Report: Company Intelligence Agent

Supratim Saha & Shaira Manandhar
Beloit College, Dec 9th, 2025

1. Introduction

1.1 Company Analyzed

The public company selected for this project is **EssilorLuxottica**, the global leader in eyewear products and eye care services. This project develops a company-specific AI system that can read corporate documents, search within them, and answer questions using only the information contained in those sources. In effect, the system functions like a custom-built version of ChatGPT, restricted to knowledge about a single company. Rather than training a new language model from scratch, this project implements a structured retrieval and reasoning pipeline that organizes EssilorLuxottica's documents and allows a large language model (LLM) to generate answers strictly supported by the retrieved text as evidence.

1.2 Document Collection

To build the knowledge base, the following documents were collected:

- Annual report excerpt
- ESG / sustainability report excerpt
- External news summary
- FactSet financial data
- FactSet ESG data

1.3 Purpose of the Assistant

The assistant is designed to answer company-specific questions using only the documents in the knowledge base. All answers strictly rely on the provided text to generate responses. It answers the following types of questions:

Company Snapshot: What does the company do? What are its main business segments? How does it generate revenue?
Risk Factors: What major risks does the company identify?
Environmental, Social, and Governance (ESG) Priorities: What environmental or social goals does the company describe? What sustainability targets does it mention?
Custom Questions: Any additional questions entered by the user

2. Data Preparation & Chunking

2.1 Document Sources

To create a robust knowledge base, several document types were collected to cover the required business areas:

Source File	Page/ Size	Key Sections Covered	Purpose
annual_report_excerpt.txt	14 pages	Business description, Revenue drivers, Risk factors	Company Snapshot
esg_excerpt.txt	10 pages	Environmental goals, Social priorities	ESG Analysis
external_summary.txt	2 pages	Market positioning, Competition	External Validation & Context
factset_esg.txt	9 pages	Emissions, Board diversity, Targets	Structured KPI Extraction +

			Additional ESG Information
factset_financials.txt	2 pages	Revenue, Growth, Business segment breakdown	Structured KPI Extraction + Additional ESG Information

2.2 Text Extraction

The original reports were provided primarily in PDF format. Each document was converted into plain text using an online PDF-to-TXT conversion tool to enable programmatic processing. This conversion removed visual layout elements such as tables, columns, and graphics while preserving the readable textual content required for downstream analysis. Minor challenges included inconsistent formatting and fragmented text caused by the original PDF structure.

2.3 Cleaning and Preprocessing

Document noises can severely degrade the performance of subsequent retrieval and LLM stages. A customized cleaning pipeline was implemented for the documents. This process yielded five cleaned files, which served as the input for chunking. The following cleaning steps were taken:

- Removed Page Numbers & Markers: Regular expressions were used to strip standalone numbers, and common patterns like "Page 3," "3 of 9," and "4/12."
- Removed Repeating Boilerplate: Specific text from report footers (e.g., © 2025 FactSet Research Systems Inc.) and repeating headers (e.g., EssilorLuxottica 2024 Annual Report) were targeted and removed.
- Removed PDF Artifacts: Lines containing only broken characters or extreme letter-spacing (e.g., E S S I L O R L U X O T T I C A) were eliminated to prevent noise.
- *Normalized Whitespace:* All excessive newlines, and spaces were standardized into a single space, ensuring all text was in a consistent encoding (e.g., UTF-8)

2.4 Chunking Process

After text cleaning, all documents were segmented into smaller sections using a custom Python-based chunking function, since a single LLM cannot process an entire long document at once,. The chunking logic targets a size range of 500–1,000 words per chunk to preserve meaningful context while remaining within the practical input limits of an LLM.

The algorithm first computes an optimal set of chunk sizes for each document based on total word count. Larger documents are split into near-maximum-sized chunks (1,000 words), while any remaining text is either kept as a single chunk or evenly divided to prevent excessively long or short segments. This ensures consistent chunk sizes and balanced information density.

All chunks were compiled into a structured DataFrame and saved as a CSV file *(essilor_chunks_clean.csv)* for downstream retrieval. In total, 17 coherent chunks were generated across all documents. Each chunk is stored with the following metadata:

