

Company Intelligence Agent

An LLM + Retrieval System for Corporate Analysis

DSDA 310 Case Study 2

Shoumika Annanyo and Moosa Sherwani

Toyota

- One of the world's leading automobile manufacturers
- Market leader in hybrid, fuel-efficient, and emerging technology mobility
- Operations span includes vehicle manufacturing, mobility, R&D, and sustainability across various international markets



TOYOTA

Agent Agenda

Our agent is designed to help users quickly extract reliable insights from large corporate documents. The agent is designed to answer questions related to:

- Company operations and segments
- Strategic priorities
- Risks
- ESG and sustainability commitments
- Financial KPIs

Documents Collected & Methodology

1. Toyota Integrated Report (65p)
2. Toyota Sustainability Data Book (148p)
3. Reuters Market Analysis (1 article)

3-Phase Processing Pipeline:

- Collect → Official + External Sources
- Convert → PDF → Text → Structured CSV
- Verify → Complete, Traceable, Clean

Design Principles:

- Transparency: Full source tracking
- Completeness: No content loss
- Consistency: Uniform processing

Data Preparation and Cleaning

✓ RAW TEXT EXTRACTED

"TOYOTA MOTOR CORPORATION\n\n\nIntegrated Report 2022\n\n\n\n\nMessage from\nthe President\n\n\n\n\nThe Source..."\n\n\n"

↓ CLEANING PIPELINE ↓

✓ CLEANED TEXT READY

"Message from the President\nThe Source of Our Value Creation: What Makes Us Toyota\nOur Founding Spirit\nThe Toyota Principles..."

```
def clean_text(text):
    text = re.sub(r'\n+', '\n', text)      # Fix newlines
    text = re.sub(r' +', ' ', text)        # Fix spaces
    text = re.sub(r'\n\d+\n', '\n', text)   # Remove page nums
    return text.strip()
```

Cleaning Challenges:

- Excessive whitespace & page breaks
- Repeating headers/footers
- Page number artifacts
- Broken words from line endings

Our Solution Pipeline:

1. Whitespace Normalization: `\n+` → single `\n`
2. Page Marker Removal: Isolated page numbers
3. Header/Footer Stripping: Pattern-based removal
4. Word Reconstruction: Fix hyphenated line breaks

Results:

- 15-20% size reduction via cleanup
- 292 clean chunks created
- Preserved semantic meaning & proper nouns

Chunks

```
DOCUMENTS (213 pages total)
↓
SPLIT INTO CHUNKS
↓
292 SEARCHABLE UNITS
↓
CHUNK DISTRIBUTION:
[REDACTED] ESG Report: 170 chunks
[REDACTED] Annual Report: 121 chunks
[REDACTED] External: 1 chunk
```

Chunk Size Rationale:

- 500-1000 words per chunk
- Matches LLM context window limits
- Balances detail with search efficiency
- Project requirement compliance

```
def chunk_text(text, chunk_size=500):
    words = text.split()
    chunks = []
    for i in range(0, len(words), chunk_size):
        chunk = ' '.join(words[i:i + chunk_size])
        chunks.append(chunk)
    return chunks
```

Design Choices:

- No overlap: Simpler system, TF-IDF finds multiple chunks
- Word-boundary splits: Preserves sentence integrity
- Metadata retention: Each chunk tracks source, type, position
- Equal treatment: All documents processed identically

