VRIJE UNIVERSITEIT AMSTERDAM

Master Thesis

A Hybrid Information Retrieval Framework for Partner Matching: A Case Study on the B Corp Ecosystem

by

Yitong Tang

(2798775)

First supervisor:	Victor de Boer
Daily supervisor:	Chris van Aart
Second reader:	Ronald Maria Siebes

July 7, 2025

Submitted in partial fulfillment of the requirements for the VU degree of Master of Science in Information Sciences

DECLARATION OF AUTHORSHIP

Use one declaration for each author.

I, **Yitong Tang**, declare that this thesis titled, **[Enhancing B Corps Collaboration Through Web Data and LLM: A Design Science Perspective]** and the work presented in it are my own . I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Yitong Tang

Date: 2025-06

A Hybrid Information Retrieval Framework for Partner Matching: A Case Study on the B Corp Ecosystem

Yitong Tang
Vrije Universiteit Amsterdam
Amsterdam, The Netherlands
y.tang3@student.vu.nl

ABSTRACT

This thesis presents the design, implementation, and evaluation of a hybrid information retrieval system that leverages structured filtering and large language models to improve partner matching within the B Corp ecosystem. As businesses increasingly form value-driven networks, the challenge of identifying suitable strategic partners from large, unstructured datasets has become a critical information science problem. This thesis addresses this challenge by using the B Corp ecosystem, a global network of over 9,000 mission-driven companies, which are certified companies committed to balancing profit with high standards of social and environmental performance, as a compelling case study. It presents the design, development, and evaluation of a novel partner-matching system that leverages a hybrid information retrieval framework.

Following a Design Science Research methodology, this project developed and iteratively refined a system through four distinct prototypes. The final architecture combines the efficiency of a traditional search engine (Elasticsearch) for structured data filtering with the deep semantic understanding of an advanced Large Language Model for analyzing unstructured text like company descriptions and mission statements.

A mixed-methods evaluation, including quantitative performance metrics (precision, recall, response time) and qualitative user interviews with B Corp stakeholders, was conducted. The results demonstrate that the final hybrid model outperforms baseline systems. It provides more accurate, relevant, and well-reasoned recommendations, particularly in interpreting complex, natural language queries and handling geographical constraints.

This research contributes a validated, transferable blueprint for building intelligent matching systems for private or domain-specific knowledge bases. It demonstrates that combining traditional search technology with modern LLMs is a highly effective approach for transforming raw data into actionable partnership opportunities, with broad applicability to other mission-driven or complex business networks.

1 INTRODUCTION

In our increasingly interconnected digital society, the fundamental challenge of finding the "right" match—whether between individuals, organizations, or ideas—has become a central problem in information science. Effectively navigating vast and complex networks to identify synergistic partnerships is critical, especially within mission-driven ecosystems where alignment is based not only on functional criteria but also on shared values, goals, and culture. This thesis tackles this universal matching problem by developing a hybrid information retrieval framework designed for complex, multi-dimensional entity matching. To ground this research in a tangible and impactful context, the global B Corp movement is utilized as a compelling case study of a network of like-minded organizations.

As businesses face increasing pressure to address pressing social challenges such as climate change, income inequality, and ethical supply chains, sustainable development has become integral to corporate growth. Over the past decade, Certified B Corporations and Benefit Corporations, often referred to as B Corps, have gained significant global attention[3]. B Corps are committed to generating social, environmental, and economic value for their stakeholders. Till the June of 2025, according to the B corps website[1] there are 9892 certified B corps from 104 countries all around the world.

Any for-profit business can pursue B Corp certification. To achieve this certification, companies must complete the B Impact Assessment (BIA) and score above 80 out of a possible 200 points. The BIA, developed by B Lab, evaluates performance across five key categories: Governance, Workers, Customers, Community, and the Environment. Recertification is required every three years.

B Corps collaborate in various ways to achieve shared goals, which is to benefit all people, communities, and the planet. Participating corporations hold annual summits on various subtopics[2]. Enterprises not only share common goals, but also cooperate in practical ways to solve practical problems. Notable examples of collaboration include resource-sharing, where 15 B Corp brands sold their products at a flexible working space in London that also hosted events for other B Corps[17]. Another strategy is leveraging economies of scale, seen in initiatives like B Locals—regional collectives of B Corps across the UK supporting local communities. Industry-specific alliances also drive progress; the B Corp Beauty Coalition, with 26 member brands, aims to enhance sustainability standards, while a B Corp-certified travel collective unites tourism companies to promote sustainable practices[17].

The collaborations observed among B Corps are part of a broader landscape of corporate partnerships, which academic literature has extensively examined. Collaboration between companies takes various forms, shaped by industry needs, sustainability goals, and innovation strategies. Majava et al. (2013) compare business collaboration models, such as networks, clusters, and ecosystems, analyzing their impact on innovation and new product development[12]. Supplier collaboration is particularly significant for sustainability. Ukko et al. (2022) examine how small e-businesses leverage supplier collaboration to enhance sustainability and market performance[18], while Chen et al. (2017) provide a comprehensive review of supply chain partnerships for sustainability[4]. In service-based collaborations, Schmidt et al. (2015) investigate successful partnerships between Swedish occupational health service (OHS) providers and

client companies[16]. Companies also collaborate on social and environmental projects, where personal relationships can be crucial for facilitating joint initiatives[7]. These studies demonstrate that cooperation is multifaceted. This complexity presents a significant challenge: identifying suitable partners is no longer a simple matter of matching similar profiles. For example, for a retail company, potential partners may exist not only within its own value chain but also in adjacent industries that share common development goals.