- chunk_id
- company name
- source file
- chunk text
- word count

Figure 1. Chunks CSV Table

essilor_chunks_clean

chunk_id	company	source_file	chunk_text	word_count
0	EssilorLuxottica	factset_esg_clean.txt	ESG Report - EssilorLuxottica SA €306.00 ESG Overview ESG Snapshot Emissions & Targets Board Diversity Suppliers 33.7091 No 40% 2/97 Total GHG Emissions / Lacks Paris-Aligned Percent of female board Exposure to	544
1	EssilorLuxottica	factset_esg_clean.txt	1+2+3 Carbon Intensity Current Year Temperature 2030 Temperature 2050 Temperature Company Name Ticker Emissions 1 2 3 Emissions (EVIC) (EVIC) (EVIC) Alignment Alignment Alignment EssilorLuxottica SA EL-FR Warb	544
0	EssilorLuxottica	factset_financials_clean.txt	EssilorLuxottica SA (EL-FR) Entity Type: Public Company Currency: LOCAL Report as of: 23 Nov '25 Closed: Price | Nov 21, 2025 17:55 Exch €306.00 -5.20 Fully Diluted Mkt Cap as of Nov 21, 2025 €143.27B Performance T	561
1	EssilorLuxottica	factset_financials_clean.txt	29.78 P/FCF 25.09 29.35 40.48 35.19 EssilorLuxottica SA (EL-FR) Entity Type: Public Company Currency: LOCAL Report as of: 23 Nov '25 Ownership Profile Recent Recent EssilorLuxottica SA engages in the design, manufa	561
0	EssilorLuxottica	esg_excerpt_clean.txt	Sustainability Report General Disclosures 1 General Disclosures Sustainability at EssilorLuxottica Doing good for its employees, customers, consumers, and communities in outreach initiatives, its Eyes on the Planet commur	1000
1	EssilorLuxottica	esg_excerpt_clean.txt	into its business model. By advancing its Eyes on • Eyes on World Sight: Based on its belief that good vision is a the Planet program, EssilorLuxottica contributes to its Mission basic human right, the Group has an ambition t	1000
2	EssilorLuxottica	esg_excerpt_clean.txt	Company's multi-annual strategic include the assessment of company and executive guidelines on social and environmental responsibility. performance, including the definition of the sustainability criterion for the long-term i	1000
3	EssilorLuxottica	esg_excerpt_clean.txt	timelines to ensure a consistent and reliable data and consolidation process. Besides, the open As part of its commitment to mitigate environmental, social and communication line with the different functions involved allows	1000
4	EssilorLuxottica	esg_excerpt_clean.txt	companies within the scope of the Corporate Sustainability 4. Validation of the final list of material matters and reporting. Reporting Directive (CSRD) to report on sustainability matters based on a Double Materiality Assessm	1000
5	EssilorLuxottica	esg_excerpt_clean.txt	benchmarking analysis of sustainability reports from 16 peers • scale: how severe are the (potential) negative and positive in the healthcare, fashion and retail industries; impacts; • analysis and mapping of the ESG risks ident	674
0	EssilorLuxottica	external_summary_clean.txt	Essilor Stellest is the first and only FDA market authorized spectacle lens in the United States EssilorLuxottica's lens is clinically proven to slow myopia progression in children1,2 Paris, France (25 September 2025) – EssilorLu	838
0	EssilorLuxottica	annual_report_excerpt_clean.txt	EssilorLuxottica 2024 Annual Report €26.5 bn Seeing At EssilorLuxottica, we see the world differently. As a global leader in vision care innovation, we are revolutionizing the way people see in revenue and experience the wor	1000
1	EssilorLuxottica	annual_report_excerpt_clean.txt	place in the Dow are creating a transformative platform that reimagines the eyes as Jones Best-in-Class Europe Index. Sustainability remains key to our a gateway to new possibilities – the most seamless and immediate Cor	1000
2	EssilorLuxottica	annual_report_excerpt_clean.txt	a future where human potential is empowered through better vision and hearing – for all. j j j j A Board driving impact and innovation Séjb+astjienj +Brojwn Marie-Christine José Gonzalo Virginie Mario Notari 7 9 Director repre	1000
3	EssilorLuxottica	annual_report_excerpt_clean.txt	pioneering In 2021, the acquisition of GrandVision business model. Rooted in the visionary completed the visionary project of Underpinning our vision for the future is a platform of strategic decision of Leonardo Del Vecchio t Politecnico di Milano. From eye-tracking to camera and sensor and optical integration, the Lab focuses on developing the next generation of smart glasses integrating hardware – such as electronics, sensors and	1000
4	EssilorLuxottica	annual_report_excerpt_clean.txt	processors – alongside algorithms and software featuring advanced signal and data processing, AI and machine learning. The Lab works as part of the Group's global R&D platform, in close synergy with our hubs in Boston,	655
5	EssilorLuxottica	annual_report_excerpt_clean.txt	the ever- evolving needs of consumers and fashion partners worldwide with passion, purpose and unparalleled excellence. With direct oversight of every step in the eyewear production – from raw materials to finished eyegla	655

Distribution of the 17 chunks by source file shows that the ESG report and annual report each contributed 6 chunks. The FactSet ESG file contributed 2 chunks, the FactSet financial file contributed 2 chunks, and the external summary file contributed 1 chunk. This distribution reflects balanced coverage across financial, ESG, and external market sources.

3. Retrieval System

3.1 Retrieval Method

The retrieval system operates on the cleaned and chunked dataset stored in *essilor_chunks_clean.csv*, which contains 17 document chunks. These chunks are loaded into a structured document store using a custom function (*build_document_store*), where each record includes the chunk ID, text, source file, document type, and year. TF–IDF artifacts are then generated using the function *build_tfidf_artifacts()*. This process converts the 17 document chunks into a TF–IDF matrix with a shape of (17, 10,609), where each column represents a unique term, and each row represents a document chunk.

To retrieve relevant information, the function *retrieve_chunks(query, k, allowed_doc_types)* is used. For a given user query, the query is compared against all stored chunk vectors. A similarity score is then computed for each chunk. The chunks are ranked by similarity. Finally, the top *k* chunks (k = 5) are returned as the most relevant results. Only the highest-ranking, filtered chunks are passed to the LLM as supporting evidence for answer generation. The retrieval process also supports document-type filtering using the *allowed_doc_types* parameter. For example:

- ESG-related questions restrict retrieval to ESG chunks
- Financial questions restrict retrieval to financial and annual report chunks

3.2 Retrieval Examples

To evaluate the performance of the retrieval system, multiple test questions were submitted to the TF–IDF–based retrieval module. For each query, the system returned the top five most relevant chunks based on cosine similarity. Two representative examples, one ESG-focused and one financial, are presented below.

Figure 2. ESG-Focused Question

```python
from retrieval import retrieve_chunks

# ESG-focused question
q1 = "What are EssilorLuxottica's total greenhouse gas emissions and Scope 1, 2, and 3 breakdown?"
res1 = retrieve_chunks(q1, k=5, allowed_doc_types=["esg"])
res1[["chunk_id", "source", "doc_type", "similarity"]]
```

Python

	chunk_id	source	doc_type	similarity
0	0	factset_esg_clean.txt	esg	0.113095
1	1	factset_esg_clean.txt	esg	0.041232
2	2	esg_excerpt_clean.txt	esg	0.040445
3	5	esg_excerpt_clean.txt	esg	0.019436
4	4	esg_excerpt_clean.txt	esg	0.010889

All retrieved chunks originate from ESG-specific documents, which are the primary sources containing emissions data. These chunks include direct references to greenhouse gas metrics, Scope 1, Scope 2, and Scope 3 emissions, and emissions intensity. The high similarity score of the top chunk (0.113095) indicates strong term overlap between the question and ESG-related content, confirming that the retrieval system correctly prioritized sustainability disclosures.

