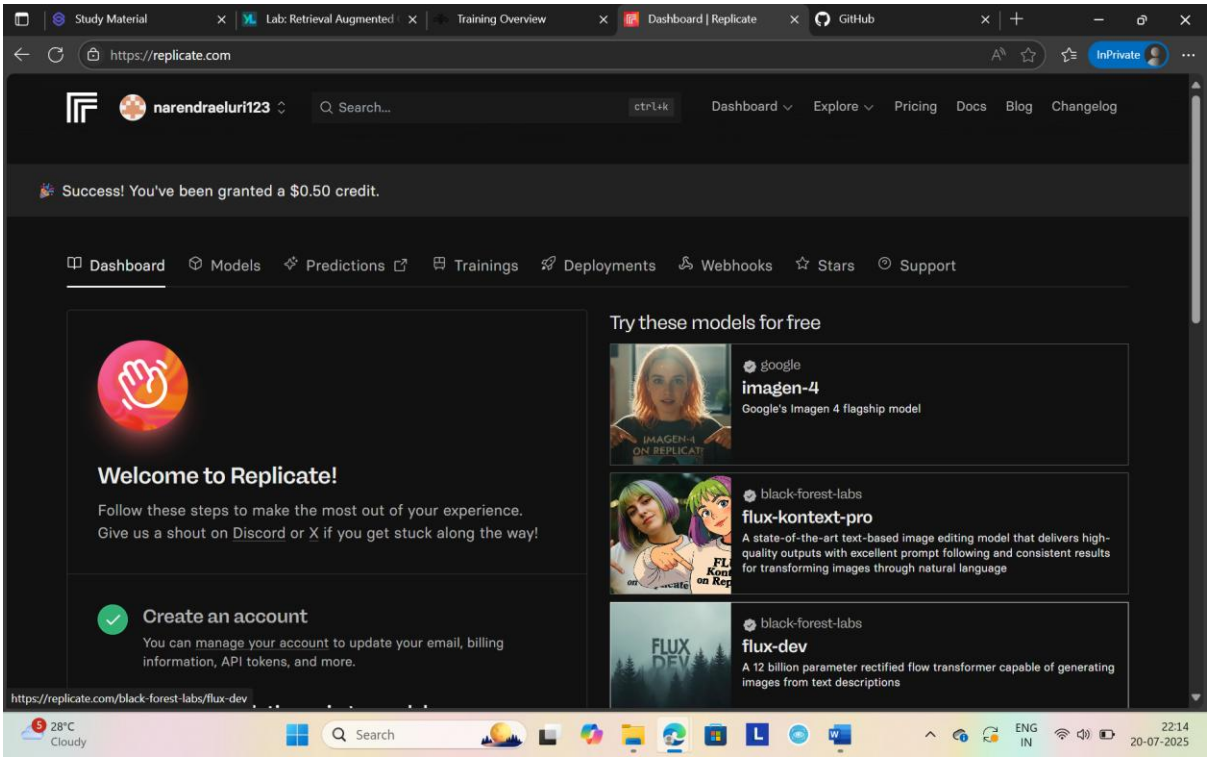
27. It shows to login with Github. Click on **Sign in with Github button.**



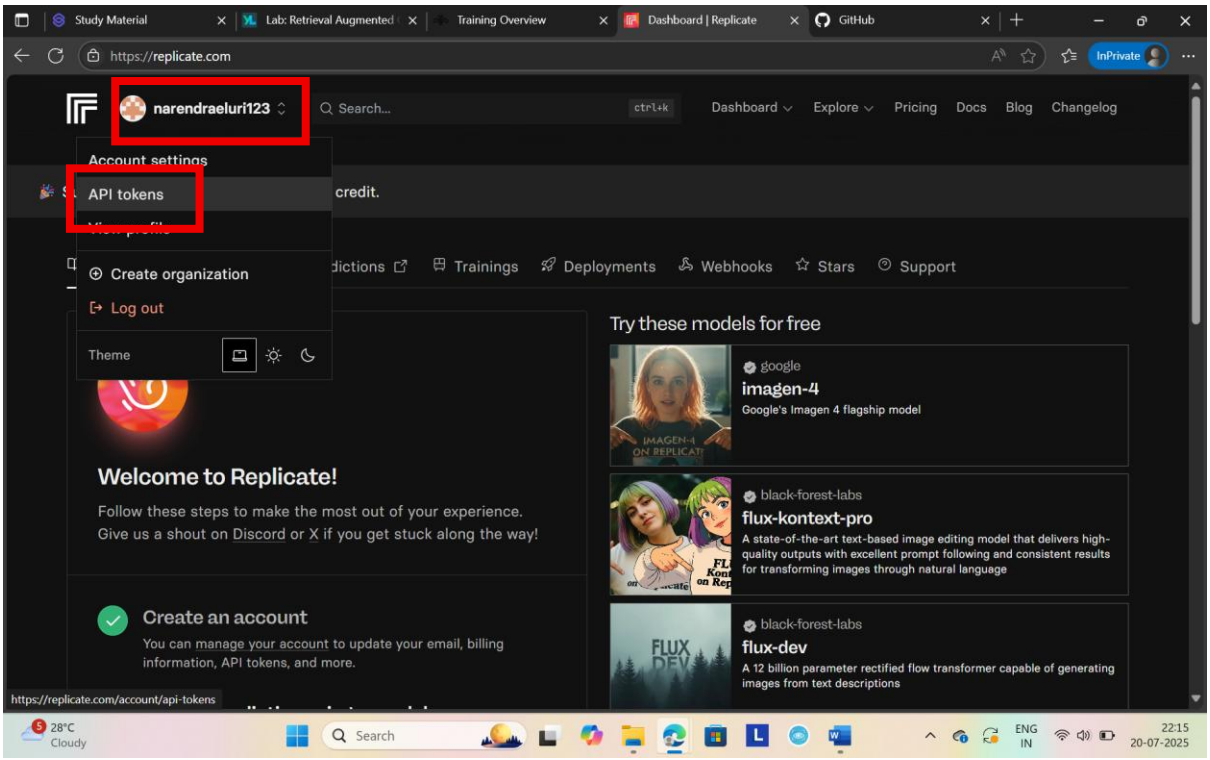
28. Click on **Accept Credit** button to accept 0.5 dollar credit.



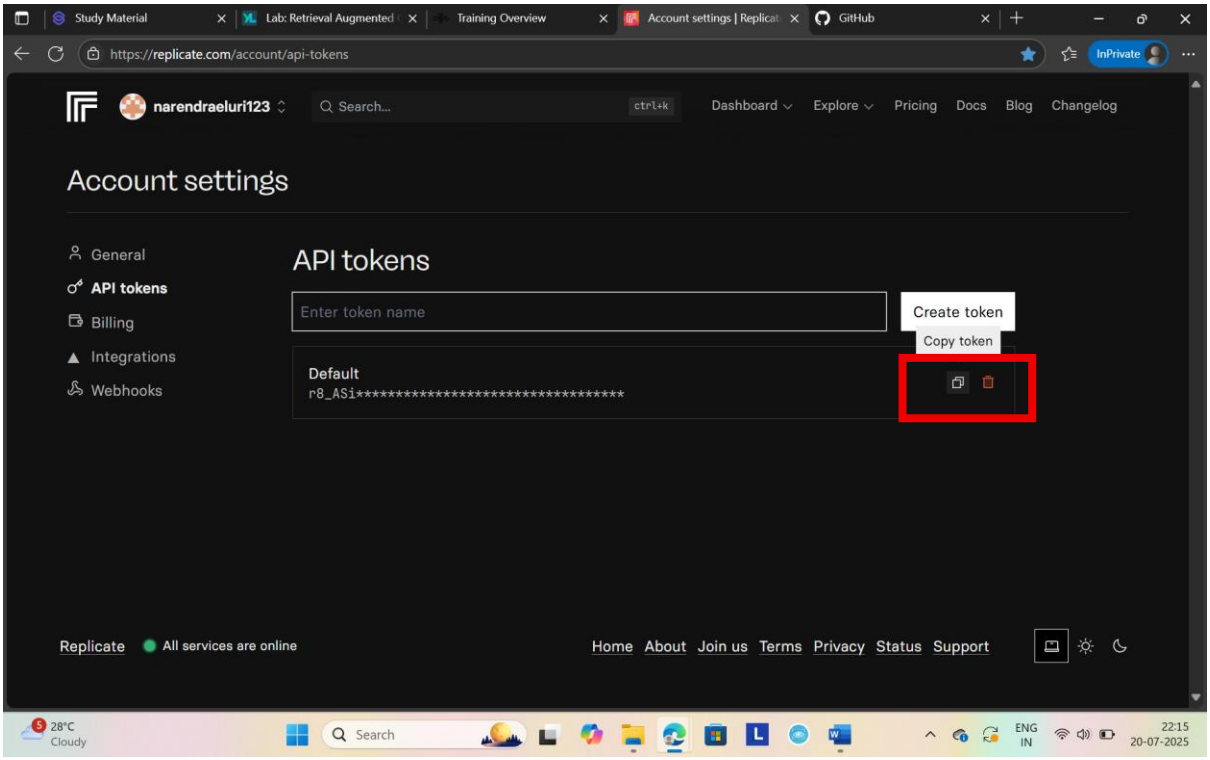
29. You can see this interface from where you need to get API Token.