Correct and efficient presentation of company information is a critical prerequisite for identifying and recommending potential partners. In this regard, the product classification systems of e-commerce platforms offer a valuable model. Product classification is essential in e-commerce for organization, search optimization, and recommendation systems. Machine learning algorithms have become the leading solution, using either flat or hierarchical approaches[5]. Li et al. (2020) show that incorporating both product titles and attributes significantly outperforms traditional models, emphasizing the importance of multi-modal data[11]. Additionally, distributional semantics improves document representation by capturing contextual similarities between product categories[9].

This project addresses the complex challenge of fostering enterprise cooperation within the sustainable development sector. At its core, this research contributes to the field of information science by developing a novel framework that automates the identification of strategic partners from complex, domain-specific datasets. By applying this framework to the B Corp ecosystem, the proposed system not only enhances collaborative efficiency but also unlocks new opportunities for collective impact, ultimately demonstrating how advanced information retrieval can accelerate the sustainable development agenda.

1.1 Motivation

This project addresses the challenge of identifying optimal partnership matches within large-scale databases. As digitization advances, companies accumulate vast datasets that are impractical to review manually, making intelligent, tailored search methods essential. This project aims to develop an efficient information retrieval solution to solve this problem. The B Corp community, with its shared value system but diverse industries and needs, provides a perfect testbed for developing a sophisticated matching algorithm that balances both shared mission and business needs factors.

The B Corp database provides an ideal case study for this challenge. It is not only public, large-scale, and of high quality, but also represents a community facing a tangible need. With nearly 10,000 B Corps worldwide, manually evaluating potential partners is a monumental task, compounded by the lack of a well-structured platform for partner discovery. Preliminary interviews with B Corp leaders confirm that many struggle to identify suitable partners from this vast pool. This gap presents a clear opportunity to develop a technological solution that facilitates efficient and effective collaboration among B Corporations. While the initial focus is on B Corps, the methodologies developed in this project could ultimately be adapted for other value-aligned networks or ecosystems.

The B Corp framework inherently aligns with sustainable development goals. However, academic attention has predominantly focused on the certification process itself—how companies become B Corps—rather than on fostering collaboration among them. A review of the literature reveals a focus on certification standards[8], with limited exploration of collaborative models. This is a significant oversight, as numerous proven collaborative strategies exist for advancing sustainability, such as supply chain partnerships[4], co-created platforms[15], and supplier alliances[18]. Given that all B Corps share a foundational commitment to sustainability, collaboration offers a powerful pathway to achieving greater collective impact. This project is novel in that it moves beyond the 'why' of B Corp collaboration to the 'how,' by developing a data-driven tool to operationalize these partnership opportunities.

1.2 Problem Definition

The challenge of identifying suitable collaboration partners among B Corporations can be framed as a classic information science problem, encompassing three core stages: information acquisition, organization, and recommendation. The rapid growth of the B Corp network has resulted in an overwhelming volume of dispersed and often unstructured data, hindering effective partnership navigation. This macro-problem can be broken down into three key information science sub-challenges.

First is the problem of data acquisition and structuring. While a public directory of all certified B Corps exists, this information is dispersed and not immediately usable. The foundational task for this project, therefore, is to systematically gather this public data and organize it into a clean, structured, and queryable database that can serve as the backbone for any subsequent analysis.

The second challenge lies in information presentation and usability. A raw database of nearly 10,000 companies is overwhelming. Decision-makers within the B Corp community currently lack a straightforward way to sift through this vast network, to efficiently filter, compare, and evaluate potential partners based on criteria that matter to them. This project, therefore, must address this usability gap by designing a clear and intuitive interface for exploring the data.

Finally, and most critically, is the task of intelligent filtering and recommendation. A simple search tool that just presents data is not enough to foster meaningful partnerships. The ultimate goal must be to move beyond single-attribute matching and provide multi-dimensional, intelligent recommendations. This requires building an analytical engine capable of a more holistic analysis, suggesting optimal partners based on a complex combination of factors. It is this third challenge—the development of an intelligent recommendation engine—that forms the primary focus of this thesis.

1.3 Research Question

To this aim, the following research question has been formulated:

1. How can a hybrid computational framework be designed to identify potential B Corp partners by integrating structured data filtering with semantic search capabilities powered by Large Language Models?

2. To what extent does this hybrid framework improve the overall quality of partnership recommendations, when evaluated against keyword-based search, in terms of relevance, rationale, and perceived user utility?

1.4 Scientific Contribution

The primary contribution of this project is to highlight the potential for collaboration among B Corps as a means to enhance their collective impact. While existing literature largely focuses on the certification process, the individual benefits of becoming a B Corps, and the role of B Corps in promoting sustainable development [8], there is limited emphasis on inter-firm collaboration as a strategic tool for achieving sustainability goals. This research shifts the focus toward collaborative opportunities by exploring how structured information, intelligent filtering, and data-driven approaches can facilitate more effective partnerships among B Corps. By addressing information accessibility and matchmaking challenges, this study contributes to both the academic discourse on sustainable business networks and the practical development of tools that support collaboration within the B Corp ecosystem.

In particular, this study makes significant contributions to the field of information retrieval by proposing and validating a technical framework designed to translate complex user needs into actionable recommendations. The research first empirically identifies the information dimensions B Corps consider most critical when evaluating partners. It then operationalizes these findings into a scalable support system, demonstrating how interdisciplinary methods from natural language processing and information science can bridge the gap between unstructured data and relevant search results in a real-world sustainability context.

Furthermore, this thesis contributes a novel hybrid architectural blueprint that combines the efficiency of a traditional search engine (Elasticsearch) with the deep semantic understanding of an advanced LLM. The rigorous evaluation of this model, using real-world data and metrics such as precision and recall, provides empirical evidence for its superiority over baseline approaches. This demonstrates a key design principle: that intelligent system integration is often more effective than relying on a single technology.