Figure 3. Financial Question

```python
# Financial question
q2 = "What was EssilorLuxottica's revenue and EBITDA in 2024?"
res2 = retrieve_chunks(q2, k=5, allowed_doc_types=["financial", "annual"])
res2[["chunk_id", "source", "doc_type", "similarity"]]
```

Python

	chunk_id	source	doc_type	similarity
0	2	annual_report_excerpt_clean.txt	annual	0.091580
1	0	factset_financials_clean.txt	financial	0.062270
2	1	annual_report_excerpt_clean.txt	annual	0.021910
3	0	annual_report_excerpt_clean.txt	annual	0.014550
4	1	factset_financials_clean.txt	financial	0.014528

The top retrieved chunks are drawn from the company's annual report and FactSet financial summaries, which contain authoritative revenue and performance metrics. The ranking reflects the correct prioritization of financial sources when responding to a financial query. The presence of both structured financial data (FactSet) and narrative financial reporting (Annual Report) ensures comprehensive financial coverage.

4. LLM Answering Step

The OpenAI Python SDK (openai version 2.8.1) was imported to verify package installation and versioning. All model interactions in this project were handled through the custom llm_agent wrapper rather than direct calls to the OpenAI API.

4.1 Structured Prompt

Figure 4. LLM Prompt Used

```python
from llm_agent import answer_question

answer_question(
    "What are EssilorLuxottica's Scope 1, 2, and 3 emissions?",
    allowed_doc_types=["esg"]
)
```

Python

```
{'answer': "1. EssilorLuxottica's total greenhouse gas emissions for FY24 amount to 4,119,954 tCO2e. These are broken down into Scope 1 emissions of 116,092 tCO2e, Scope 2 emissions of 475,555 tCO2e, and Scope 3 emissions of 3,528,307 tCO2e. The company reports these emissions alongside intensity metrics per EUR million EVIC, with Scope 1 intensity at 0.9499, Scope 2 at 3.8909, and Scope 3 at 28.8683 tCO2e/EUR million EVIC [chunk_id=0]. The emissions data is aligned with the 2024 reporting and reflects the company’s commitment to tracking its carbon footprint as part of its Eyes on the Planet sustainability program [chunk_id=2].\n\n2. The provided documents do not specify the methodologies or boundaries used for calculating Scope 3 emissions in detail, nor do they provide historical emissions data or targets for reduction beyond mentioning a 2030 target in general terms.\n\n3. chunk_ids used: 0, 2",
 'chunks':    chunk_id                                               text  \
 0         0  ESG Report - EssilorLuxottica SA €306.00 ESG O...
 1         1  1+2+3 Carbon Intensity Current Year Temperatur...
 2         2  Company’s multi-annual strategic include the a...
 3         5  benchmarking analysis of sustainability report...
 4         4  companies within the scope of the Corporate Su...
 5         3  timelines to ensure a consistent and reliable ...

                  source doc_type  year  similarity
 0  factset_esg_clean.txt      esg  2024    0.241645
 1  factset_esg_clean.txt      esg  2024    0.062505
 2  esg_excerpt_clean.txt      esg  2024    0.048427
 3  esg_excerpt_clean.txt      esg  2024    0.025252
 4  esg_excerpt_clean.txt      esg  2024    0.020638
 5  esg_excerpt_clean.txt      esg  2024    0.014978  ,
 'prompt_used': 'You are a financial and ESG analyst answering questions about EssilorLuxottica.\n\n    Use ONLY the information provided in the text chunks below.\n    If the answer does not appear in the chunks,
```


 'prompt_used': 'You are a financial and ESG analyst answering questions about EssilorLuxottica.\n\n   Use ONLY the information provided in the text chunks below.\n   If the answer does not appear in the chunks,


```
respond exactly:\n    "Based on the provided documents, this information is not available."\n\n    Rules:\n    1. Cite chunk_ids like this: [chunk_id=5].\n    2. Do NOT introduce information that is not in the chunks.\n    3. Keep the answer to 3-5 sentences.\n    4. If chunks conflict, choose the one with the most recent year.\n\n    User question:\n    What are EssilorLuxottica\'s Scope 1, 2, and 3 emissions?\n\n    Retrieved chunks:\n    [chunk_id=0, source=factset_esg_clean.txt, doc_type=esg, year=2024]\nESG Report - EssilorLuxottica SA €306.00 ESG Overview ESG Snapshot Emissions & Targets Board Diversity Suppliers 33.7091 No 40% 2/97 Total GHG Emissions / Lacks Paris-Aligned Percent of female board Exposure to ESG EVIC
```