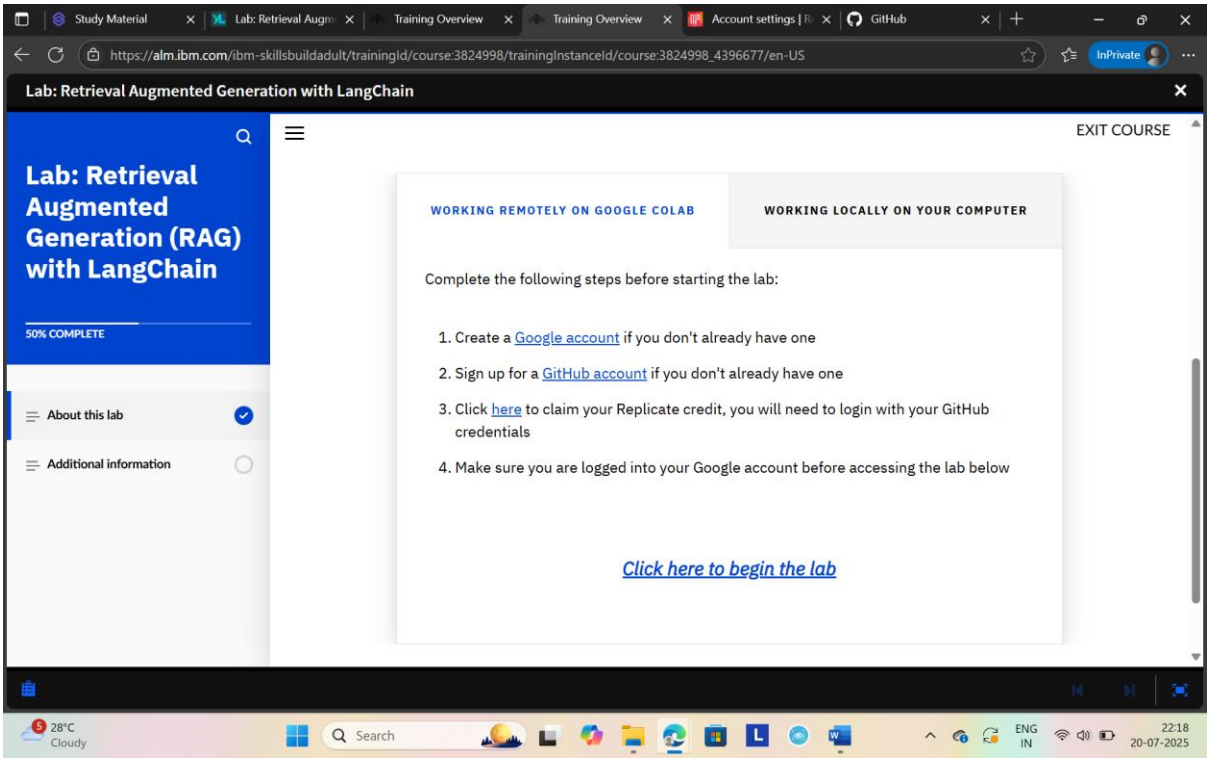
30. Click on Profile button on top left corner, select API tokens.



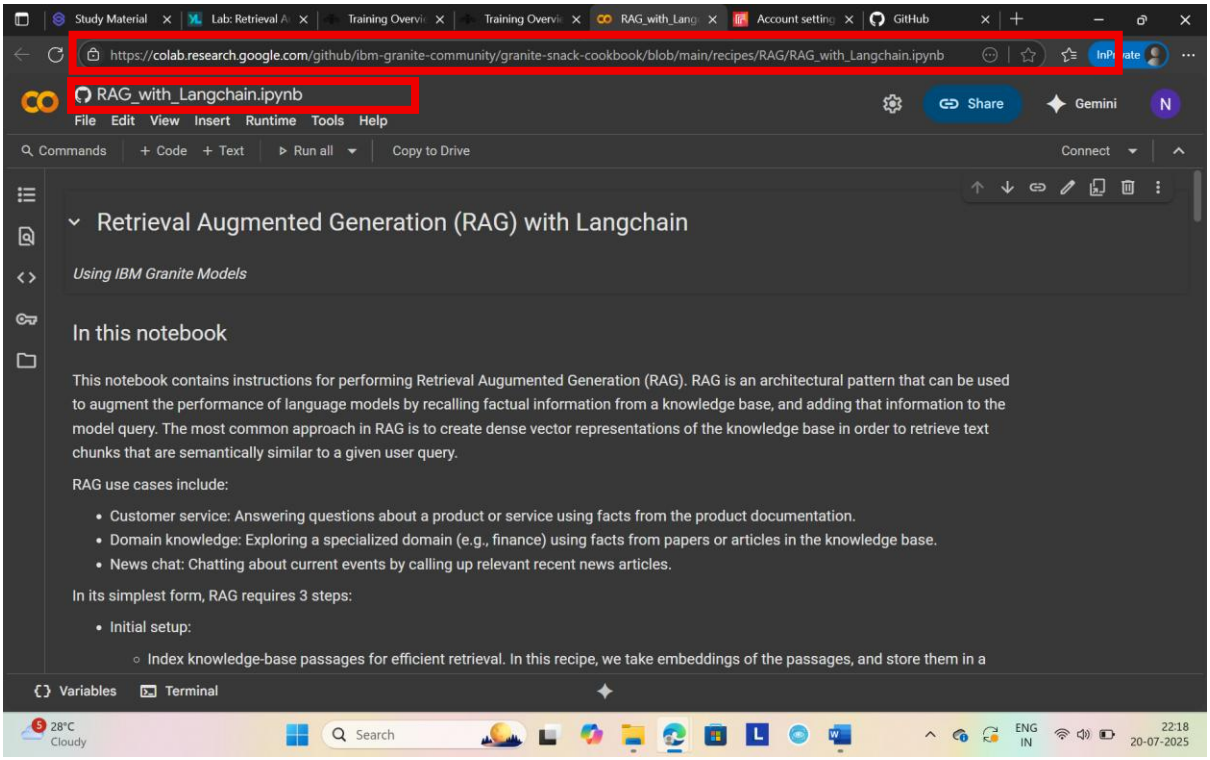
31. Copy the API token by clicking on copy icon. Save it in notepad for future reference.



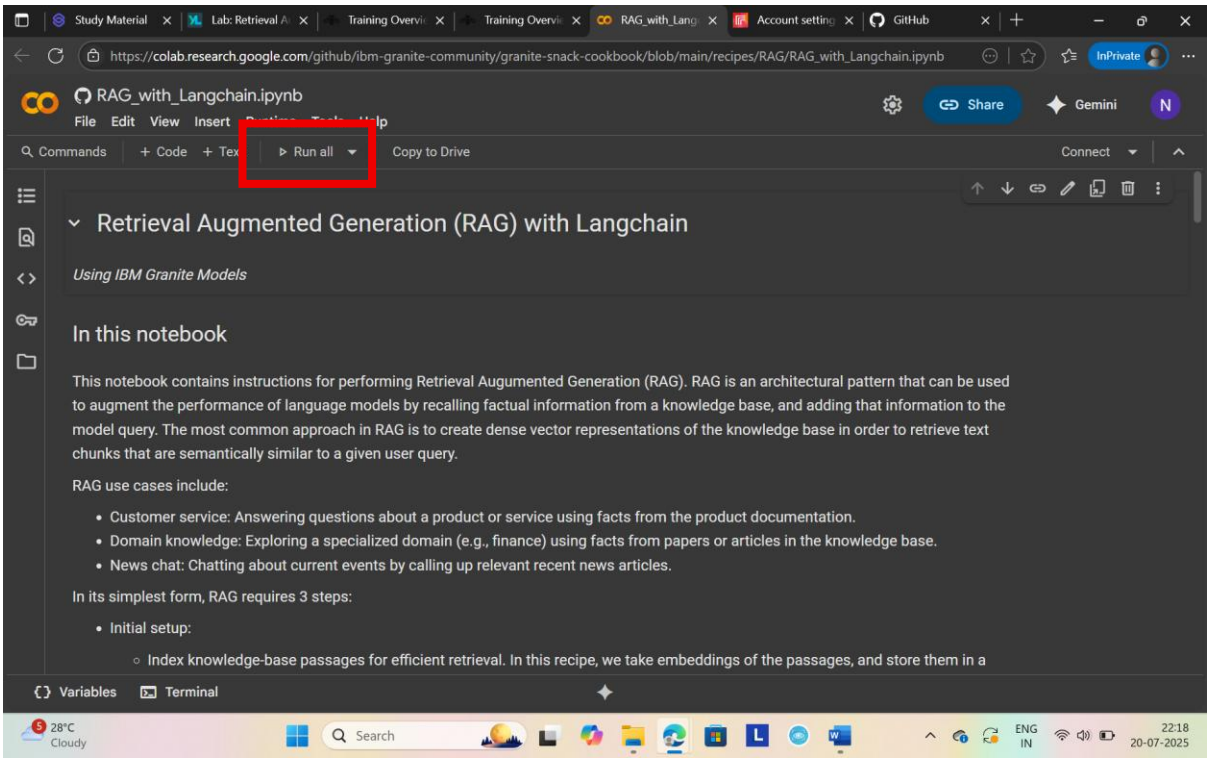
32. Come back to the Lab page, click on **Click here to begin the lab.**