In doing so, the study offers more than just a solution for the B Corp community. It lays a foundation for future research on inter-organizational collaboration and provides a reproducible and generalisable model for building intelligent matching systems in other purpose-oriented business ecosystems.

We summarize our contributions as follows:

⋆ **Highlighting and Operationalizing Inter-Firm Collaboration**: This research provides a novel perspective by empirically demonstrating the untapped potential of collaboration among B Corps. It moves beyond theoretical discussion by designing a tangible system, thereby shifting the research focus from 'why' B Corps are important to 'how' they can work together more effectively.

⋆ **A Design Blueprint for Hybrid Partner Matching Systems**: This thesis answers the question of how to build an effective matching system for this context. It contributes a validated design pattern that demonstrates the necessity of a hybrid approach. We found that combining structured filtering with Large Language Model-based semantic search is critical for success. This blueprint offers specific, transferable knowledge about designing information systems that must balance explicit business logic with implicit, meaning-based relevance.

⋆ **Empirical Validation of an AI-Augmented Matching Approach**: In response to the question of what methods work best, this study provides empirical evidence of the proposed system's superiority. Through evaluation using precision and recall metrics, we demonstrate that our integrated framework significantly outperforms systems relying solely on traditional filtering. This finding validates that AI-driven semantic analysis is not just an add-on but a core component for achieving high-accuracy recommendations in complex, purpose-driven domains, offering a reproducible methodology for future research.

⋆ **The Power of Hybrid Architecture in System Design**: One of the key findings of this research is that the effectiveness of a core model, such as an advanced LLM, is significantly amplified by thoughtful engineering. The comparison between the Gemini-only system (in my project, which is Model 3) and the final hybrid system (Model 4) empirically demonstrates that integrating a specialized tool like Elasticsearch for structured search allows the LLM to focus on its primary strength of semantic understanding. This principle of a hybrid architecture—offloading specific tasks to the most suitable tool—is a critical design insight, proving that superior performance often comes from intelligent integration rather than relying on a single, monolithic model.

⋆ **Context-Dependent Advantages and Design Trade-offs**: This study further reveals that there is no single "best" model across all possible dimensions. While the final hybrid model (Model 4) achieved the highest overall performance, other models offer distinct advantages in specific contexts. The locally deployed system (Model 2), for instance, provides superior data privacy and security, making it the optimal choice for organizations with highly sensitive data. Similarly, the Gemini-only system (Model 3) offered the fastest response times, ideal for applications where low latency is critical. This highlights a crucial design trade-off between accuracy, performance, security, and cost, demonstrating that the optimal solution is highly dependent on the specific requirements of the deployment scenario.

2 RELATED SCIENTIFIC WORK

The academic discourse on B Corporations, while growing, is still in a nascent stage. A systematic literature review by Diez-Busto et al. (2021)[6], which analyzed 50 relevant studies up to 2020, confirmed this. Their findings revealed that existing research has predominantly focused on the motivations for seeking B Corp certification and the subsequent economic and social impacts on individual firms. The review concluded by highlighting several key gaps and calling for future research on topics such as the contribution of B Corps to the Sustainable Development Goals, case studies on the certification process, and methods for measuring social impact.

Building directly upon the research gaps identified by Diez-Busto et al.[6], this thesis addresses the underexplored area of inter-firm collaboration within the B Corp ecosystem. While their review called for better impact measurement, this study posits that enhancing collaborative efficiency through technological tools is a critical prerequisite for amplifying that collective impact.

In global value-driven organizations such as B Corp, efficient partner matching faces unique challenges. On one hand, member companies are closely connected by common sustainable development goals; on the other, geographical dispersion and information asymmetry result in a lack of in-depth understanding between

them. According to interviews with B Corp practitioners, finding partners with aligned values and complementary capabilities is a recognized difficulty.

This dilemma stems from multiple factors. First, traditional search engines, which rely on keyword matching, struggle to address complex "goal alignment" needs, such as deep cooperation on ESG indicators, corporate missions, or non-physical services. Second, mainstream recommendation algorithms often focus on superficial similarities like industry or region, limiting their applicability in value-oriented matching scenarios. More importantly, users' collaborative needs often evolve from vague to specific, placing higher demands on the system's intelligence and flexibility. Therefore, in the context of sustainable business, achieving efficient, accurate, and reliable partner matching based on users' dynamic goals has become a critical issue in the fields of recommender systems and information retrieval.

Currently, the official B Corp website only provides a filtering-based search function, where users obtain company lists through preset categories and keywords. While this method can ensure precision for well-defined queries, it is inherently passive and fails to proactively uncover potential collaboration opportunities. To overcome this limitation, early academic research explored recommendation technologies, notably Collaborative Filtering (CF).

Early recommendation systems often employed content-driven or collaborative filtering methods. CF, one of the earliest technical paths in recommended systems, emphasizes weighted predictions based on "similar users" or "similar items." However, as Pennock et al. (2000)[14] demonstrated from a social choice theory perspective, CF possesses inherent structural limitations in expressing individual, nuanced goals. Their framework, which formalizes CF as a "preference aggregation problem," reveals why such methods are ill-suited for scenarios like B Corp, where value alignment and accommodating innovative, cross-industry needs are paramount.