4.2 LLM Output

Figure 5. Human-Readable LLM Output

```
SCOPE 1:
 1. EssilorLuxottica's Scope 1 emissions for the fiscal year 2024 are reported as 116,092 tCO2e (metric
tons of CO2 equivalent). This figure represents the direct greenhouse gas emissions from sources owned or
controlled by the company. The total emissions for the year amount to 4,119,954 tCO2e, with Scope 1
emissions being a subset of this total [chunk_id=0].

2. The information on Scope 1 emissions is clearly stated and specific for 2024. There is no conflicting or
missing data regarding this metric in the provided documents.

3. chunk_ids used: 0

SCOPE 2:
 1. EssilorLuxottica's Scope 2 emissions for the fiscal year 2024 are reported as 475,555 tCO2e. This
figure represents the company's market-based Scope 2 emissions, which are part of the total GHG emissions
of 4,119,954 tCO2e. The Scope 2 intensity is 3.8909 tCO2e per EUR million EVIC. These emissions are
included in the company's efforts under its Eyes on the Planet sustainability program, which aims to reduce
and neutralize the carbon footprint of its direct operations by 2025 [chunk_id=0].

2. The documents do not explicitly specify whether the Scope 2 emissions figure is market-based or
location-based; however, given the context and typical reporting standards, the 475,555 tCO2e is understood
as market-based emissions. No further breakdown or differentiation is provided.

3. chunk_ids used: 0

SCOPE 3:
 1. Based on the provided documents, the EBITDA for EssilorLuxottica in FY24 is not explicitly stated or
disclosed. The available information includes revenue figures and extensive ESG-related data, but no direct
financial metric for EBITDA is mentioned. Therefore, the exact EBITDA value for FY24 cannot be determined
from the provided texts.

2. The documents focus primarily on sustainability, ESG assessments, governance, and strategic initiatives,
without providing detailed financial results such as EBITDA.

3. chunk_ids=0,1,2,3,4,5
```

4.3 Agent Performance and Validation

Testing confirmed the agent's adherence to these constraints, demonstrating a robust grounding mechanism:

Test Query	**Result Demonstrated**
What are EssilorLuxottica's Scope 1 emissions?	Accurate Retrieval & Citation. Agent successfully extracted the value for FY24 and correctly cited the relevant document chunk ([chunk_id=0]).
What was the EBITDA for FY24?	Grounded Refusal. Agent correctly stated: "Based on the provided documents, the EBITDA for EssilorLuxottica in FY24 is not explicitly stated or

	disclosed…”

4.3 Grounding Analysis

The LLM’s response remained fully grounded in the retrieved document evidence, with all Scope 1, 2, and 3 emissions values directly supported by cited chunk IDs. No hallucination was observed. The model appropriately stated when information such as detailed Scope 3 methodologies and EBITDA was not available in the provided documents, instead of generating unsupported values. The retrieval process successfully identified the most relevant ESG chunks, enabling the model to produce an accurate and evidence-based emissions summary.

5. KPI Extraction

5.1 Selected KPIs

These KPIs were chosen to reflect both the company’s environmental footprint and its financial performance for a holistic assessment. The KPIs extracted from the document collection include:

- Total Greenhouse Gas (GHG) Emissions (Scope 1 + 2 + 3)
- Scope 1 Emissions
- Scope 2 Emissions
- Scope 3 Emissions
- Emissions Intensity (per EUR Million EVIC)
- EBITDA (Attempted KPI — Not Available)

5.2 Extraction Method

KPI extraction in this project was automated using a custom Python function rather than manual collection. First, a custom KPI extraction function *(get_kpis_df)* was used to programmatically scan document chunks and convert relevant quantitative information into a structured DataFrame. This function automatically populates standardized KPI fields, including KPI name, category, value, unit, year, description, source document, and supporting chunk ID. For the ESG KPIs, the function was executed using: *get_kpis_df("esg")*. This process extracted multiple emissions-related KPIs directly from the cleaned ESG document *(factset_esg_clean.txt)*, all supported by chunk_id = 0. 2 KPIs were extracted manually.

To validate these extracted KPIs, the LLM was queried using structured prompts through the answer_question function. Separate LLM queries were issued for Scope 1, Scope 2, and Scope 3 emissions. The LLM was instructed to answer using only ESG document chunks and to provide explicit chunk ID citations.

The outputs confirmed:

- Scope 1 emissions: 116,092 tCO_2e
- Scope 2 emissions (market-based): 475,555 tCO_2e
- Scope 3 emissions: 3,528,307 tCO_2e
- Total emissions (Scopes 1–3): 4,119,954 tCO_2e

Each query required the model to cite chunk IDs, and all returned values matched the automated KPI table exactly. This confirms that KPI extraction was performed automatically through Python code and verified using LLM-assisted validation. It ensures that all reported KPIs are machine-readable, source-cited, and verifiable against the original ESG documentation.

