Assignment 3 for Senior Design

With the commercialization of large language models and companies of varying magnitudes feeling the need to adopt them to stay relevant, the first question they ask is 'what data do I have?' and 'what can I use AI to do?'. These two questions can only be answer through an industry expert that understands the strengths and limitations of AI but there are very few people that can claim to be this. This is where we see the use for an AI agent to automate pipeline design, monitor them, and optimize them. As a CS student with experience in both AI/ML research (published two first author papers) and has had 3 coops in Full Stack where I had to create data pipelines, I understand how difficult it is to first find the relevant data within data lakes and then transform so they are usable by an AI model. My senior design project seeks to address this problem by creating an intelligent agent that reduces the barrier to building efficient, reliable, and scalable data pipelines. In doing so, I aim to contribute a tool that empowers companies to accelerate AI adoption without requiring every team to have deep in-house expertise in both data engineering and machine learning.

Through the University of Cincinnati, I have been able to learn AI from multiple classes like Soft Computing, Deep Learning, and Intelligent Systems. These classes have given me the fundamental understanding behind how an AI model works and, as such, what models are actually capable of. On the database side of this, I have taken Data Science and Database Design, which has allowed me to understand how data is structured, stored, and queried efficiently across different systems. Courses in Software Engineering and Algorithms also taught me how to break down large problems into smaller, manageable steps, which is essential for designing scalable pipelines. Beyond technical content, many of these courses required group projects and presentations, which developed my collaboration and communication skills. Together, these experiences have given me both the theoretical background and the practical teamwork foundation necessary to design AI agents that can automate the construction and optimization of data pipelines.

For my co-ops, I worked at Motz Engineering for two semesters where I focused on software development around a billing forecast database. A big part of that role involved creating a pipeline to transform raw data into usable numbers that could then be pulled into reports, which gave me my first real-world exposure to data engineering challenges. After that, I joined Lightship Capital, where I integrated AI tools like Gemini and MistralOCR with a Snowflake database provided by Nasdaq Evestment. That project taught me how to connect large language models with structured data and essentially give them the ability to "see" and interact with the database. Alongside my co-ops, I've also been doing research with Dr. Kelly Cohen at the AI Bio Lab, where I focus on explainable AI and have published two first-author papers. Through this work, I learned how to train AI systems that are not only accurate but also

transparent and trustworthy, which I believe is especially important for enterprise adoption. Altogether, these experiences shaped how I approach problems and gave me a strong mix of technical and research-driven skills to apply in my senior design project.

This project excites me because this is a large problem and the solution can actually be applied to more that data pipelines like workflows in Apple Shortcuts or Power Automate. Furthermore, data pipelines are the backbone of every industry since making sure there is a reliable flow of correct data is essential to keep the business operating, and it is important for companies that are not there yet to pursue this. This project will be able to allow these companies to try out data pipelines without having to invest as much into expensive talent. By lowering this barrier, our project could open the door for more organizations to leverage data effectively. On a personal level, the project perfectly combines my interests in automation, AI, and large-scale data systems. I see this as a chance to apply what I've learned in both class and research to a problem that matters, while also developing skills that will serve me well in my career.

My approach to this project will start with researching what tools already exist for building and managing data pipelines and seeing how AI agents can fit into that space. I want to begin by experimenting with smaller prototypes that can handle tasks like pulling in data, cleaning it, and loading it into a simple database. From there, the plan is to build up toward an agent that can generate and manage more complex pipelines on its own. The goal is to show that this system can save time and reduce the amount of manual work that usually goes into setting things up. I will know I have done a good job if the agent can handle real data without constant human intervention and if the pipelines it builds actually work the way they are supposed to. On a personal level, I will measure my contribution by how much I am able to push the project forward with both technical ideas and teamwork.