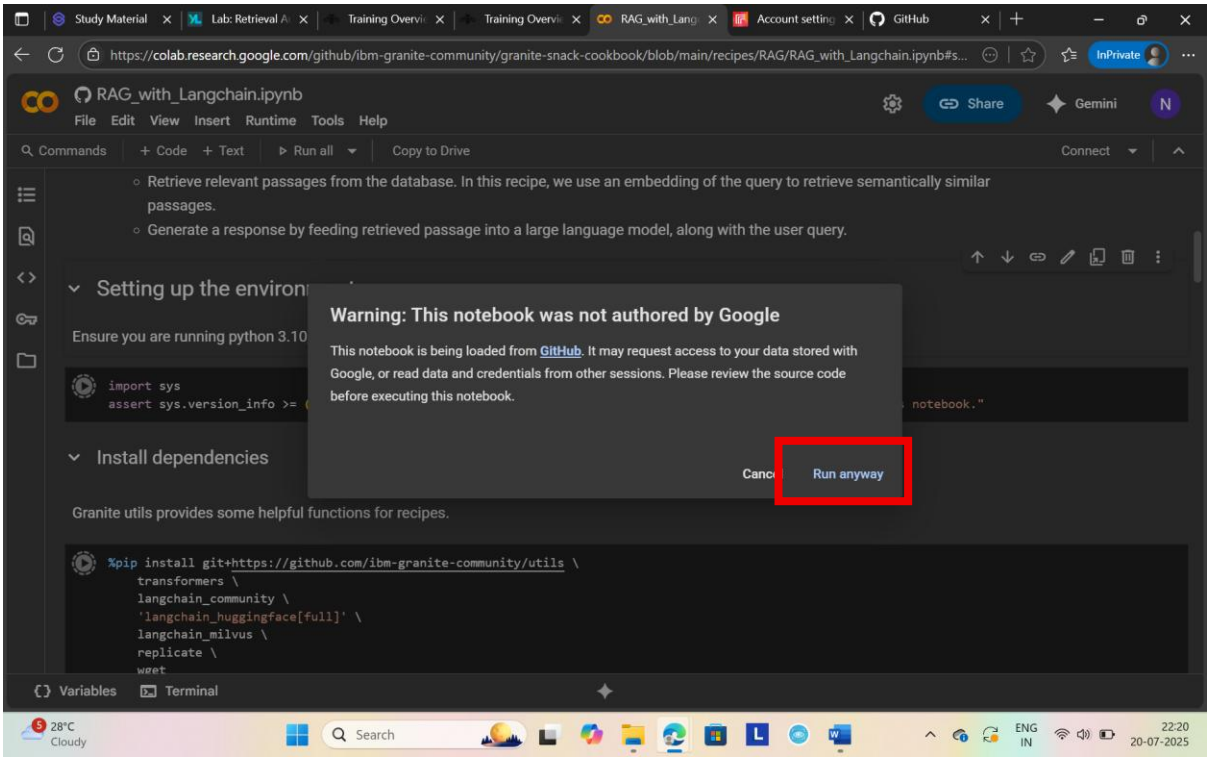
33. A Jupyter notebook with Python coding for the Lab will be opened in colab.



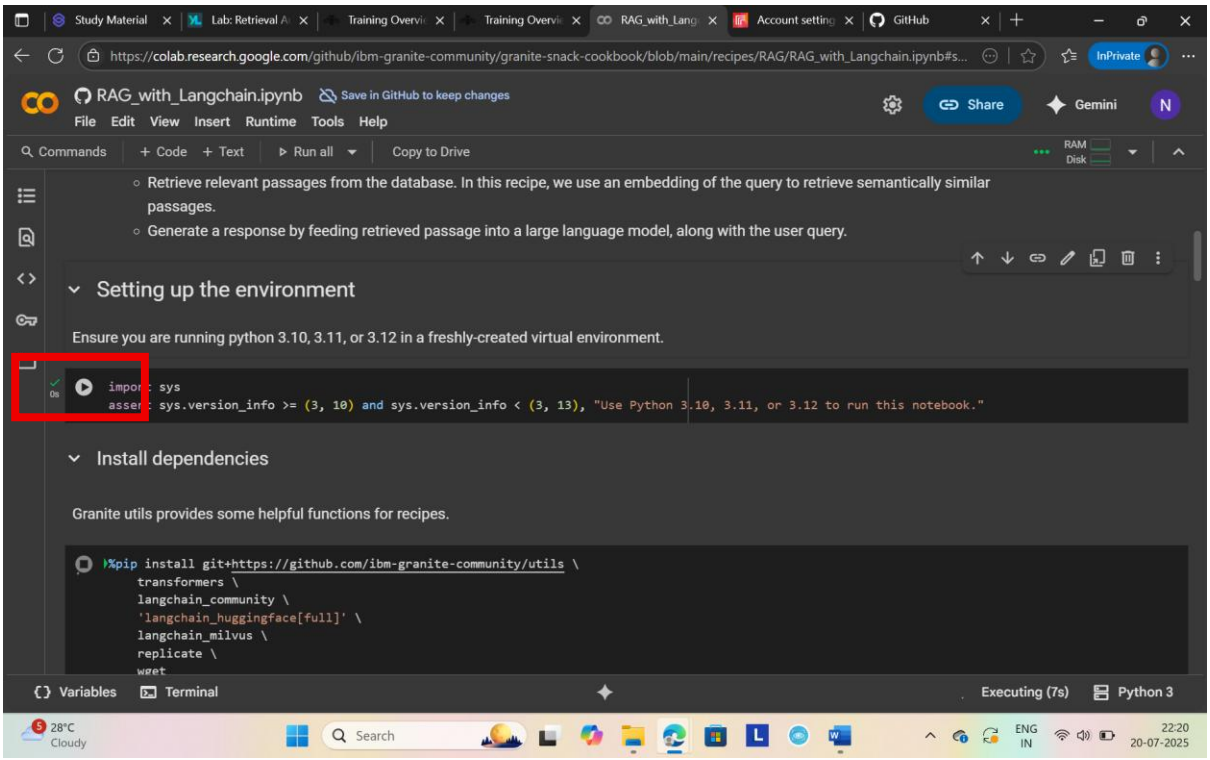
34. Find an option **Run all** and click on it.



