Video Audition Outline

Audition Information	
Title	Local Retrieval Augmented Generation in Python
Author Name	Alexander Harris
Level *(Beginner, Intermediate, Advanced)*	Intermediate

Additional Notes (optional)	I also created a Github repository for this audition. It can be seen here: https://github.com/alexanderusaf/local_rag_python

Course Planning	
Learner Prerequisites *What do they already know?*	Basic Knowledge of Python Basic Knowledge of Command Line Interface Basic Knowledge of Client and Server Interactions Basic Knowledge of Large Language Models
Storyline *Create a realistic story or scenario and explain how the learner is going to solve it.*	Imagine a software developer working on a legal research team, his name is Linus. He's working on a project and needs all the intelligence of Chat GPT but wants improved security (without having to send his data over the network) and enhanced accuracy on the highly domain specific knowledge his legal work requires. He's tried utilizing online tools to no avail. Chat GPT seems to hallucinate legal nuances when writing prompts that are specified to a particular jurisdiction or provide out-of-date information all together. Instead of waiting around for the Chat GPT developers to solve the problem for him, he can solve this problem independently by utilizing Ollama, locally served large language models, Retrieval Augmented Generation (RAG) and Python programming language. Let's take a closer look at how Linus can solve this problem.
Description *Introductory statement, general overview, and main learning point, what the learner will know by the end of the audition.*	In this audition you'll learn the fundamentals of retrieval augmented generation (RAG), how to set up a large language model served locally through Ollama, and how to configure the large language model for retrieval augmented generation using Python Programming Language.

Course Organization		
1	Audition Overview No action required. Do not edit or remove. Include a brief introduction with a title slide and overview. Briefly lay out what you will cover in this audition and what we should learn.	60 - 90 seconds
2	Slide portion This is where you present the concepts of what you are teaching. Remember "what's in it for me" and avoid weighing us down with too much text. Module Layout: *(Explain how you will be teaching the learning objective and incorporating your storyline.)* ● What is RAG ● Why use RAG, WIIFM ● Examples of real world RAG use cases	3 Min *(2-5 min)*
3	Demo This is where you will show us your teaching in practice. Remember to clean up the clutter of your screen as much as possible, and please use callouts (please do not point at objects with your cursor). Module Layout: *(Explain how you will be teaching the learning objective and incorporating your storyline.)* ● Getting the source code of the sample application ● Set up the local environment ● Generate embeddings ● Retrieval ● Prompt ● Program execution	4 Min *(3-5 min)*
	Other Portions (optional) If you want to have more slide or demo sections, add them here. Do keep in mind that the entire audition should be less than 10 minutes in length, so don't try to cover more than is needed.	
	Conclusion Provide a *brief* conclusion restating what you just covered and what we should know, and optionally what would come next.	< 1 Min