Results Summary

Total Chunks: 292

Avg. Words/Chunk: 750

Storage: toyota_chunks.csv

chunk_id	company	source_file	chunk_text	chunk_number	document_type
annual_report_0	toyota	annual_report_excerpt.pdf	Integrated Report 2022 Integrated Report 2022 fiscal year ended March 31, 2022 TOYOTA MOTOR CORPORATION INTEGRATED REPORT Message from the President The Source of Our Value Creation: What Makes Us To	0	annual_report
annual_report_1	toyota	annual_report_excerpt.pdf	creative for the betterment of lives and society. Using our technology, we work toward a future of convenience and happiness available to all. This is our mission, producing happiness for all, and the core of what makes us To;	1	annual_report
annual_report_2	toyota	annual_report_excerpt.pdf	2 Message from the President 4 The Source of Our Value Creation: What Makes Us Toyota 4 Our Dream Sprint 5 The Toyota Principles and Toyota Philosophy 6 Toyota Production System (TPS) 7 Toyota and Sports 8 Value	2	annual_report
annual_report_3	toyota	annual_report_excerpt.pdf	each country's energy and infrastructure situation while keeping open an array of technological options to accelerate CO2 reduction and zero emission efforts. ¹ I have continued to emphasize these core points. To move forward	3	annual_report
annual_report_4	toyota	annual_report_excerpt.pdf	the preferred choice of customers around the world. I believe this is because the people working at Toyota have changed. Toyota's efforts at the geria are underpinned by the many people, both inside and outside the Comp	4	annual_report
annual_report_5	toyota	annual_report_excerpt.pdf	his first loom, the Toyota Hand Loom, which could be operated with only one hand and greatly increase efficiency. He patented the loom in May 1891. Seeking to more dramatically increase capacity, Sakichi turn	5	annual_report
annual_report_6	toyota	annual_report_excerpt.pdf	completed, and the Toyota Model G1 Index was announced.	6	annual_report
annual_report_7	toyota	annual_report_excerpt.pdf	The very next year, in 1903, mass production of Model AA passenger cars commenced. Toyota Motor Co., Ltd. was established in 1937, with Kichisaburo	7	annual_report
annual_report_8	toyota	annual_report_excerpt.pdf	the time was imposed—navigating tremendous social changes as he built the Company and the foundations of Japan's automotive industry. The spirit they embodied—of striving to stay ahead of the times and endeavor	8	annual_report
annual_report_9	toyota	annual_report_excerpt.pdf	there was a brief time when we turned our focus to numbers and gave less thought to people. Primarily due to our rapid expansion in the late 20th century, we faced many problems, including quality concerns and trade fric	9	annual_report
annual_report_10	toyota	annual_report_excerpt.pdf	people Stakeholders TOYOTA MOTOR CORPORATION INTEGRATED REPORT Message from the President The Source of Our Value Creation: What Makes Us Toyota Value Creation Story: Working toward the Mobility Socie	10	annual_report
annual_report_11	toyota	annual_report_excerpt.pdf	Main Purpose The Type G automatic loom is the machine that helps drive a redesign of Toyota's business. Automatic looms back then were always monitored by one operator, based on a mindset of "one person, one mach	11	annual_report
annual_report_12	toyota	annual_report_excerpt.pdf	"lead time," the amount of time required for products or services to be delivered after they are ordered. Toyota What comes to mind when you think about Just-In-Time? Taking a "what is needed when needed" approach, to	12	annual_report
annual_report_13	toyota	annual_report_excerpt.pdf	"never giving up" and the spirit of working "for the team," which encourages effort on the behalf of others—I believe these were exactly the mindsets the company had when they recklessly took on the chal	13	annual_report
annual_report_14	toyota	annual_report_excerpt.pdf	International Olympic and Paralympic Committees. Approximately 300 Global Team Toyota Athletes from 50 countries and regions competed at the recent Olympic and Paralympic Games in Tokyo and Beijing. Toyota not onl	14	annual_report
annual_report_15	toyota	annual_report_excerpt.pdf	process of the era every two decades. Why is that? I think it is because Toyota treats car development as the front line for developing the skills and knowledge that will be passed down as well as for human resource c	15	annual_report
annual_report_16	toyota	annual_report_excerpt.pdf	centered Management. Sports 800 Publica Sports 2000GT Supra Celica Levin/Turbo MR2 10 AT 10 Years Making Ever-better Cars Initiatives to Achieve Carbon Neutrality Software and Connected Initiatives Commercial Secti	16	annual_report
annual_report_17	toyota	annual_report_excerpt.pdf	and other OEMs don't), is a master driver in top management. A president who can take responsibility for the "flavor" of the products we put out. A president who is able to definitely say "no" to projects, even ones that our	17	annual_report
annual_report_18	toyota	annual_report_excerpt.pdf	strong opposition and criticism. No latent technology gy was ignored in the Crown's creation, including a double-wishbone suspension for the front wheels. Reminiscing about the launch, Nakamura said, "It was like all of Japan entered hard times from the ninth generation onward. The Crown's Transformation: Generations 9 to 15 To start with, the Crown's positioning within Toyota changed. In 1978, Toyota launched the Lexus LS as the Toj	18	annual_report
annual_report_19	toyota	annual_report_excerpt.pdf	entered the vehicle market and the driving performance at the Notburying. Doing so, we created the 14th-generation "Reborn Crown" and "Connected Crown." Over the past 20 years, we constantly challenged ourselves to	19	annual_report
annual_report_20	toyota	annual_report_excerpt.pdf	be the best in town, the best in the world. ¹ I realized that the Crown is a long-time seller because the chief engineers constantly challenged themselves to create an ever-better Crown with a seat-in-town focus; the	20	annual_report