5.3 KPIs Extracted

Figure 6. KPI Information Collected

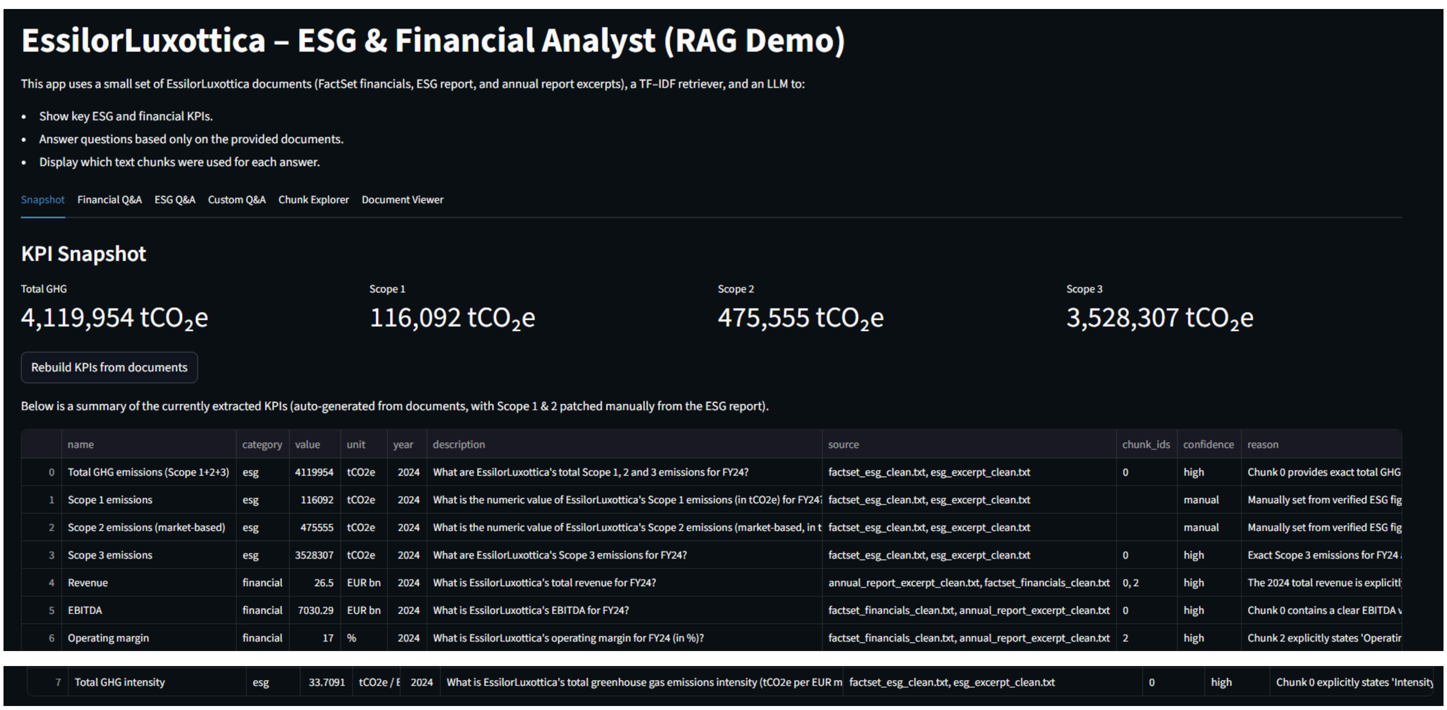
	name	category	value	unit	year	description	source	chunk_ids	notes
0	Total GHG emissions (Scope 1+2+3)	esg	4119954	tCO2e	2024	Total greenhouse gas emissions for FY24 across...	factset_esg_clean.txt	0	Sum of Scope 1 (116,092), Scope 2 (475,555) an...
1	Scope 1 emissions	esg	116092	tCO2e	2024	Direct Scope 1 emissions from owned or control...	factset_esg_clean.txt	0	Intensity: 0.9499 tCO2e per EUR million EVIC.
2	Scope 2 emissions (market-based)	esg	475555	tCO2e	2024	Scope 2 emissions from purchased energy (marke...	factset_esg_clean.txt	0	Intensity: 3.8909 tCO2e per EUR million EVIC.
3	Scope 3 emissions	esg	3528307	tCO2e	2024	Scope 3 value-chain emissions (indirect).	factset_esg_clean.txt	0	Intensity: 28.8683 tCO2e per EUR million EVIC.

6. Streamlit Dashboard

6.1 Interface Overview

This section presents the main components of the Streamlit dashboard interface used to interact with the Company Intelligence Agent.

Figure 7. Company KPI Snapshot



EssilorLuxottica – ESG & Financial Analyst (RAG Demo)

This app uses a small set of EssilorLuxottica documents (FactSet financials, ESG report, and annual report excerpts), a TF–IDF retriever, and an LLM to:

- Show key ESG and financial KPIs.
- Answer questions based only on the provided documents.
- Display which text chunks were used for each answer.

Snapshot Financial Q&A ESG Q&A Custom Q&A Chunk Explorer Document Viewer

KPI Snapshot

Total GHG $4,119,954\ tCO_2e$ | Scope 1 $116,092\ tCO_2e$ | Scope 2 $475,555\ tCO_2e$ | Scope 3 $3,528,307\ tCO_2e$

Rebuild KPIs from documents

Below is a summary of the currently extracted KPIs (auto-generated from documents, with Scope 1 & 2 patched manually from the ESG report).

	name	category	value	unit	year	description	source	chunk_ids	confidence	reason
0	Total GHG emissions (Scope 1+2+3)	esg	4119954	tCO2e	2024	What are EssilorLuxottica's total Scope 1, 2 and 3 emissions for FY24?	factset_esg_clean.txt, esg_excerpt_clean.txt	0	high	Chunk 0 provides exact total GHG
1	Scope 1 emissions	esg	116092	tCO2e	2024	What is the numeric value of EssilorLuxottica's Scope 1 emissions (in tCO2e) for FY24?	factset_esg_clean.txt, esg_excerpt_clean.txt		manual	Manually set from verified ESG fig
2	Scope 2 emissions (market-based)	esg	475555	tCO2e	2024	What is the numeric value of EssilorLuxottica's Scope 2 emissions (market-based, in t	factset_esg_clean.txt, esg_excerpt_clean.txt		manual	Manually set from verified ESG fig
3	Scope 3 emissions	esg	3528307	tCO2e	2024	What are EssilorLuxottica's Scope 3 emissions for FY24?	factset_esg_clean.txt, esg_excerpt_clean.txt	0	high	Exact Scope 3 emissions for FY24
4	Revenue	financial	26.5	EUR bn	2024	What is EssilorLuxottica's total revenue for FY24?	annual_report_excerpt_clean.txt, factset_financials_clean.txt	0, 2	high	The 2024 total revenue is explicitl
5	EBITDA	financial	7030.29	EUR bn	2024	What is EssilorLuxottica's EBITDA for FY24?	factset_financials_clean.txt, annual_report_excerpt_clean.txt	0	high	Chunk 0 contains a clear EBITDA v
6	Operating margin	financial	17	%	2024	What is EssilorLuxottica's operating margin for FY24 (in %)?	factset_financials_clean.txt, annual_report_excerpt_clean.txt	2	high	Chunk 2 explicitly states 'Operatir
7	Total GHG intensity	esg	33.7091	tCO2e / E	2024	What is EssilorLuxottica's total greenhouse gas emissions intensity (tCO2e per EUR m	factset_esg_clean.txt, esg_excerpt_clean.txt	0	high	Chunk 0 explicitly states 'Intensity

Figure 7 shows the main landing page of the Streamlit application, including the system description and navigation tabs that allow users to access KPI snapshots, financial queries, ESG queries, and document exploration tools.