The next stage of recommendation models emphasized a more holistic understanding of user expression, focusing on the comprehension and matching of cooperation goals rather than just company attributes. Researchers began to enhance system sensitivity through techniques like multi-scenario modeling. Moving beyond traditional behavior-based similarity, researchers started to model recommender systems from the perspective of users' goals, tasks, or long-term motivations. Papadimitriou et al. (2018) [13] proposed a recommendation framework based on "goal-action" mapping. This model posits that a user's behavior sequence is a means to achieve certain underlying goals, rather than a series of isolated interest-based choices. By constructing a "goal-action set" correlation map and offering distinct strategies like Focus, Breadth, and Best Match, it caters to diverse user preferences.

This approach—identifying goal intentions from past behaviors to recommend future actions—is structurally different from collaborative filtering and is closer to the "task-driven matching based on cooperation goals" logic emphasized in my research. This logic is particularly suitable for complex situations with vague user goals or cross-category collaborations, providing a clear mechanism for "goal-sensitive" recommendations, especially when matching partners for ESG or industrial integration objectives. While related research introduced concepts like "intent embedding" or "multi-scenario processing" to help systems perceive user goals, they still fall short of understanding truly complex needs.

The rise of Large Language Models (LLMs) is bringing a paradigm shift to the recommender systems field. As noted by Wu et al. (2024) [20], LLMs, with their superior natural language understanding, reasoning, and generation capabilities, offer novel solutions to long-standing challenges such as the cold-start problem and lack of interpretability. The application of LLMs in recommendation primarily follows two paradigms: a fine-tuning model requiring domain-specific data, and a prompt engineering approach that directly guides the LLM.

This project focuses on the latter paradigm. It offers significant advantages, including high flexibility and independence from massive annotated datasets and intensive computational resources. This makes it particularly suitable for small and medium-sized enterprises to rapidly deploy and iterate on recommendation functions in general-purpose scenarios. Pioneering works in this area, such as the InstructRec framework proposed by Zhang et al. (2023) [21], reshape the recommendation task as "Instruction Following." Through sophisticated prompt design, LLMs can directly comprehend complex user needs and perform diverse tasks. Crucially, InstructRec's design inherently supports the generation of explainable recommendation reasons, meaning the LLM can not only "make" a decision but also "explain" the logic behind it.

Despite the broad prospects of LLMs in recommendation, existing research has limitations. Most studies either focus on using LLMs to enhance traditional retrieval models or employ them merely as general-purpose text generators. In collaborative matching scenarios that require synthesizing multi-party information, there is a research gap in how to systematically and practically engineer prompts to guide LLMs toward achieving both accurate matching and high-quality explanations. Therefore, the core contribution of this project, beyond using LLMs to analyze user input and generate reasons, is to design a reusable, engineering-focused prompt framework for cooperative matching to enable precise guidance of the LLM.

In terms of interaction design, I will adopt a conversational and explanatory recommendation system, allowing users to effectively articulate their goals in an open-ended format. As Lei et al. (2023)[10] argue, unlike traditional systems that work statically from past interaction history, a conversational approach enables more accurate and personalized recommendations by providing better user-system interaction.

Finally, most current LLM-based recommendation prototypes rely on heavy APIs or end-to-end model training, lacking a lightweight, deployable architecture suitable for prototyping. This project will investigate how to achieve optimal results by exploring engineering implementations and architectural designs for deploying LLMs on local or limited-database environments.

3 RESEARCH STRATEGIES AND RESEARCH METHODS

This research employs a pragmatic, time-constrained strategy combining a Design Science Research (DSR) methodology with a mixed-methods approach to address the challenge of facilitating B Corp

collaborations. The project, conducted between February and May 2025, prioritized the rapid and iterative development of a functional partner-matching system. The overall strategy involved four distinct phases: foundational data collection and baseline system development, enhancement with a local LLM, integration of an advanced cloud-based LLM, and final validation and refinement through user feedback.

3.1 Research Methodology: Design Science Research

This project is fundamentally anchored in the Design Science Research (DSR) methodology. DSR is a problem-solving paradigm that aims to advance human knowledge through the creation of innovative artifacts [19]. In this case, a sophisticated B Corp partner-matching system. This approach was chosen for its direct alignment with the research goals, as it supports the iterative design, implementation, and evaluation of a technological solution to a real-world problem.

An evolutionary prototyping approach was used to implement the DSR framework. The project progressed through a series of system artifacts, starting with a traditional classification system and advancing to a hybrid system integrating advanced language models. Each iteration was rigorously evaluated for its effectiveness, allowing for a comparative analysis that identified the most efficient and accurate solution to answer the research questions.

3.2 Research Procedure

The research was executed through a systematic and iterative process, following the principles of Design Science Research. The development unfolded across four distinct, sequential phases, evolving from a simple baseline system to a sophisticated, user-validated final artifact.

The project commenced with a foundational analysis of the official B Corp directory, which identified critical usability flaws such as a complex industry classification and unlinked search filters. To address these issues, an initial prototype (P1) was developed, featuring a hierarchical classification system and interconnected filtering logic. This served as an improved baseline for all subsequent development. To overcome the keyword-dependency of this first version, a second prototype (P2) was built, integrating a local LLM (Microsoft's Phi-2) and a vector database (ChromaDB) to enable semantic search. While this version successfully introduced a conversational interface, it also revealed limitations in recommendation accuracy and the quality of its explanations.

Building on these findings, the third phase focused on creating a more powerful hybrid system (P3) by integrating a cloud-based LLM (Google's Gemini). This iteration featured a refined data flow, a hybrid retrieval approach combining structured search with semantic matching, and advanced prompt engineering to generate well-reasoned, categorized recommendations. The final phase of the research was dedicated to qualitative validation and refinement. This P3 prototype was presented to four B Corp stakeholders in semi-structured interviews. The invaluable feedback gathered during these sessions prompted a final round of significant technical enhancements, most notably the integration of Elasticsearch to improve search robustness and the development of a template-guided interface to better meet user needs. After a final UI polish, the entire application was containerized using Docker to ensure its deployment readiness and scalability.

