# GM AI and Cybersecurity Ecosystem Project Update Report

### 1. Project Overview

This project aims to develop an **AI-powered company locator dashboard** using **Streamlit**. The dashboard will allow users to **search for AI & cybersecurity companies**, apply filters, and **visualize their locations on an interactive map**.

The goal is to provide valuable insights for **policymakers, businesses, researchers, investors, and job seekers**. The project currently relies on **CSV-based data storage**, with plans to integrate a database in the future.

### 2. Work Completed in Weeks 1 & 2

#### 2.1 Understanding Data City Platform

* Researched **Data City** and its capabilities.
* Studied **available data sources and APIs** to integrate relevant insights.

#### 2.2 Defining Users & Use Cases

| **User Type** | **Needs / Use Cases** |
| --- | --- |
| **Policymakers** | Identify industry trends, investment needs, and policy impacts. |
| **Businesses** | Find potential partners, investors, and market opportunities. |
| **Researchers** | Study AI & cybersecurity clusters, analyze trends, and predict growth. |
| **Investors** | Discover funding opportunities, track emerging startups. |
| **Job Seekers** | Identify companies hiring in the AI & cybersecurity sector. |

#### 2.3 Finalizing Data Sources

To provide **comprehensive AI & cybersecurity insights**, we have selected the following data sources:  
✅ **Companies House** – Provides details on all registered UK companies, including financial filings and company status.  
✅ **Crunchbase** – Funding data and startup insights.  
✅ **TechCrunch, Statista, Tech Manchester** – Web scraping for tech industry trends.  
✅ **LinkedIn, Twitter** – Social media data for company activity and hiring trends.

#### 2.4 Tech Stack Decision

* **Backend:** Python
* **Frontend:** Streamlit
* **Data Processing:** Pandas, Geopy, Folium

Initially, we planned to use **Django (backend) and React (frontend)**. However, due to:  
1️ **Time constraints** – Faster development is needed.  
2️ **Focus on AI & ML features** rather than app structure.  
3️ **Scalability considerations** – Can migrate to Django & React for a production-level application later.

#### 2.5 ETL Process & Data Handling

* **ETL will be a separate script** scheduled to run at night.
* This prevents **slowdowns** in the Streamlit app.
* **Option to update data in real-time** if needed in the future.

#### 2.6 MVP Development

✅ Built a **basic prototype** with:

* **Company locator map** for Greater Manchester
* **Search bar** with filtering options

✅ **Data Extraction & Processing**

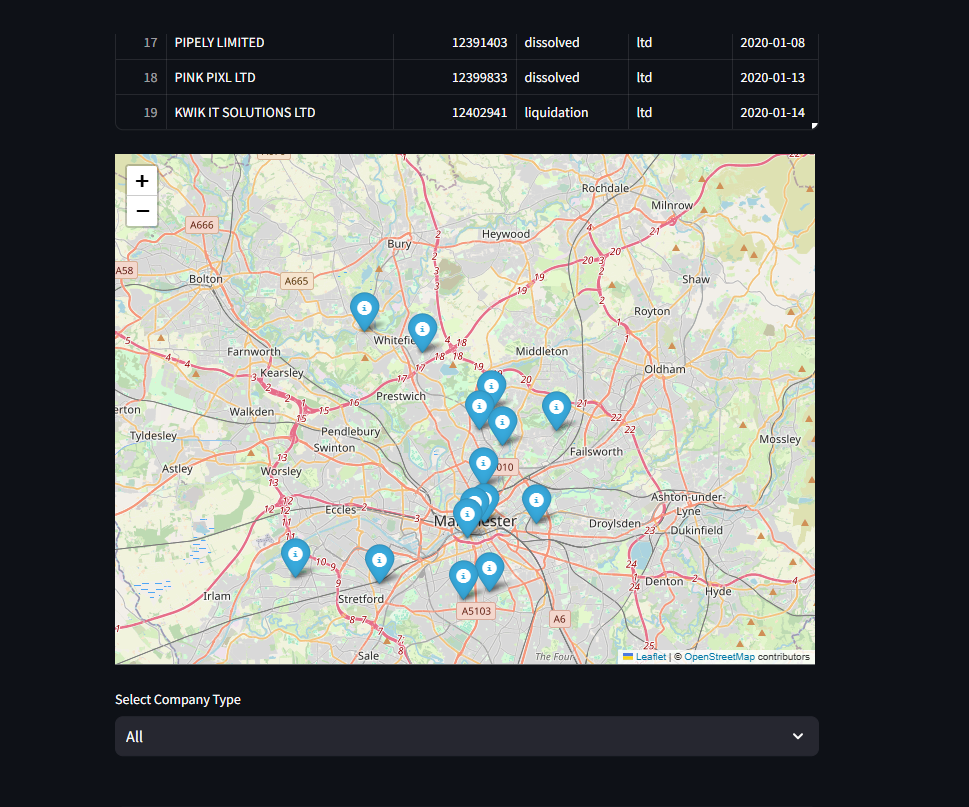
* Extracted **company data from Companies House API**
* Used **Geopy** to fetch **latitude & longitude** for mapping
* Stored extracted data in **CSV format** (no database integration yet)

✅ **Visualization**

* Used **Folium** to **plot companies on a map**
* Displayed company details in interactive popups

A screenshot of a computer

AI-generated content may be incorrect.



Github link : <https://github.com/Isha-2001/GM_Ecosystem>

### 3. Challenges & Optimization

| **Challenge** | **Solution Implemented** |
| --- | --- |
| **Geopy is slow for geocoding** | Exploring **caching methods** to reuse previous results. |
| **API Rate Limits** | Added **time.sleep()** to prevent throttling. |
| **Large dataset slows processing** | Planning to **limit data size** for local testing. |
| **No database yet** | Using CSV files for now, with plans to integrate **PostgreSQL or Firebase**. |

### 4. Next Steps

🔹 **Improve Search & Filtering**

* Implement advanced filters (e.g., **company status, incorporation date**).

🔹 **Optimize Geolocation Processing**

* Implement **batch processing** for latitude/longitude extraction.
* Store geocoding results to **avoid redundant API calls**.

🔹 **Database Integration**

* Transition from CSV-based storage to **a database** (SQL/NOSQL).

### 5. Conclusion

🚀 **Key Milestones Achieved**  
✅ Data sources finalized  
✅ ETL pipeline structured  
✅ MVP with **search & map visualization**  
✅ Optimized decision on **tech stack**

🔍 **Challenges to Solve**  
⚡ **Improve geolocation speed**  
⚡ **Enhance search & filtering and add more features and analysis**   
⚡ **Plan future database integration**

📌 **Seeking Guidance On:**

* Optimizing **geolocation data handling**.
* Transitioning from CSV to **database integration**.