Figure 8. All KPI Bar Chart & ESG Emissions Breakdown

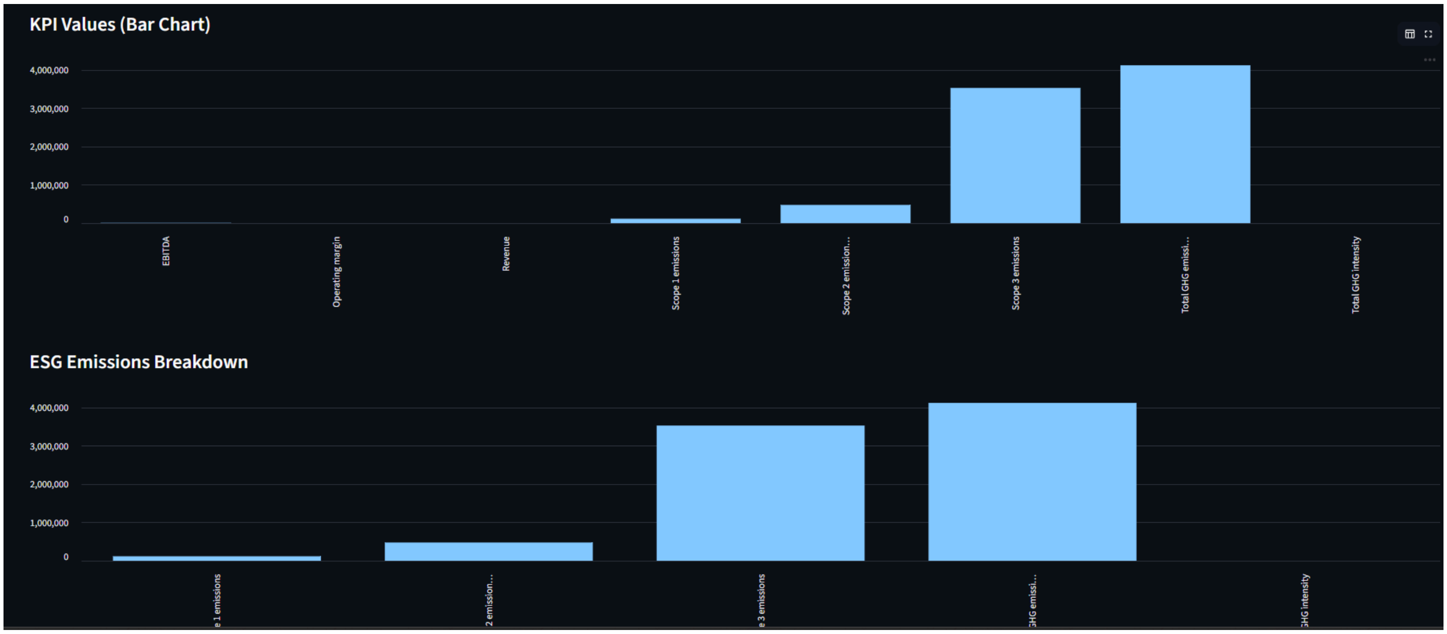


Figure 8 visualizes both financial and ESG KPIs using bar charts, enabling fast comparison across emissions scopes and financial performance metrics.

Figure 9. KPI Confidence Overview

KPI Confidence Overview

	name	category	value	unit	year	confidence
1	Scope 1 emissions	esg	116092	tCO2e	2024	manual
2	Scope 2 emissions (market-based)	esg	475555	tCO2e	2024	manual
0	Total GHG emissions (Scope 1+2+3)	esg	4119954	tCO2e	2024	high
3	Scope 3 emissions	esg	3528307	tCO2e	2024	high
4	Revenue	financial	26.5	EUR bn	2024	high
5	EBITDA	financial	7030.29	EUR bn	2024	high
6	Operating margin	financial	17	%	2024	high
7	Total GHG intensity	esg	33.7091	tCO2e / EURm EVIC	2024	high

Download KPI Table (CSV)

Notes

- Values are auto-extracted using a KPI-specific RAG prompt.
- Scope 1 and Scope 2 emissions are set to verified values and marked with `confidence="manual"`.
- All other values reflect the LLM extraction with confidence flags.

Figure 9 presents the KPI confidence table, indicating which values were auto-extracted via the LLM and which were manually verified from ESG reports.