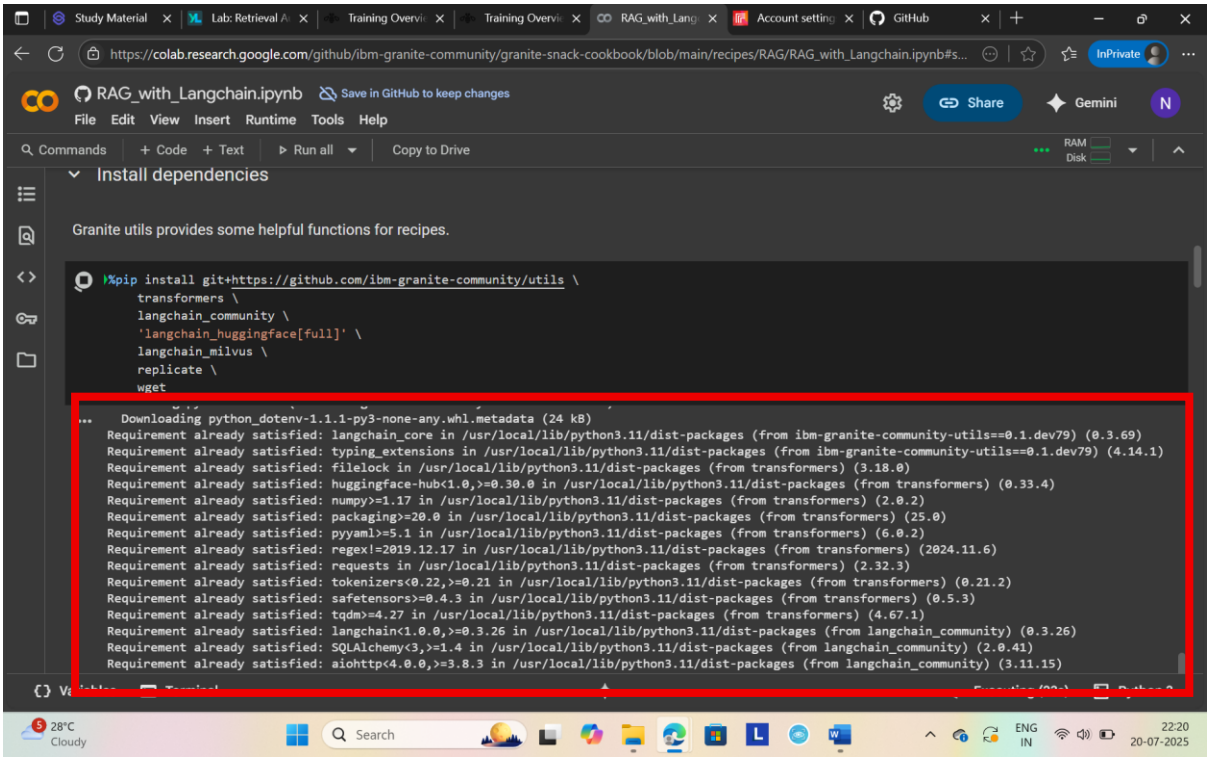
35. click on **Run anyway**



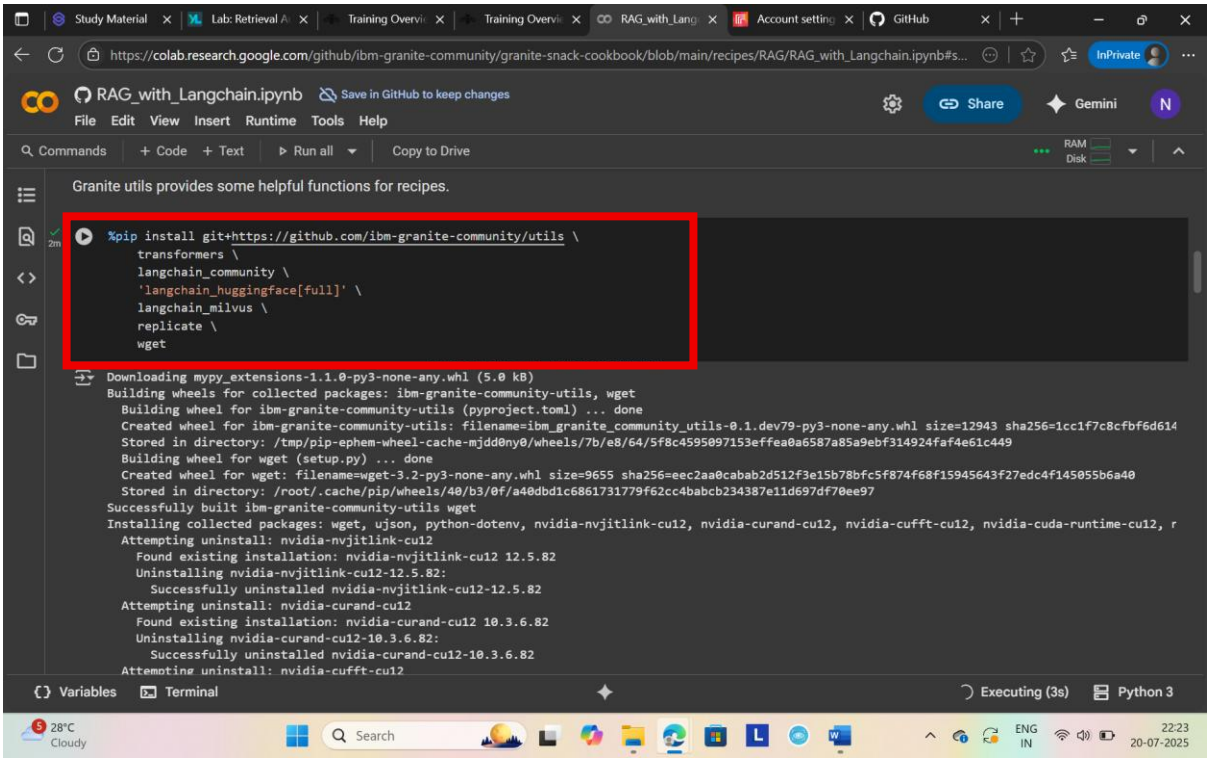
36. Your code will be start executing cell by cell.



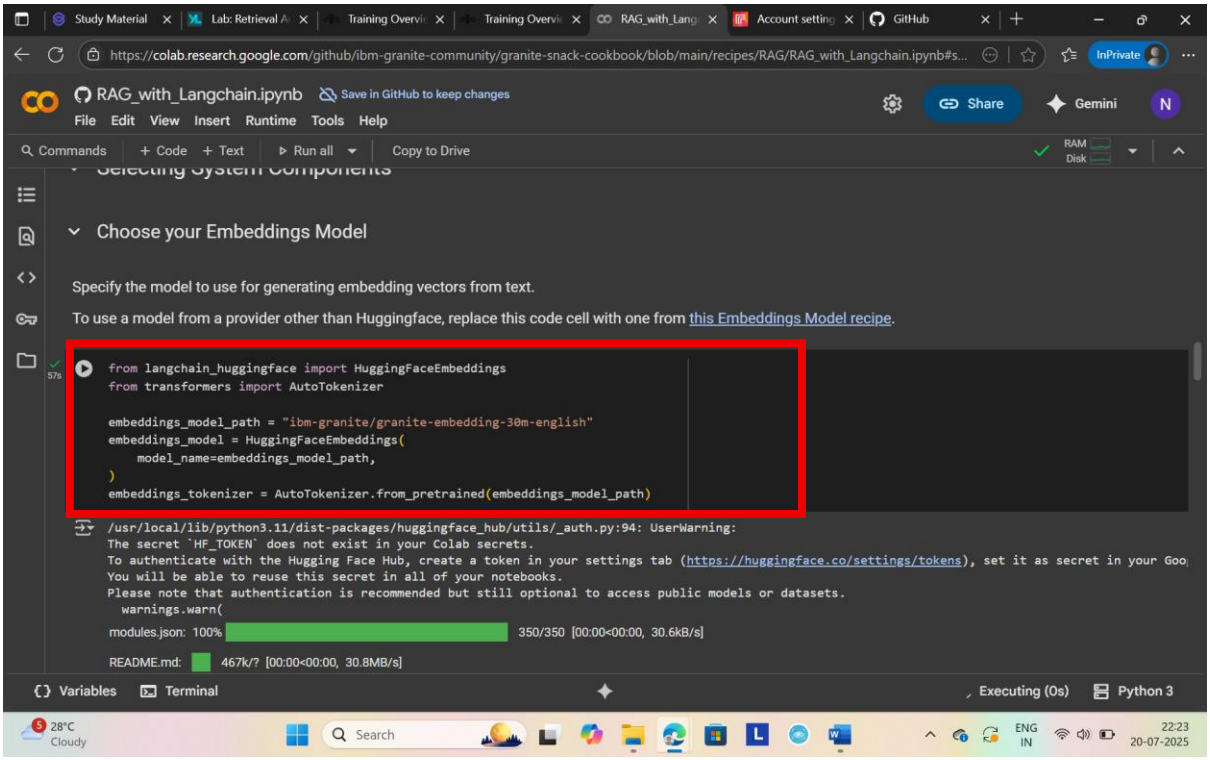
37. You can environment is setup, dependencies are installed. It will take 2-3min time.



