Build anything you want using, on top of, or inspired by any of the following resources – the only restriction is the submitted take-home should be a colab notebook leading through what you have built.

We expect that your take-home showcases:

* Familiarity with open-source libraries and toolkits
* Technical excellence in coding and engineering
* A thoughtful approach highlighting different aspects of the method you aim to achieve

**Details on Take Home:**

Name your submitted take home \*\*Firstname\_Lastname\_Maxim\_Takehome\*\*. Once you’ve completed the colab notebook, save and pin your revisions so the version is clear. Please make sure permissions are set so anyone with the link can view the colab.

We recommend taking 3-5 hours for this challenge. We are more interested in understanding how you think about communicating your learning and ideas rather than having an exhaustive walkthrough of your project.

Resources:

1. Optimal context:
   1. Think of ways to automate providing the necessary information from a resource to help answer relevant questions for a human.
   2. Retrieval Augmented Generation: <https://learnbybuilding.ai/tutorials/rag-from-scratch>
2. Evaluations: Build an interesting dataset on which you see [ChatGPT-4o-mini](https://chatgpt.com/) failing (Everyone has free access to the portal)
3. Tokenization: Learn more about tokenizers in the [video](https://youtu.be/zduSFxRajkE?si=CVwA-X7Wo3HCO1uQ). Play on the [website](https://tiktokenizer.vercel.app/) and find interesting tokenizer outputs that do not make sense.
4. Open-source libraries: <https://github.com/huggingface>, <https://github.com/langchain-ai/langchain>
5. N-gram model: <https://towardsdatascience.com/text-generation-using-n-gram-model-8d12d9802aa0>
6. Introduction to GPTs: [Repo](https://github.com/karpathy/nanoGPT), [Video](https://youtu.be/kCc8FmEb1nY?si=2Og6e8fTb0fkT0KY)
7. Synthetic Data Generation: <https://github.com/sdv-dev/SDV>, <https://www.turing.com/kb/synthetic-data-generation-techniques>
8. Any ideas or libraries mentioned on <https://blog.getmaxim.ai/>