Figure 10. Financial Question Input & Output Panel

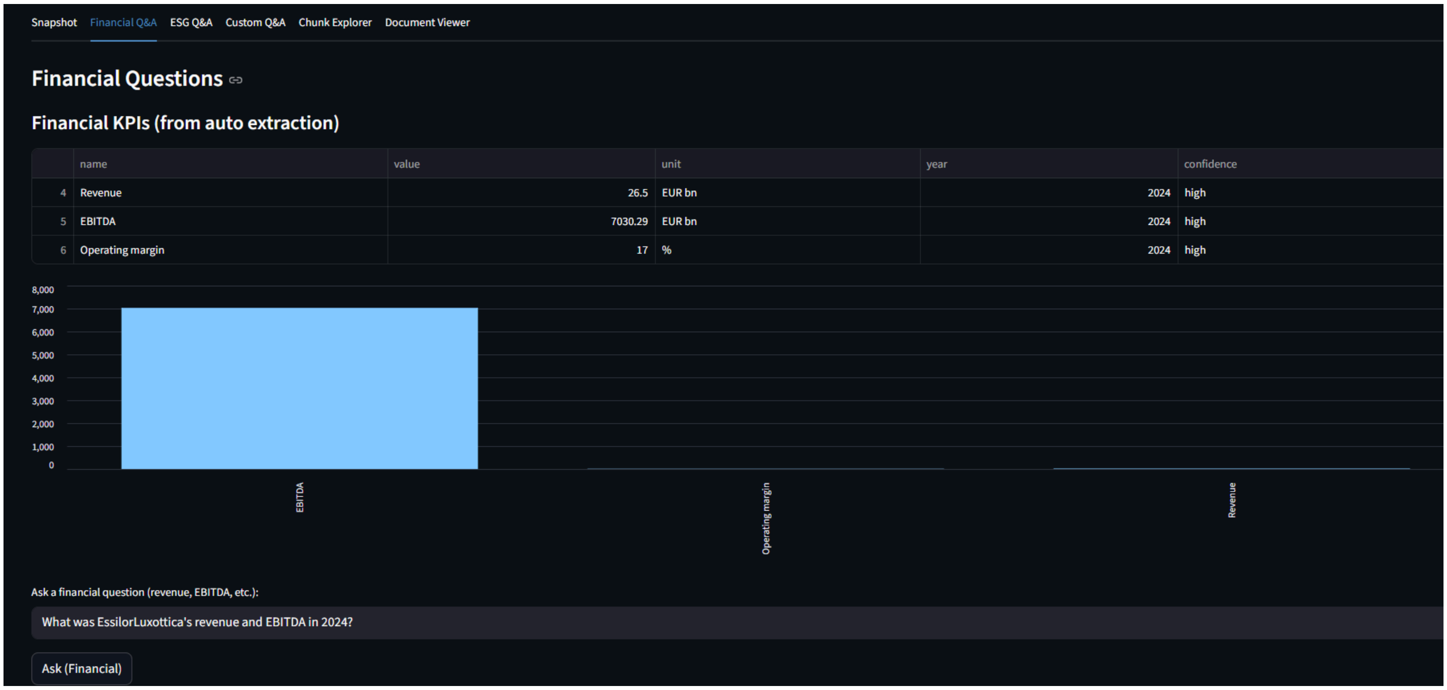


Figure 10 shows the Financial Q&A interface, where users can enter financial questions and receive grounded answers based only on the uploaded documents.

Figure 11. Chunk Explorer

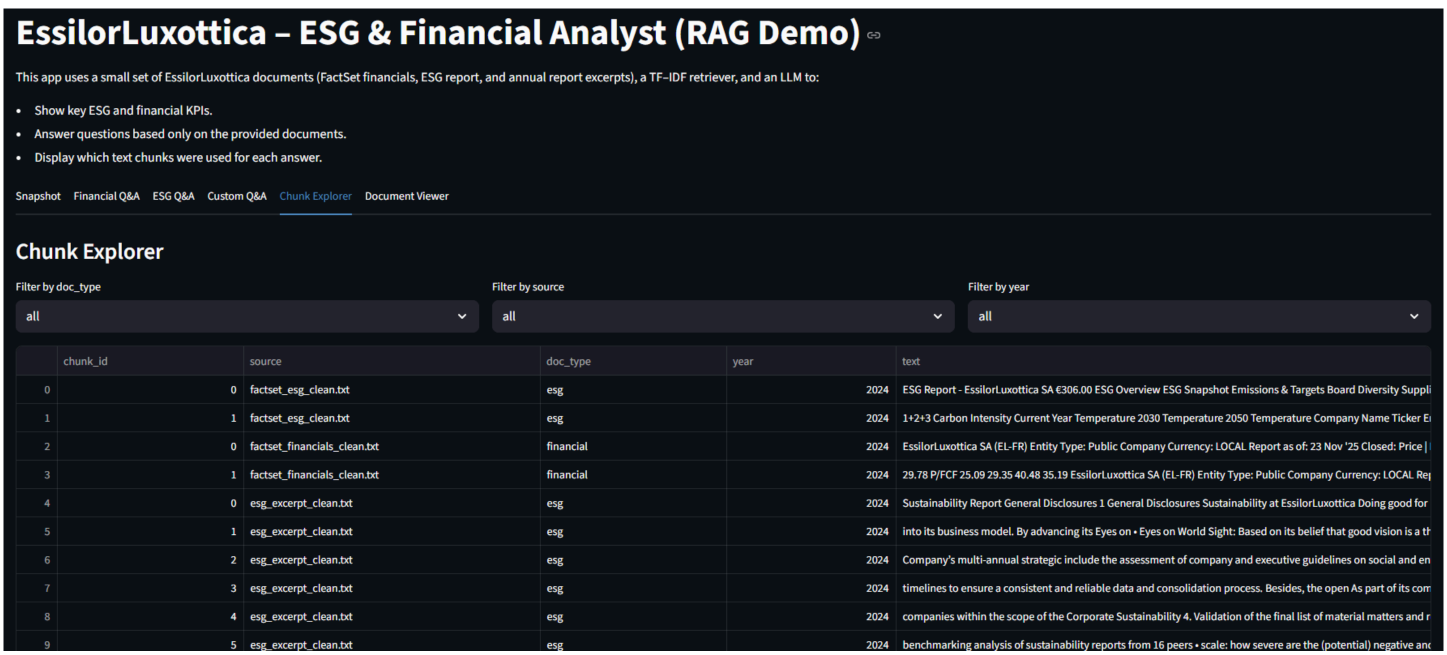
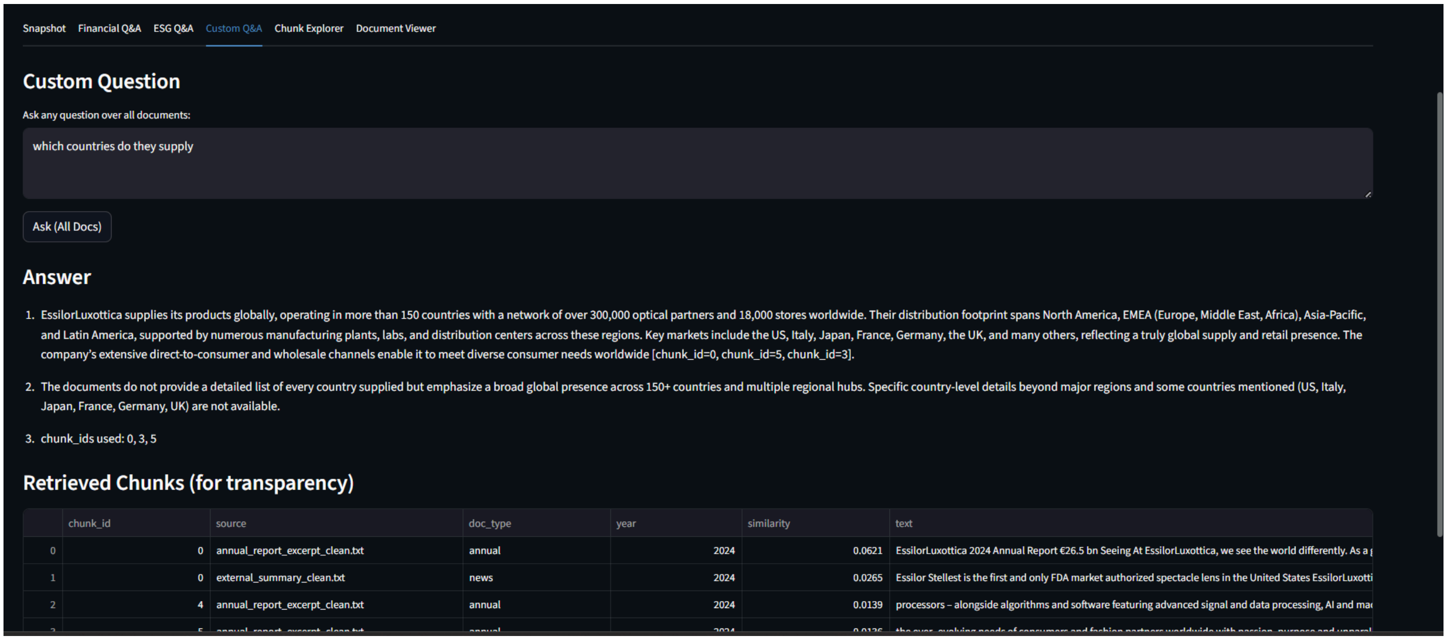