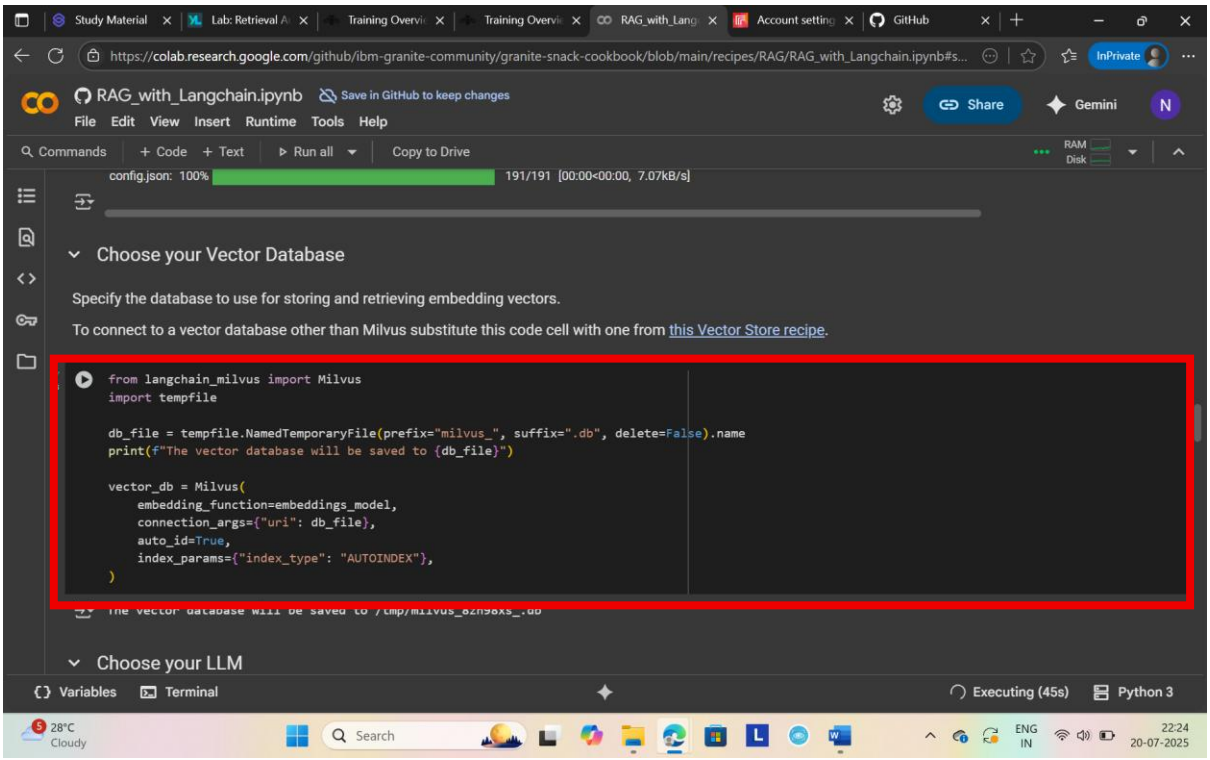
38. Executing cell by cell continues… installing all libraries, modules required for the lab activity.



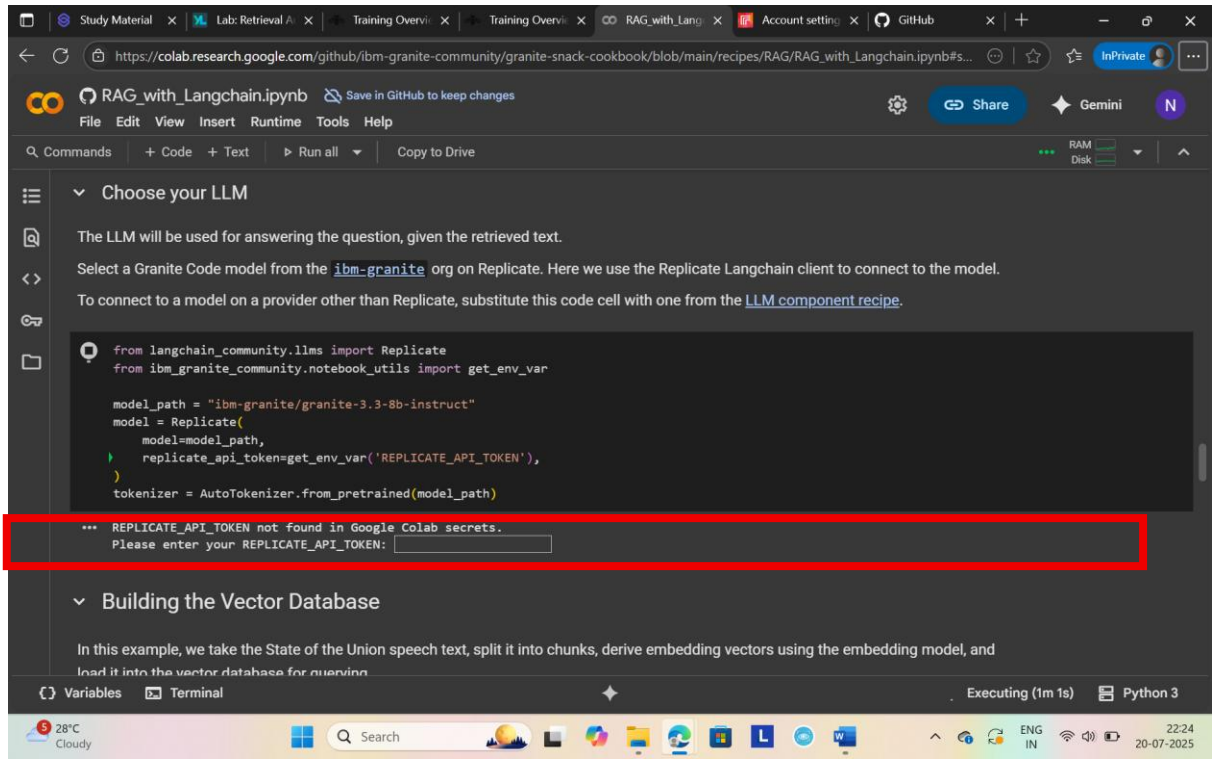
39. Here model for the lab will be chosen.



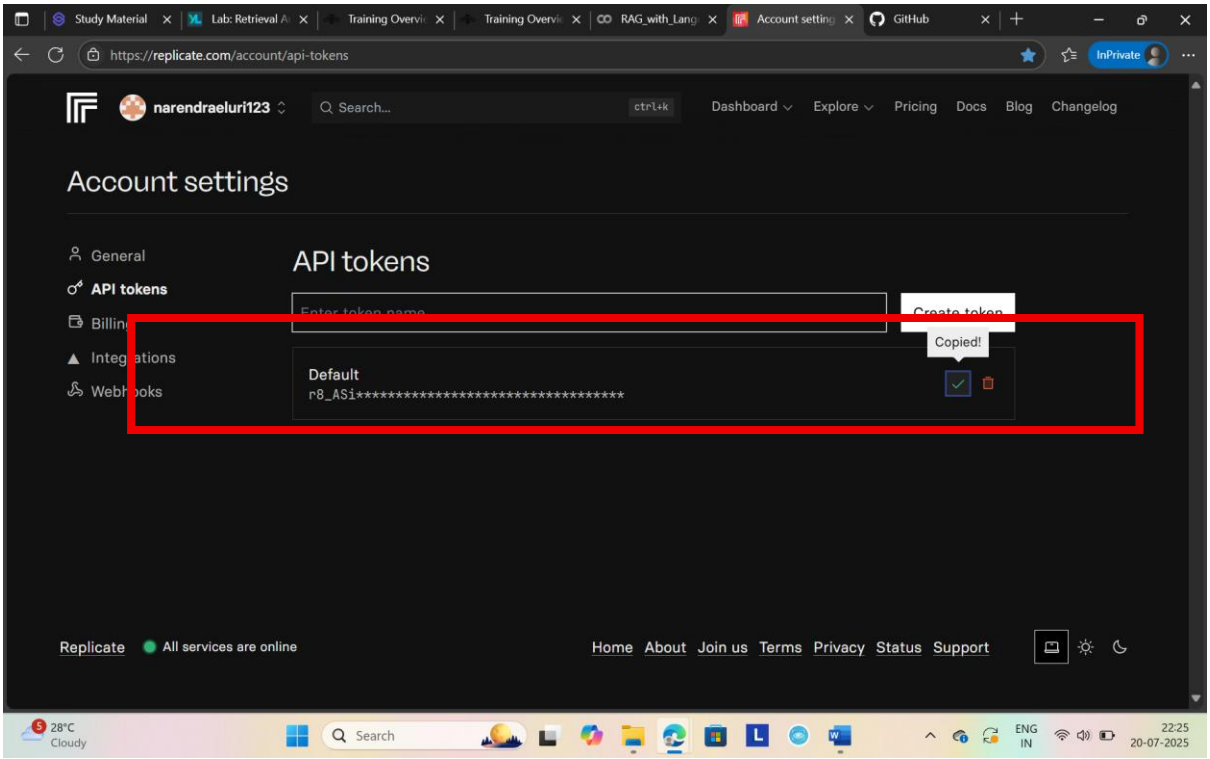
40. Vector Database to store text chunks will be created.