3.3 Data Collection and Analysis Methods

3.3.1 ***Data Collection.*** A two-pronged data collection strategy was employed. The primary quantitative dataset was acquired via web scraping of the official B Corp public directory. This method was necessary as no public API was available and it provided the foundational data (company profiles, industries, descriptions) needed to build and train the system. The primary qualitative data was collected through semi-structured interviews with four potential end-users. This method was chosen to gather in-depth, contextual feedback on the system's usability and practical value, which cannot be captured by quantitative metrics alone.

3.3.2 ***Data Analysis.*** The analysis methods corresponded to the type of data collected. The scraped quantitative data was subjected to data preprocessing, which included cleaning, normalization, and feature engineering, most notably the vectorization of textual descriptions for semantic analysis. The qualitative data from interview transcripts was analyzed using a thematic analysis approach. This involved identifying, coding, and interpreting recurring themes related to user needs, usability issues, and suggestions for improvement.

3.4 Evaluation Strategy

A comprehensive evaluation strategy was designed to assess the project's outcomes from both a technical and a user-centric perspective.

3.4.1 ***System Performance Evaluation.*** The technical performance of the system prototypes (Model 2, 3, and 4) was quantitatively evaluated to answer the research question concerning the most effective methods. This was achieved through a comparative analysis that measured and contrasted the models on several key dimensions. Computational efficiency was assessed using metrics of system response time and resource (CPU, memory) consumption. Recommendation quality was analyzed by observing the relevance and accuracy of responses to two standardized test queries, one broad and one specific.

3.4.2 ***Stakeholder Evaluation.*** The practical applicability and real-world value of the final system were assessed through qualitative stakeholder evaluation. This method was chosen because technical performance metrics alone do not capture user satisfaction or whether a system truly solves a user's problem. Through usability testing sessions and semi-structured interviews, direct feedback was gathered from four representative end-users. This evaluation focused on assessing the system's usability, the perceived quality of its interactions, and its overall potential to facilitate the partner discovery process in the B Corp ecosystem.

4 STUDY EXECUTION

4.1 Data Collection Methods

The project commenced with data acquisition from the official B Corp online directory (https://www.bcorporation.net/en-us/find-

a-b-corp/). This directory was selected as the primary data source for its comprehensiveness and authority. It is the official registry, containing all certified B Corporations, and is the primary public resource used by the B Corp community for partner discovery. The directory provides the essential structured and unstructured information required to build a robust recommendation system.

To gather this data, a Python-based web scraping script was developed and executed in February 2025. The script systematically collected key attributes for each company, including: Headquarters, Certified Since, Industry, Sector, Operating Regions, Website, Overall B Impact Score, and Company Description.

Following the data collection, a critical preprocessing phase was conducted to prepare the raw data for analysis. This phase involved cleaning, structuring, and normalizing the data to ensure consistency and system compatibility. Crucially, the unstructured text from the Company Description field was vectorized using natural language processing techniques. This process converted the qualitative descriptions into a quantitative format, enabling the semantic analysis required for the advanced recommendation models in later phases. The final output of this phase was a clean, structured database stored in a JSON file, which formed the foundational layer for all subsequent system development and evaluation.

4.2 Data Analysis

4.2.1 ***System Design and Development.*** **Iterative development approach:** Progressive enhancement of the system through four distinct methodological phases. the development of the system followed a clear, iterative trajectory across four distinct phases. Phase 1 established a foundational system (P1) that addressed the critical usability flaws of the official directory by introducing a hierarchical classification and interconnected filtering. Phase 2 advanced this baseline by integrating a local LLM and vector database (P2), enabling semantic search but revealing limitations in accuracy and reasoning. Phase 3 overcame these issues by creating a sophisticated hybrid system (P3) that leveraged a powerful cloud-based LLM (Gemini) for superior analysis and response generation. Finally, Phase 4 closed the loop by incorporating direct user feedback from interviews, which led to the integration of Elasticsearch for robust search, significant UI enhancements, and the final containerization of the system, resulting in a user-validated and deployment-ready artifact.

Phase 1 - Traditional classification and search methods: The initial phase of the project began with a critical analysis of the official "Find a B Corp" website to identify existing usability problems. This analysis revealed two primary issues that hinder effective partner discovery. First, the industry classification was overly complex and unintuitive. The original system includes more than 100 distinct industry categories. This granularity forces users into a time-consuming navigation process; for example, locating the "Sporting Goods" category could require over 20 clicks. The alphabetical arrangement does not solve this, as users may be unsure of the exact terminology used, such as whether to search for "sporting goods" or "physical exercise." Such inconsistencies make it difficult for users to reliably locate the appropriate industry.

Second, the system lacked interconnected filters. Selecting a specific industry does not dynamically influence the options available in other filters. For instance, if a user selects "agricultural products," the location filter should ideally update to display only regions where such companies are present. The original website does not implement this logic, often leading users to empty result pages.

Based on these identified issues, the first version of the system was developed with specific functional adjustments. To address the classification problem, the database was optimized by implementing a hierarchical industry classification system. This system organizes industries into three levels, ranging from general to more detailed categories. The top tier includes 11 major categories: Construction and Real Estate; Consumer Goods and Retail; Education and Culture; Food and Agriculture; Health and Healthcare; Manufacturing; Others; Professional Services; Sustainability and Environment; and Technology and Digital. This layered structure enables users to perform more intuitive and targeted filtering.