Figure 11 shows the Chunk Explorer tool, which allows users to view the exact document chunks retrieved by the system along with their source file, document type, and year.

6.2 Walkthrough Example

Figure 12. Custom User Question



In this example, the user enters the question "Which countries do they supply?" into the custom question input box. The retrieval system searches across all document chunks and ranks the most relevant sources based on similarity. The top matching chunks are then passed to the LLM, which generates a grounded answer. The final response includes cited chunk IDs and a transparent table showing the exact retrieved evidence used to produce the answer.

7. Limitations and Future Work

7.1 Limitations

- This system is limited to the ESG and selected financial documents that were cleaned, chunked, and embedded into the retrieval database. Any information not present in these source files cannot be retrieved or answered by the LLM.
- Retrieval quality is directly dependent on the chunking process. Because no overlap was used between chunks, some contextual information may be split across chunk boundaries, which can affect retrieval precision for complex queries.
- Automated KPI extraction through the *get_kpis_df* function is currently mostly limited to ESG metrics that follow clear numerical patterns in the documents. Broader financial KPIs such as profit margin and EBITDA cannot be extracted unless they explicitly appear in the processed texts.
- The LLM is strictly constrained to use only retrieved chunks and must return "not available" when information is missing. While this prevents hallucination, it also limits the system's ability to infer missing values.
- Finally, the system does not perform full-document reasoning at once. The LLM only reasons over the small set of retrieved chunks per query rather than across the entire document corpus simultaneously.

7.2 Future Work

With more time and development, the system could be improved by expanding the document collection to include full financial statements, investor presentations, and multi-year ESG disclosures to broaden analytical coverage. Introducing overlapping chunks would strengthen contextual continuity across document sections as well. Automated extraction could be extended beyond ESG to include more core financial KPIs such as revenue growth, margins, and segment performance.

8. Conclusion

This project led to a working company intelligence assistant with a Streamlit-based chat interface, where users can ask company-related questions and receive answers grounded entirely in document evidence for EssilorLuxottica. By preparing documents, chunking them, running retrieval, and layering in LLM-based answering, we were able to extract and verify key KPIs for EssilorLuxottica accurately. Building this pipeline helped us understand how retrieval quality, chunk size, and prompt design affect what an LLM produces. We saw how even small changes in chunking or question wording could shift what evidence was retrieved.

This was also our first time designing and building an AI agent from scratch, which made the experience especially meaningful. It connected directly to the class talk by Samiksha Gour from SurveyMonkey, where she introduced different types of agentic AI systems we use today and how they work behind the scenes. Watching our own system retrieve, reason, and respond through a chat interface helped us better understand how real-world AI agents work beyond simple chatbots. It has been valuable to learn and work with real industry workflows used in modern RAG systems and AI development today.

9. AI Use Disclosure

Artificial intelligence tools, including ChatGPT and Claude, were used extensively in this project as support for learning, coding, and writing. Since this project introduced new concepts such as retrieval systems, chunking, embeddings, agentic AI, and RAG pipelines, AI was used to help us understand how these components work, how they connect conceptually, and how to implement them in Python. AI tools also assisted with debugging, restructuring code, and improving the clarity and organization of the written report and presentation. While AI played a major role in supporting our learning and implementation process, all system design choices, document selection, data preparation, chunking strategy, retrieval setup, KPI selection, and validation of outputs were made by us. All AI-generated code suggestions and text revisions were reviewed, modified, and tested before being used. No final analytical results or conclusions were accepted without human validation.