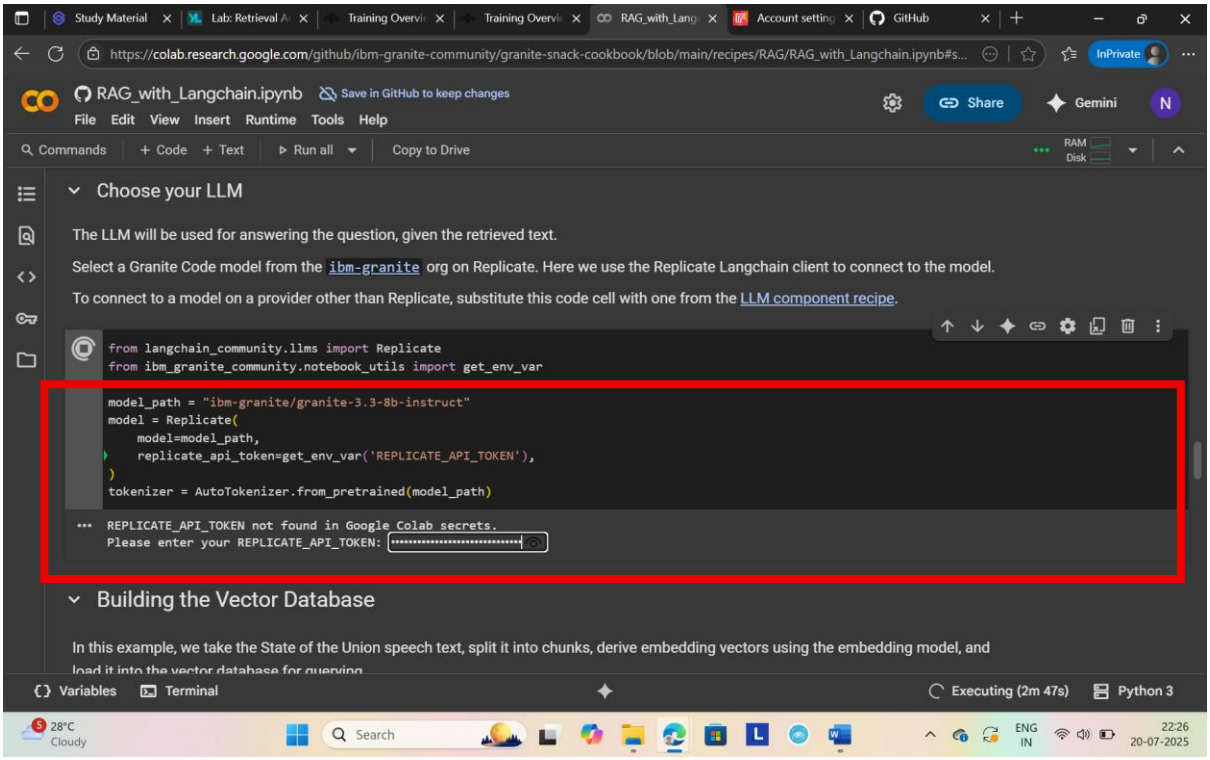
41. This code trying to load the language model from **Replicate** using the LangChain and ibm_granite_community libraries. It requires Replicate API token. Since API token is available on colab, manually it is asking you to provie the token.



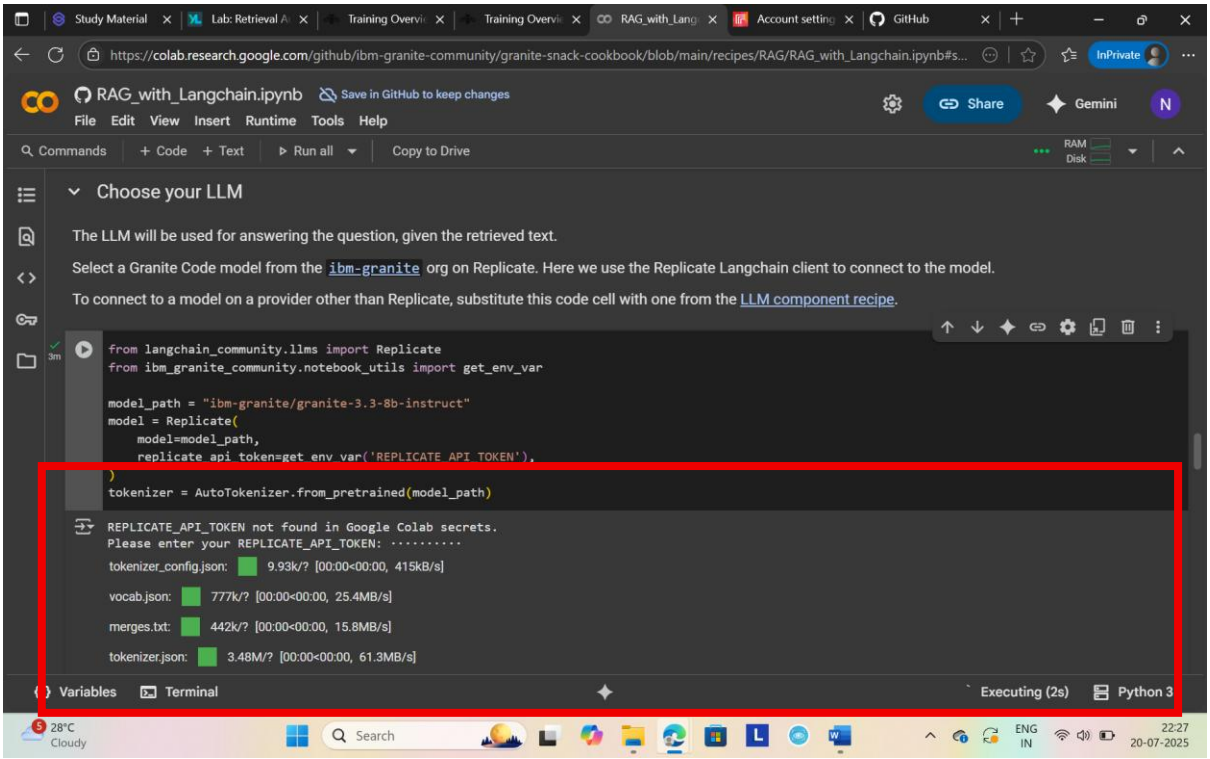
42. You can visit replicate page, copy the token.(or copy from notepad also)



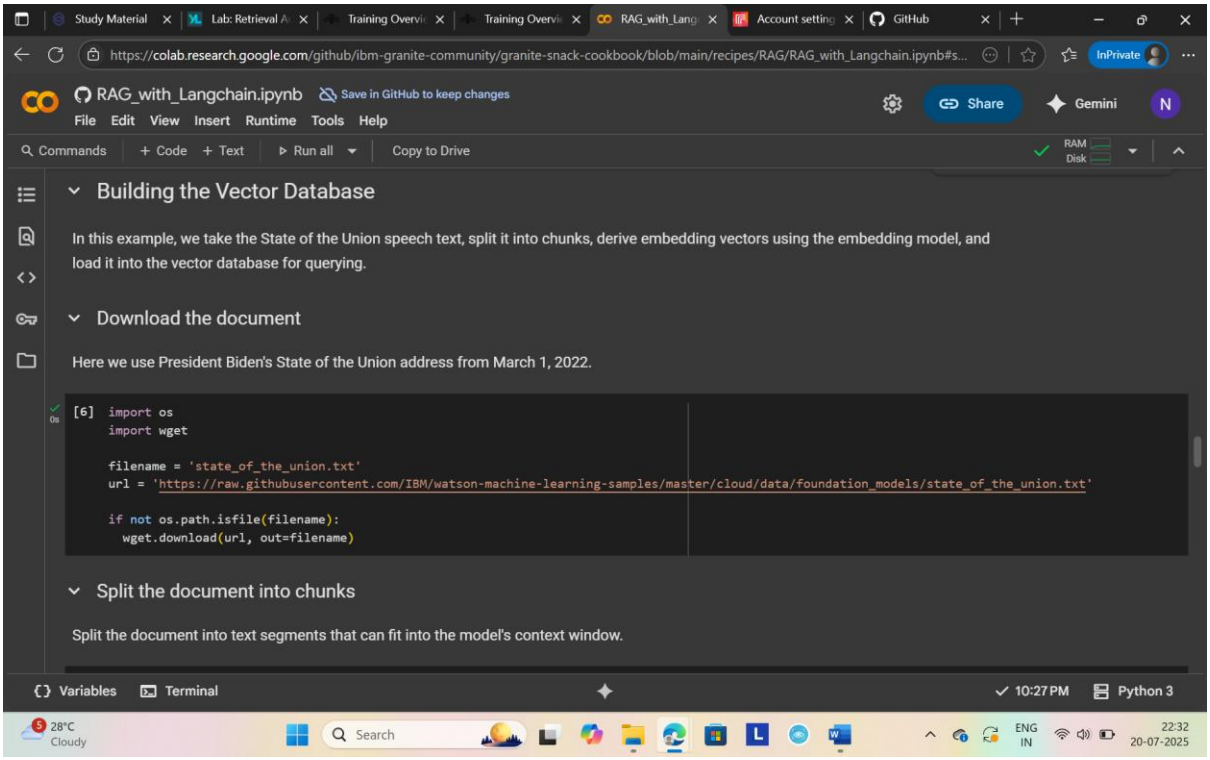
43. Replicate API token is taken, and starts creating the model using Langchain and IBM Granite community models.