Building on this improved data structure, a platform was developed using the React framework for the front end and Node.js for the back end, with a MongoDB database and a RESTful API for data access. The system implements several key features for user filtering and results display. Users can filter by company name with exact or fuzzy matching, select from the hierarchical industry categories, or filter by country and region, distinguishing between operating regions and headquarters. The filtered results are then presented in a clear table that includes the company name, industry, region, B Corp score, B Corp website, and the company's official website. Additionally, the interface includes a "quick filter" function, allowing users to click on a location card to see all companies in that area, and a "one-click reset" button to clear all active filters.

Together, these functions form a comprehensive system for screening B Corp companies, allowing users to quickly identify partners that meet specific criteria. The main advantage of this foundational system is that it directly addresses the problems of disorganized classification and unlinked filters. However, a key limitation remains: it is not an ideal solution for a truly user-centric search, as users who do not input the correct, pre-defined classification keywords may still struggle to find what they are looking for. To address this limitation, the next phase of development focused on integrating a large language model to support more accurate and flexible search functionality.

Phase 2 - Local language model implementation: To overcome the keyword-dependency limitations of the foundational system, the second development phase focused on implementing semantic retrieval capabilities. This resulted in a second prototype, which integrated a local Large Language Model and a vector database.

The first step in this phase was to embed the corporate data. The company descriptions from the database were processed using a pretrained SentenceTransformer model. This model vectorized each description, converting the unstructured text into high-dimensional embedding vectors suitable for semantic comparison. Concurrently, a complete record structure was established for each company, linking the newly generated vector to its original text, metadata, and a unique ID. This structured approach is essential for efficient indexing and retrieval.

After generating the embeddings, the vector database ChromaDB was employed to store the vectors along with their corresponding

text and metadata. This setup enables rapid, similarity-based retrieval. A semantic search module was then developed. This module takes a user's natural language query, transforms it into an embedding vector using the same SentenceTransformer model, and then queries ChromaDB to retrieve the most semantically relevant company records by calculating vector similarity.

For tasks requiring natural language understanding, such as query intent analysis and generating explanations for the recommendations, a local LLM, Microsoft's Phi-2, was integrated into the system. The choice of Phi-2 was driven by its suitability for local deployment on a personal computer. To optimize performance, a quantized version of the model (Q4.K.M.gguf) was used, significantly reducing memory usage. The model was loaded and operated using the llama-cpp-python library. This architecture created a clear division of labor: SentenceTransformer and ChromaDB handled the core semantic search, while Phi-2 provided the interpretive layer.

Based on this technical stack, a dialogue-based user interface was implemented on the same front end as the previous version. Users could now input their queries in natural language, such as "Find me a company to collaborate with," and receive automatically generated results. This version represented a significant functional improvement, freeing users from the constraints of specific keywords and allowing for a more intuitive and efficient search experience.

However, evaluation of P2 revealed two persistent problems. First, the accuracy of the results was not always ideal. For instance, a query for "a company to supply food to KLM" (a Dutch airline) returned a UK-based company as the top recommendation, indicating a failure to correctly weigh geographical proximity. Second, the reasoning provided for recommendations was overly generic and repetitive, often stating merely: "This company matches your search criteria based on its industry and location." This lack of specific, insightful justification led to the decision to explore whether a more advanced model could produce superior results, thus motivating the progression to the next stage of development.

Phase 3 - Remote Google LLM integration: The third iteration of system development was aimed at overcoming the limitations of the local model by integrating a more powerful, state-of-the-art Large Language Model. This led to the creation of the this hybrid prototype, which utilized the Google Gemini model via its API. The decision to adopt Gemini was based on three primary factors: its consistently high performance on public benchmarks, the availability of a well-documented developer platform, and a generous free-tier quota that facilitated extensive testing without significant financial cost.

In this phase, the overall system architecture was refined for clarity and efficiency. A minimalistic, dialogue-based front end was implemented using the NickGUI framework to simplify the user experience. The data flow logic was also formalized into four distinct steps: User Input, LLM-driven Query Interpretation, Hybrid Data Retrieval, and LLM-driven Response Generation.

The most significant architectural improvements were in the data retrieval and response generation stages. For data retrieval, a hybrid approach was engineered. The system first used the LLM to process the user's natural language query and extract structured search conditions, such as keywords, industry, and region. It then performed a structured search combined with fuzzy matching on company descriptions and names, assigning relevance scores to the results and ranking them accordingly. This hybrid method leverages both the precision of structured search and the flexibility of semantic matching.

For the LLM-driven analysis and generation, prompt engineering was a key technique. The base prompt was formulated as: "You are an assistant helping to analyze questions about B Corp certified companies." Within this context, the system was designed to extract three core components from user queries—industry, region, and additional keywords—which were then used to construct the search parameters.

The generated responses were strategically categorized into three recommendation types to provide users with diverse options: companies prioritized by geographical relevance, companies with a strong product or service match, and companies selected as an overall "best fit" based on a multi-factor analysis. This categorization was inspired by the core filtering logic of the official B Corp platform and was later validated by user interviews, which confirmed that industry and location are paramount criteria for partner selection. To enhance usability, auxiliary features such as clearing the search input and printing conversation history were also introduced.

This third round design resulted in a marked improvement in overall system quality. First, search accuracy was significantly enhanced as the system now considered multiple dimensions, including location, industry, company scale, and partnership goals. Second, the chat interface proved to be highly intuitive and user-friendly. Finally, and most importantly, the system could now provide informative and interpretable matching reasons, making the results more transparent and easier for users to evaluate.

Phase 4 - Remote Google LLM integration with further improvement: With a functionally complete prototype (P3) in place, the final phase first focused on qualitative validation through user interviews and a subsequent round of system refinements. Four potential users, all deeply engaged in the B Corp ecosystem as managers or founders, participated in the evaluation. Each 30-minute session included a 10-minute software demonstration followed by a semi-structured interview to gather feedback on the system's usability, relevance, and overall value.