annual_report_21	toyota	annual_report_excerpt.pdf	models at the same time was no easy task. What made it possible were Toyota in-house company and the TNGA. We couldn't have presented the new Crown today without them. The first element, the in-house co	21	annual_report
annual_report_22	toyota	annual_report_excerpt.pdf	workforce. The new Crown is full of these underlying strengths. That is why, with this series, we will once again take on the world. The new Crown will be available in approximately 40 countries and regions, with an expected	22	annual_report
annual_report_23	toyota	annual_report_excerpt.pdf	three subsequent races. For each race, they continued to improve the car, making it stronger and faster. The other car is the GR Yaris. We made this car for a specific purpose: to win the World Rally Championship. Until now, Kamui Kobayashi had seized pole position with an astounding time. He passed the trophy to me while thank- ing me. It made me want to get closer to the drivers and race alongside them. Racing, however, is hard. That	23	annual_report
annual_report_24	toyota	annual_report_excerpt.pdf	race, Kamui Kobayashi had seized pole position with an astounding time. He passed the trophy to me while thank- ing me. It made me want to get closer to the drivers and race alongside them. Racing, however, is hard. That	24	annual_report
annual_report_25	toyota	annual_report_excerpt.pdf	us that there's no way that Toyota can realize this kind of car making. Now, however, Toyota has finally changed, realizing a kind of car making in which not only engineers and mechanics, but its professional drivers, pro - i	25	annual_report
annual_report_26	toyota	annual_report_excerpt.pdf	more energetically. We want racing teams to gather here with that same motivation in mind. If all the motivation behind Forest, it would be endless. We will turn this area in Fuji into a place that those who enjoy motor	26	annual_report
annual_report_27	toyota	annual_report_excerpt.pdf	Sports 123 Lexus RX 111 Tundra SR Venza 66 Powertrain Sales ICE 1,927 HEV 602 PHEV 69 BEV 99 EV 10 FCEV 3 Total 2,591 PHV 216 FCEV 11 ICE 75% HEV 23% (Thousands of units) China Sales By Series (Thousands of u	27	annual_report
annual_report_28	toyota	annual_report_excerpt.pdf	wants to prepare as many options as possible for our customers around the world. We believe that all electrified vehicles can be divided into two categories, depending on the energy that they use. One category is that of "ot	28	annual_report
annual_report_29	toyota	annual_report_excerpt.pdf	the President The Source of Our Value Creation: What Makes Us Toyota Value Creation Story: Working toward the Mobility Society of the Future Business Foundations for Value Creation Corporate Data To achieve this goal	29	annual_report
annual_report_30	toyota	annual_report_excerpt.pdf	steady improvements toward reducing energy use and by expanding the use of innovative production engineering technology. In this diversified and uncharted era, it is important to flexibly change the type and quantity of p	30	annual_report
annual_report_31	toyota	annual_report_excerpt.pdf	annual_report_31 toyota annual_report_excerpt.pdf the growing demand for BEVs around the world. TOYOTA MOTOR CORPORATION INTEGRATED REPORT Message from the President The Source of Our Value Creation: What Makes Us Toyota Value Creation: Workin	31	annual_report
annual_report_32	toyota	annual_report_excerpt.pdf	the truth is that we are not fully committed because our goal is not 100% BEVs. We hope you understand that we would very much like to continue our work in this industry. Evaluation by Environmental Groups and the Future o	32	annual_report
annual_report_33	toyota	annual_report_excerpt.pdf	the president of the company, you can respond as driver Morizo. Also I have to answer, however, in the past I was not interested in Toyota's BEVs, but I am getting interested in the BEVs that we are now developing for the I	33	annual_report
annual_report_34	toyota	annual_report_excerpt.pdf	Morizo. I have strong expectations for and a desire to give my feedback to the development, such as "making this kind of car be for" or "as an automaker we want to autonomous driving that is unique, even if	34	annual_report
annual_report_35	toyota	annual_report_excerpt.pdf	decrease battery safety. This concept has remained unchanged since batteries were installed in the first-generation Prius. Therefore, Toyota is committed to balancing the battery capacity retention ratio when installed in a vehicle environment may vary depending on customer use conditions, usage environment, and driving methods. Therefore, a 10% battery capacity retention ratio after 10 ye	35	annual_report
annual_report_36	toyota	annual_report_excerpt.pdf	annual_report_36 toyota annual_report_excerpt.pdf latent risks when production is growing. Because of this, we have to take a risk-oriented approach to growth based on Toyota's philosophy of "making only what is needed, when it is need - ed, and only in the amount need	36	annual_report
annual_report_37	toyota	annual_report_excerpt.pdf	annual_report_37 toyota annual_report_excerpt.pdf Emissions with Few Batteries Since the introduction of the first-generation Prius in 1997, Toyota has introduced PHEVs, FCEVs, and BEVs while continuing to improve their perfor- mance. Our more than 20 years of producin	37	annual_report
annual_report_38	toyota	annual_report_excerpt.pdf	annual_report_38 toyota annual_report_excerpt.pdf b24X in the second half of the 2020s. Using Solid-state Batteries Starting with HEVs in the near future, the energy density of conven- tional lithium-ion batteries per unit of weight is expected to peak. Accordingly, vigorous ef	38	annual_report
annual_report_39	toyota	annual_report_excerpt.pdf	annual_report_39 toyota annual_report_excerpt.pdf ffectiveness of the batteries. Therefore, we have to take a risk-oriented approach to growth based on Toyota's philosophy of "making only what is needed, when it is need - ed, and only in the amount need	39	annual_report
annual_report_40	toyota	annual_report_excerpt.pdf	annual_report_40 toyota annual_report_excerpt.pdf	40	annual_report