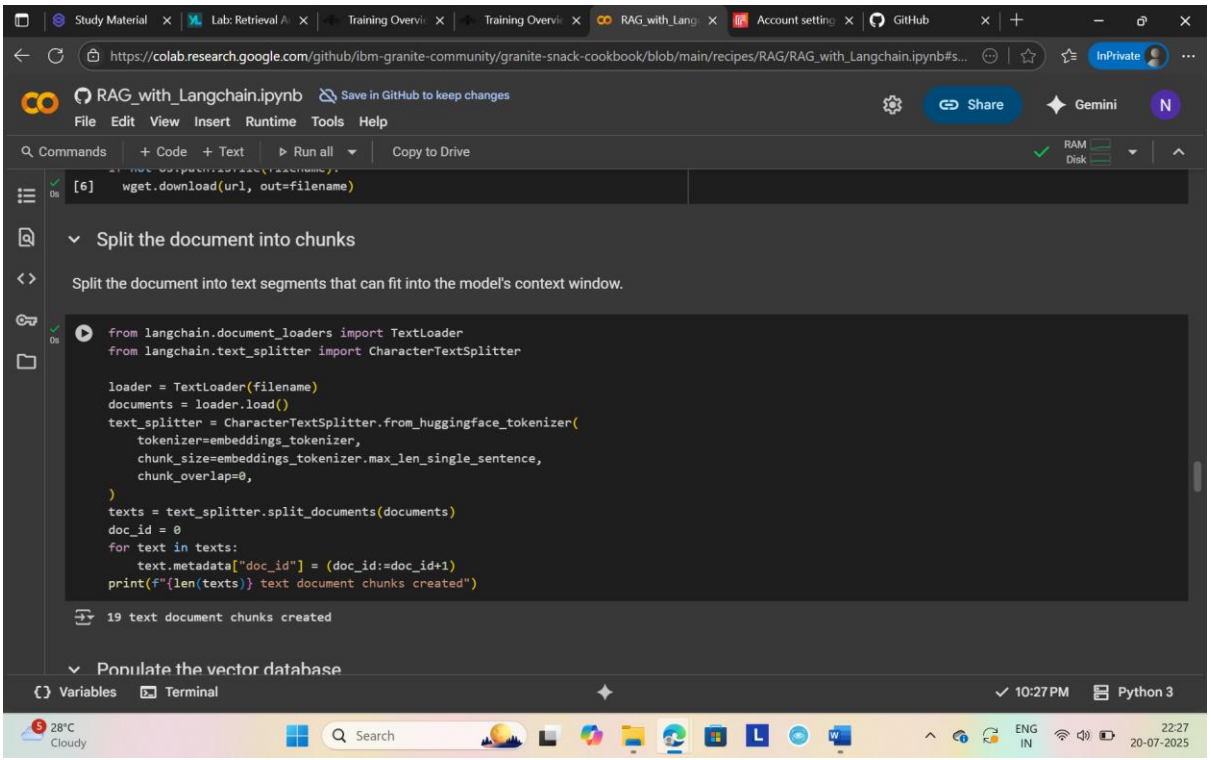
44. It starts creating model



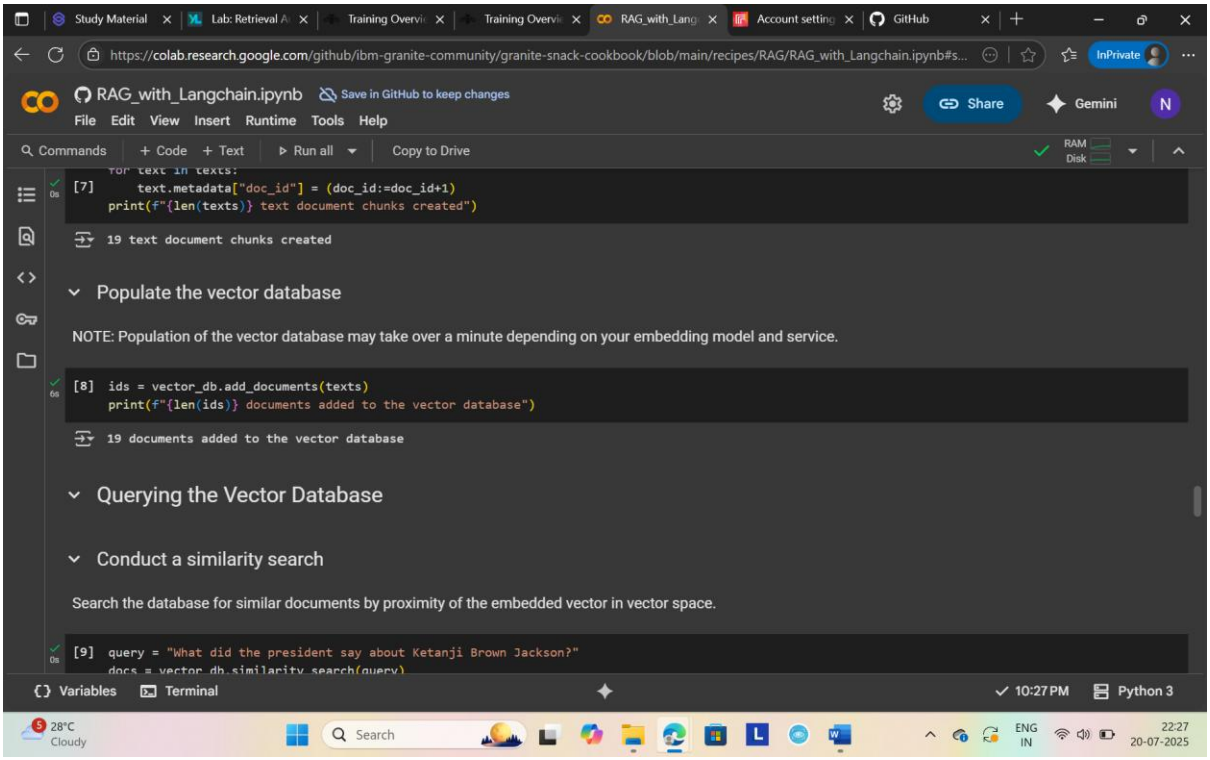
45. Starts creating vector databases



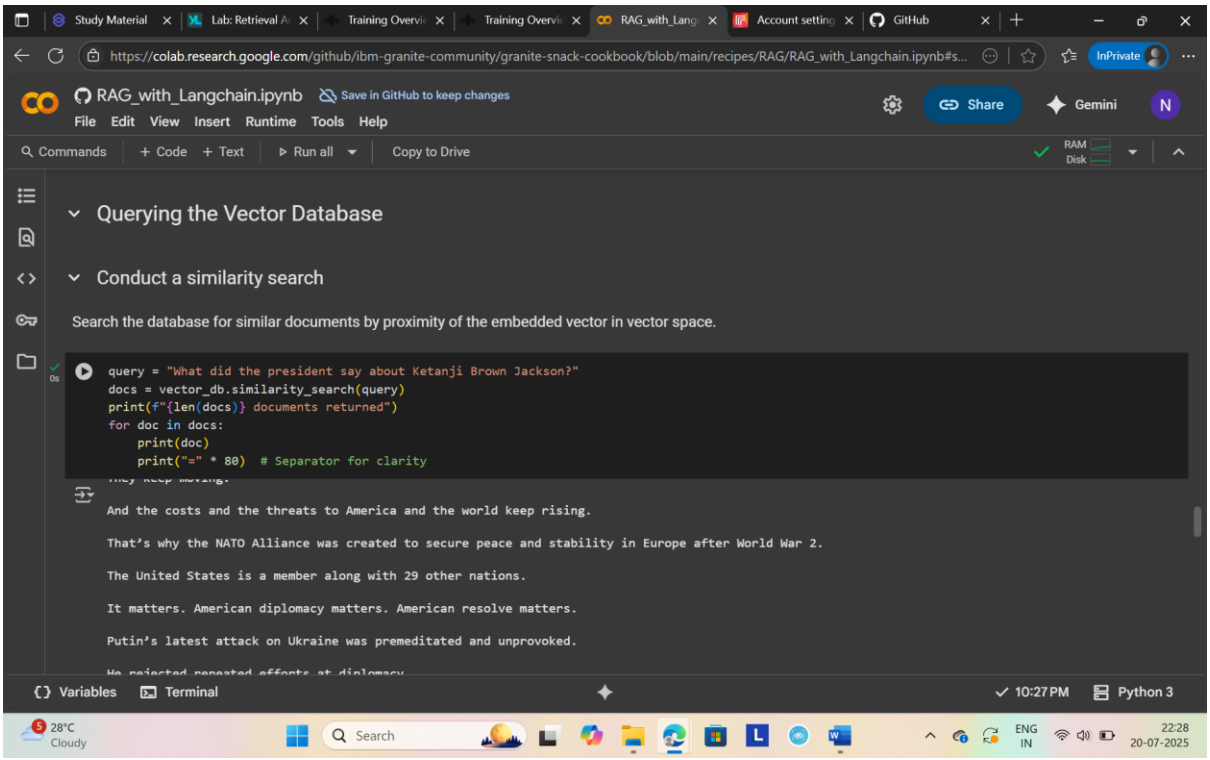
46. Splitting the text document into small chunks



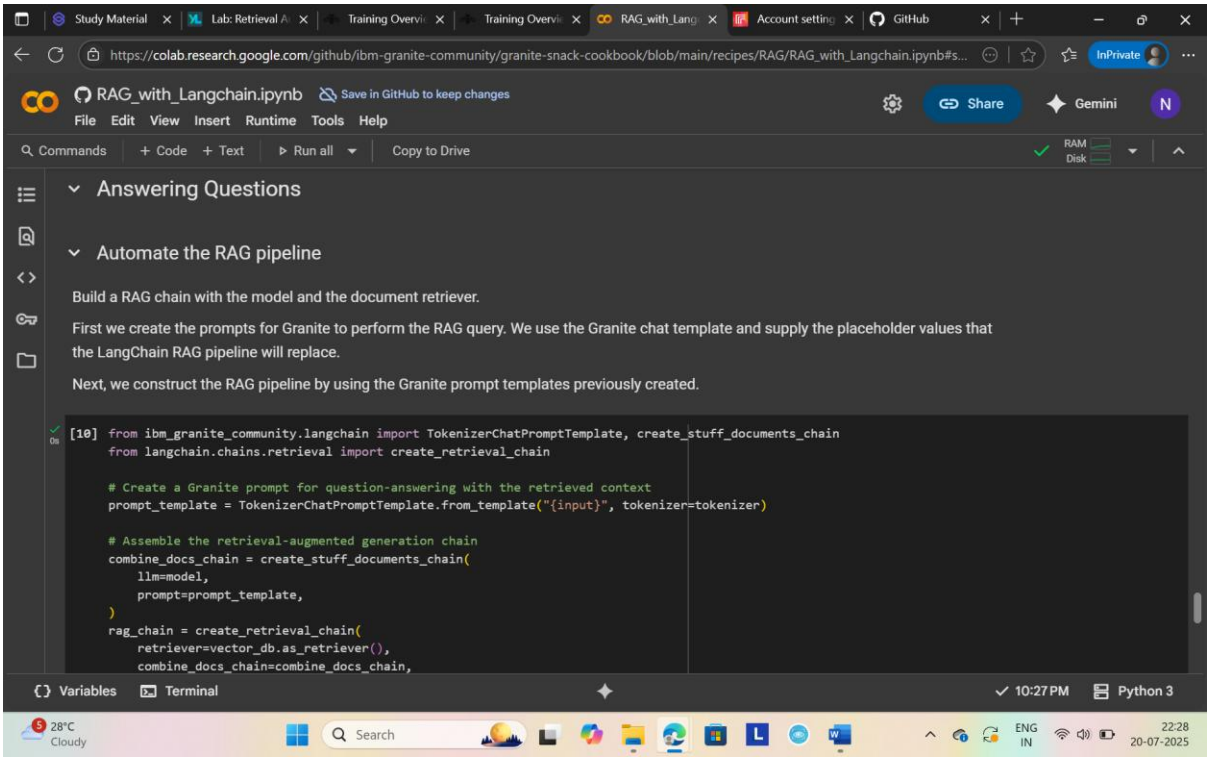
47. Populating the vector databases with the chunks