The interviews surfaced three key suggestions for improvement: the ability to iteratively refine search criteria during a dialogue, the option to input information about the user's own company to improve match quality, and the need for more detailed company information, including contact details, post-recommendation.

These insights prompted a final, significant revision of the system. While providing contact information was not feasible due to dataset limitations, the other suggestions were addressed. A template-guided search interface was developed, allowing users to input structured information about their own company and their ideal partner profile. All fields in this template were kept optional to maintain flexibility. Additionally, a "More" button was implemented, which connects to a new API endpoint to provide users with extended details and recommendation justifications for any selected company.

A major technical enhancement was also made to the search logic. To resolve issues with strict keyword matching, such as spelling errors or alternate country names, Elasticsearch was integrated into the back-end architecture. The integration of Elasticsearch brought multiple advantages, including support for full-text search,

fuzzy matching capabilities to handle typos, the ability to perform complex multi-field querying, and high scalability and speed.

The updated search pipeline now operates as follows: first, the user's natural language input is semantically analyzed. Second, Elasticsearch performs a preliminary screening, returning the top 50 results based on a relevance score. Finally, these candidate results are passed to the Gemini model, which uses prompt-based generation to produce detailed recommendation explanations and re-ranks them for final display to the user.

This final hybrid design, combining a traditional search engine with an advanced LLM, further enhanced both the accuracy and intelligence of the recommendation process. Finally, based on user feedback, the front-end interface was polished. The dialogue now uses distinct background colors to differentiate user and chatbot messages, and other visual cues like hover effects were added. An introductory page and a feedback questionnaire were also included to support broader use. The entire application was then containerized using Docker, ensuring ease of deployment, maintenance, and scalability. This refined and containerized system represents the final artifact of this research project.

4.2.2 ***Agile Development Philosophy and Project Trajectory.*** Although this project was developed independently, its structure was guided by agile development principles. The overall task was divided into multiple short-cycle iterations, each centered on delivering a specific feature or module. This iterative approach, with reviews at the end of each cycle, allowed for flexible problem-solving and enhanced responsiveness to emerging challenges and user feedback.

The evolution of the project followed a clear and logical trajectory. The initial version (P1) implemented an improved keyword-based search. The second stage (P2) introduced natural language support with a basic LLM. The third version (P3), refined in the last phase, integrated a powerful search engine with an advanced LLM to perform sophisticated hybrid retrieval and generate rich, contextual recommendations. This progressive development approach ensured that each version built upon the lessons of its predecessor, ultimately resulting in a robust, intelligent, and user-validated system.

4.3 Evaluation Methods

4.3.1 ***System Performance Evaluation.*** To assess the effectiveness and feasibility of the proposed recommendation system, a multi-dimensional evaluation framework was adopted, encompassing both quantitative performance and qualitative relevance. This evaluation aims to determine the extent to which the system can accurately and efficiently identify potential B Corporation partners aligned with specific sustainability goals.

First, a comparative analysis was conducted across four system iterations: a baseline system based on traditional keyword matching, a local embedding-based model, a Google LLM-integrated version, and a version using the same model but enhanced with an improved search engine. In this comparison, differences in model deployment methods, ease of deployment, completeness of the answers, and performance in response to two frequently asked questions were examined. A qualitative assessment was also carried out to evaluate the semantic relevance of the recommendations. This involved analyzing whether the suggested companies share similar impact goals, industry contexts, or values, thereby supporting the research objective of fostering regenerative and meaningful collaborations. Second, precision and recall metrics were used to quantitatively assess retrieval quality. By comparing system-generated recommendations with a reference set of relevant companies, the evaluation directly reflects the system's capability to support goal-aligned matchmaking. Finally, computational efficiency was assessed by measuring system response time and resource consumption. These aspects are essential for evaluating the system's practical viability for real-time or large-scale deployment.

The system was tested using two representative queries. The first question was relatively broad: "I want to import food from China." This query lacked specificity, such as the user's location, and therefore represented an open-ended and divergent use case. The second question was derived from a real-world scenario raised during an interview and was more practical and focused: "I want to supply lunch food for passengers on the plane for Dutch Royal Airlines in the Netherlands." Compared to the first, this question had a narrower scope and clearer contextual constraints.

To evaluate system performance, three of the four previously mentioned versions were tested with these two queries. For clarity, the systems are referred to as follows: Model 1, the baseline system based on manual keyword querying; Model 2, a locally deployed model using semantic embeddings; Model 3, the version incorporating the Gemini 1.5 Flash model; and Model 4, the enhanced version using the Gemini 1.5 Flash model combined with an optimized search engine. In this comparison, the results from Model 1 were excluded, primarily because its search process relies entirely on manual input, making it fundamentally different from the other three automated approaches. In practice, Model 1 is better suited as a complementary tool. Specifically, when users have a clear search target—such as known keywords or company names—Model 1 offers a direct solution. In contrast, for more ambiguous queries or when users are seeking suggestions, Models 2, 3, and 4 are more appropriate.

In terms of answer structure, Models 3 and 4—both powered by advanced large language models—demonstrate clear organization and well-articulated recommendation rationales. These models are able to directly connect the recommendations to the user's query. Model 3 tends to generate more verbose responses, while Model 4 produces more concise answers and includes a "further reading" function that allows users to explore additional information. Model 2, by contrast, generates brief outputs that merely list the company name, region, and industry. These responses lack depth, analytical reasoning, or comparative value. Moreover, some of its recommended companies lack websites or detailed descriptions, making it difficult for users to assess their relevance. From a deployment perspective, Model 2 has the advantage of being locally hosted, eliminating the need to consider external API quotas or costs. However, its deployment process is relatively complex. In contrast, Models 3 and 4 are much simpler to deploy.