Retrieval Demo

How It Works:

- Question: User asks about Toyota's business or risks
- Search: TF-IDF scans all 292 document chunks
- Match: Finds most conceptually relevant sections
- Return: Shows top 2-3 chunks with source attribution

Key Insights from the Demo:

1. Semantic Understanding: Matches concepts, not just keywords
2. Multiple Perspectives: Returns complementary information
3. Source Transparency: Every result is traceable to original documents
4. Real-time Performance: Fast enough for interactive use

Why This Matters:

- Foundation for accurate AI answers
- Ensures evidence-based responses
- Enables source citation
- Provides user confidence in results

```
TF-IDF model trained on 292 documents
=====
RETRIEVAL SYSTEM DEMONSTRATION - WITH RELEVANCE ANALYSIS
=====

1. QUESTION: "What does Toyota do?"
   Retrieved chunks:
   1. [ID: annual_report_31]
      Preview: the growing demand for BEVs around the world. TOYOTA MOTOR CORPORATION INTEGRATED REP
      ORT Message from the President The ...
      → Relevance: Annual report chunk describing Toyota's core automotive business

   2. [ID: esg_report_208]
      Preview: and dealers. ' Contents of the activities: w Multilingual web portal and application
      that provide relevant information...
      → Relevance: ESG report showing Toyota's operational scope and dealer network

2. QUESTION: "What are Toyota's risk factors?"
   Retrieved chunks:
   1. [ID: esg_report_194]
      Preview: 29 23 Japan (excluding Toyota Motor Corporation) 111 107 103 North America 47 48 65 E
      urope 11 11 10 Asia 30 35 38 Others...
      → Relevance: Geographic risk distribution across regions

   2. [ID: esg_report_277]
      Preview: CRO (DCRO) wPerson supervising risk management in each region: Regional CEO wPerson r
      esponsible/in charge of risk mana...
      → Relevance: Contains explicit risk management roles (CRO = Chief Risk Officer)

=====

RELEVANCE ASSESSMENT:
• Question 1 chunks describe Toyota's business operations
• Question 2 chunks contain risk management terminology
```