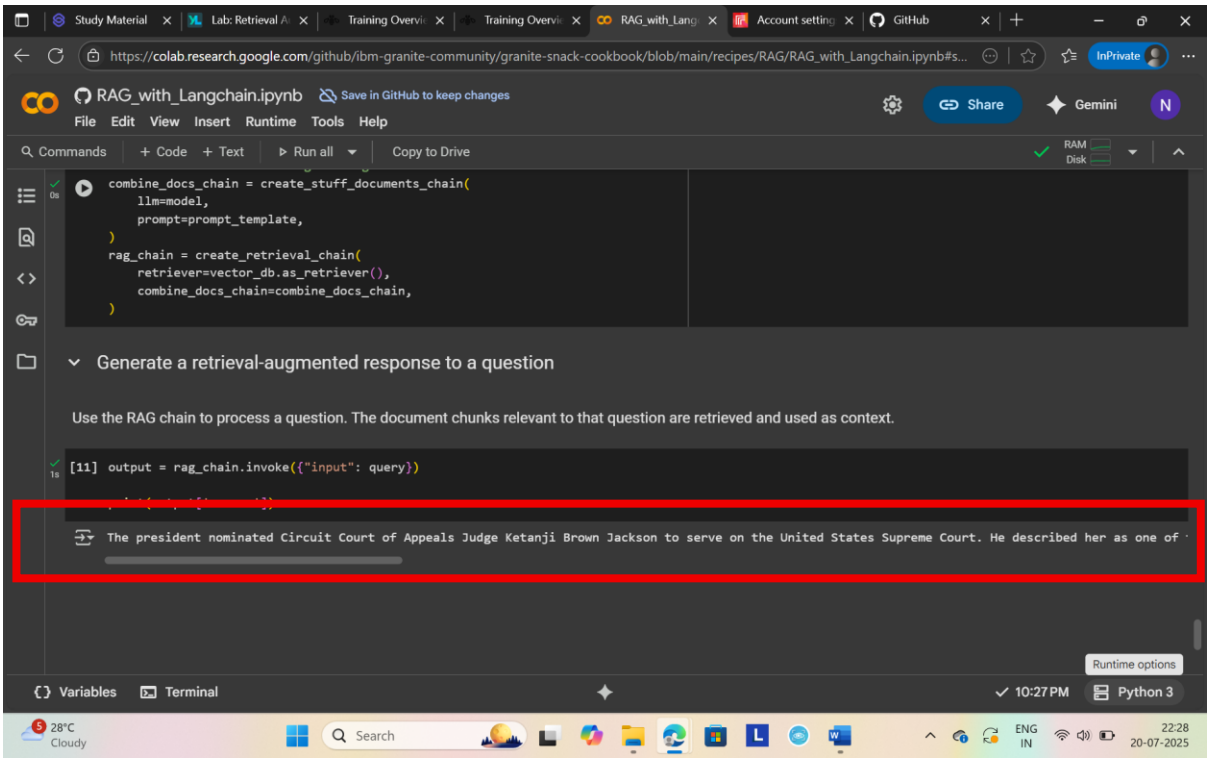
48. Performing similarity search basing on the input query will be done.



49. Prompt augmentation will be done.



50. The model will take the refined prompt and generates best suited answer from its knowledge as output.



Congratulations, you have successfully executed RAG lab with Langchain.