

## Building a GenAI Data Pipeline with Langflow and AstraDB

Arman Sandher

03/26/25

When approaching this project, I had to watch some tutorial videos on how to use DataStax. After getting comfortable and having found a document from Langflow that depicts how to take a file, vectorize it, and answer queries based on it, I was ready to start.

Documentation used:

[Vector store RAG | Langflow Documentation](#)

I loosely based this on the documentation, which is why I decided to use OpenAI for my model. I also chose to use only the nodes in AstraDB/Langflow instead of using code. I used the Amazon API since we had used it previously in the semester. Initially, I started with trying to get the model to work before working with unsorted data. After getting an API key, I was able to get this portion working. I was able to follow the documentation. I used additional nodes to load a file, split it into chunks, load it into a collection, and use that as the context for my queries.

The **challenges** I faced were due to the collection I had initialized. I had initialized it within Langflow using the “global” icon. However, I found out later that this was not properly initialized. I had to do this within the database as a vectorized collection, otherwise, I would get a “AstraDBVectorStore: Embedding Gateway failed to process request: getCredential error, integration credential is null” error.

Also ran into an issue with which embedding I was supposed to be using. The documentation I was using was a bit dated, and a few of the nodes had combined. I was switching between two of the embedded systems, AstraVectorize or whichever other embedded model I chose to use (OpenAI in this case). After playing around with it. I ended up using the AstraVectorize model, and it began to work.

For context, my unsorted data is a paper I created with some other students earlier in the week. This paper is about delivery drones. *\*Attached in my Github\**

Here are the results:



user

can you tell me the benefits to drone delivery?



AI gpt-4o-mini

Drone delivery offers several benefits, including:

1. **Minimized Delays:** Drones can navigate directly to delivery locations, potentially reducing delivery times compared to traditional methods.
2. **Reduced Operational Costs:** By streamlining the delivery process, drones can lower costs associated with transportation and logistics.
3. **Improved Delivery Accuracy:** Drones can deliver packages to secure, pre-authorized locations, enhancing the accuracy of deliveries.
4. **Enhanced Security:** Drones can mitigate the risk of package theft (porch piracy) by delivering directly to secure locations such as smart lockers or through hand-to-hand delivery with authentication technology.
5. **Increased Consumer Trust:** With improved security and reliability, consumer confidence in online shopping can be bolstered.
6. **Environmental Benefits:** Drone delivery can significantly reduce the environmental footprint associated with last-mile delivery, potentially lowering emissions per package by up to 54% compared to gasoline-powered vehicles.
7. **Sustainable Logistics:** As cities seek greener logistics solutions, drones provide an eco-friendly alternative to traditional delivery vehicles, helping to alleviate urban congestion and reduce carbon emissions.



**user**

how much knowledge do you have about drones



**AI** gpt-4o-mini

I have knowledge about various aspects of drones, including their payload capacity, flight range, speed, altitude capabilities, navigation systems, and power sources. For instance, I know that five drones can hold up to 55 pounds, and that higher altitude positions can enhance navigation and imaging capabilities. I am aware of specific drones like the Monkey King, which is a Mach 4 supersonic drone capable of flying at twice the speed of sound and at an altitude of 12.4 miles. Additionally, I understand the importance of sensors, such as cameras and microphones, for obstacle detection and navigation, as well as the use of advanced networks like Generative Adversarial Networks and Convolutional Neural Networks for real-time navigation. Lastly, I recognize the significance of having appropriate battery and power sources for effective drone operation.



**user**

who created this paper about delivery drones?



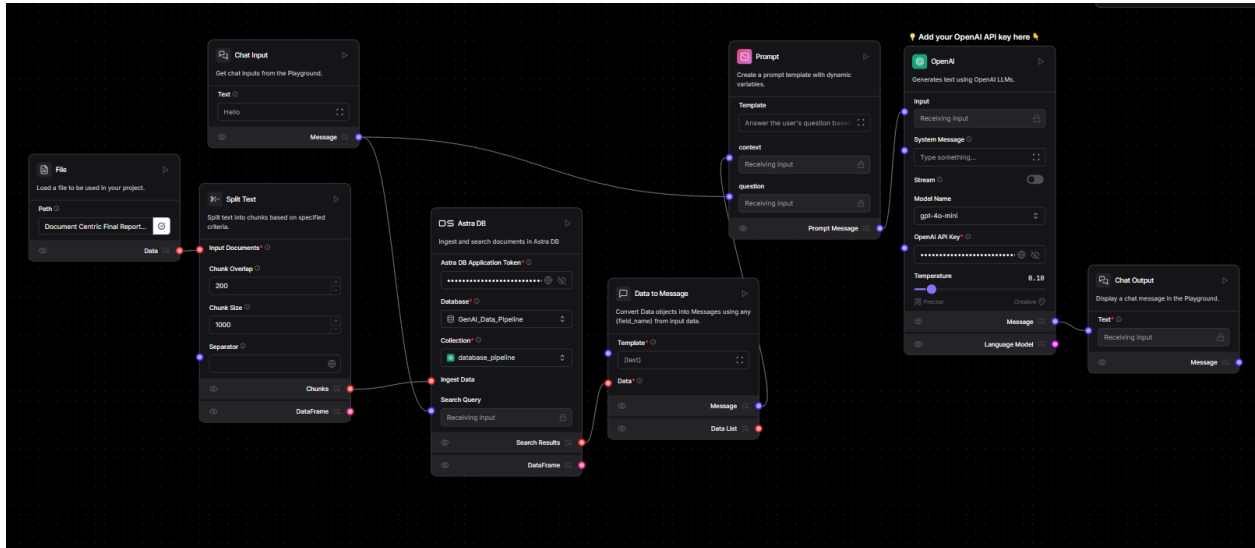
**AI** gpt-4o-mini

The paper about delivery drones was created by Justin Phan, Arman Sandher, Han Liu, and Pavithra Mahadevappa Girithimmappa, who are affiliated with the Department of Interdisciplinary Engineering at San Jose State University.

Future Improvements:

In the future, hopefully I can have it retrieve images AND text. I'll also look into real-time streaming ingestion.

The other part of this is that if it is taking information from the document, it takes quite a bit of time, improving this portion of the search would be a huge benefit.



*an enhanced picture viewable on my github.*