LLM- Powered Answer Generation

The screenshot shows a user interface for generating answers using an LLM. At the top, a header says "Ask Questions About Toyota". Below it is a text input field containing the question "What are Toyota's main risk factors?". To the right of the input field is a red "Get Answer" button with a magnifying glass icon. The next section, titled "Retrieved Evidence", displays a yellow box containing the text "Top 3 relevant chunks found: esg_report_277, esg_report_194, annual_report_108". Below this is a section titled "LLM Answer with Citations" which contains the text "Toyota's main risk factors include natural disasters like earthquakes and floods, disruptions in the supply chain, and interruptions in the delivery of essential goods [esg_report_277]". At the bottom, there are three entries under "View detailed source information": "esg_report_277" (From Toyota corporate documents), "esg_report_194" (From Toyota corporate documents), and "annual_report_108" (From Toyota corporate documents).

Key Components Shown in Screenshot:

1. User Query: "What are Toyota's main risk factors?"
2. Retrieved Evidence: Top 3 relevant document chunks
3. LLM Answer: Concise, cited response
4. Source Transparency: Full evidence trail

How It Works:

User Question → Retrieval → Evidence → LLM → Cited Answer

Why This Approach is effective:

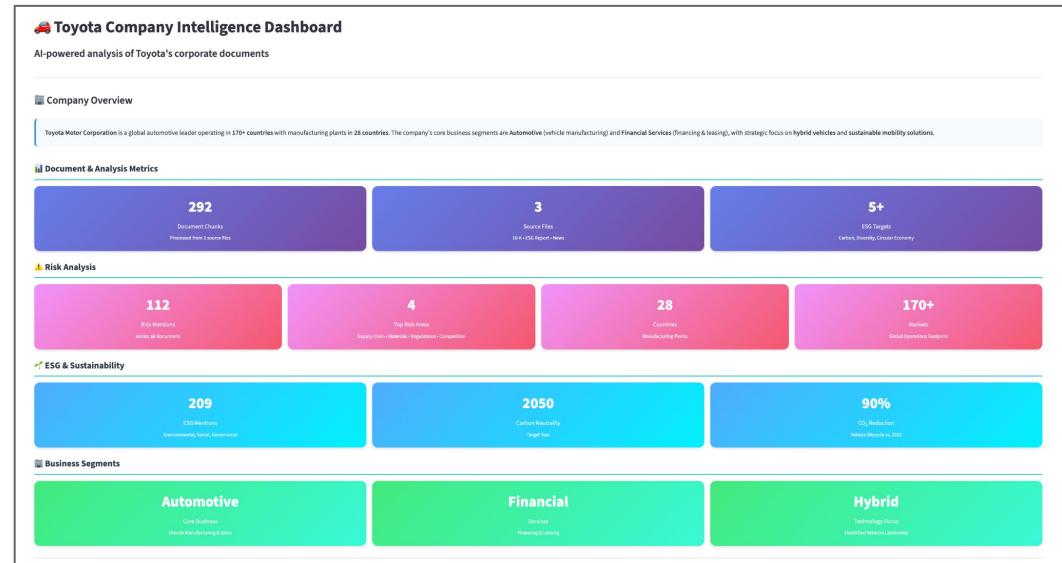
- Evidence-Based: No hallucinations
- Transparent: Full source visibility
- Auditable: Every claim traceable
- User-Trusted: Verifiable answers

KPI Extraction

KPIs extracted:

1. Risks
2. ESG
3. Business: Automotive and Financial Services
4. Global Operational Footprint: 170+ countries

KPIs were extracted using a combination of manual inspection, keyword frequency analysis, and Python scripts.



Limitations and Future Work

Limitations:

1. TF-IDF is limited in semantic understanding of terms
2. PDFs have formatting issues
3. The scope of the LLM is limited to the data provided

Future Improvements:

1. Integrate embedding-based retrieval for stronger semantic matching
2. Add confidence scoring to outputs
3. Automate document ingestion
4. Enhance UI for more interactive exploration of chunk-level evidence

Toyota Streamlit Dashboard - Live!

Thank You!