In terms of the actual recommendation results, Models 3 and 4 demonstrate relatively high accuracy. Model 4 stands out for its ability to accurately interpret and control for geographical location and offers more complete and relevant recommendations. Model 3 exhibits higher recommendation precision in the second query.

Specifically, for the first question, two out of the five recommendations produced by Model 2 were only marginally related to the food industry or the import of food from China. Among the five results generated by Model 3, the fifth recommendation concerned innovation management—though tangentially related to the food industry, it was distant from the supply sector. In contrast, Model 4's filtering results largely met the requirements. Notably, Model 3 seemed to interpret the query more from a supply chain perspective, while Model 4 focused more on food production in China. Furthermore, Model 4 demonstrated a stronger grasp of geographical factors than Model 3.

Regarding the second question, all five recommendations from Model 2 belonged to the food industry, though one was in the beverage sector in the UK and thus less relevant. Model 3's results all related to either the food industry or the supply chain, but only two were located in the Netherlands, reflecting the earlier-noted regional recognition limitation. Model 4 returned ten results, which were relatively comprehensive; however, the last two responses were less relevant in terms of geographical location.

The performance comparison reveals that Model 3 achieves the fastest response times with moderate CPU and memory usage, making it the most efficient overall. Model 4 demonstrates the lowest CPU and memory usage, indicating stable and lightweight performance, though with slightly longer response times than Model 3. In contrast, Model 2 consumes significantly more CPU resources and time, despite its low memory usage, suggesting it may not be suitable for time-sensitive or resource-constrained environments. I tested the result on my personal laptop, which will influence the performance of Model 2.

4.3.2 ***Stakeholder Evaluation***. To evaluate the system's usability and real-world applicability, feedback was gathered from four potential end-users, formal usability testing sessions were conducted with the third prototype to assess the clarity and accessibility of the interface. These methods helped ensure the system aligns with actual user needs and supports informed decision-making in collaborative contexts.

User feedback indicated a high level of satisfaction with the system, with all four participants expressing positive impressions. Regarding the front-end interface, all users, who were experienced with AI tools, were already accustomed to chat-based interactions and found the interface clean and intuitive. One participant suggested differentiating the background colors of user and AI messages to improve readability, a recommendation that was subsequently implemented.

From an interaction perspective, users expressed a desire for more dynamic exchanges, including the ability to initiate multiple rounds of interaction, supplement information iteratively, and view extended content. While an attempt was made to implement this functionality, it was ultimately unsuccessful due to technical constraints. As a compromise, structured templates were introduced to support clearer input of user requirements and enable access to more detailed information.

In terms of content, users showed significant interest in obtaining more comprehensive company data, particularly contact information such as email addresses. However, since the project is built on a closed, pre-scraped database rather than a live web-crawled corpus, this type of information is currently unavailable.

In summary, the users affirmed the project's value and practicality. Their feedback provided crucial validation and helped identify key directions for future development, particularly in enhancing system interactivity and expanding data coverage.

COMPARISON OF RECOMMENDED COMPANIES BY MODEL

Model 2: B Corp Explorer

Table 1: Model 2 Recommended Companies

Company Name	Industry	Location	Summary
Dongshi Sunflower Beijing Information Technology Co	Food & Agriculture	Beijing, China	Chinese-based food company identified via B Corp Explorer, though details are minimal.
Fly By Jing	Food & Agriculture	California, USA	Popular US-based brand inspired by Chinese flavors; location not ideal for direct import from China.
Danone Eln Greater China	Food & Agriculture	Shanghai, China	Major food manufacturer with strong operations in China.
Withheart Meat Shop Co Ltd	Food & Agriculture	Taichung, Taiwan	Meat-focused producer.
Ju Tian Cleantech Co Ltd	Others	Nantou County, Taiwan	Cleantech company, not directly related to food import.

Model 3: Structured Match

Table 2: Model 3 Recommended Companies

Company Name	Industry	Location	Summary
Good Food Technologies	Plant-based Foods	Hong Kong	Plant-based innovator with operations in Guangzhou; ideal for importing from Greater China.
Meelunie	Food Ingredients	Amsterdam, Netherlands	Global ingredient supplier with operations in China; strong B Corp credentials.
Boiron Frères	Fruit Products	Chateauneuf, France	Premium fruit puree producer with China distribution network.
Openway Food Co. Pty. Ltd.	Packaged Foods	Australia	Works with global clients including China; focuses on sustainable food innovation.
Wood for Trees Design	Innovation Consultancy	Bristol, UK	Consultancy with food innovation focus; indirect but experience with China-based clients.

Model 4: Template-based Regional Recommendation

Table 3: Model 4 Recommended Companies

Company Name	Industry	Location	Summary
Danone Waters China	Food Products	China	Established food manufacturer with strong Chinese market presence and supply chain.
Good Food Technologies	Plant-based Foods	Hong Kong	Asian food innovator with extensive plant-based product line and China ties.
Nutricia AMN Greater China	Medical Nutrition	China	Operates across China; good candidate for importing niche nutrition products.
LiLi Time	Restaurant / Food Service	Shanghai, China	Focuses on sustainability; potential for niche food collaborations.
Camps Food	Wholesale / Retail	Breda, Netherlands	Netherlands-based distributor, could assist with local import logistics.

Table 4: Model 2 Recommended Companies

Company Name	Industry	Location	Summary
Albert Heijn Nederland	Food & Agriculture	North Holland, Netherlands	Large supermarket chain; strong logistics and supply chain, suitable for distribution.
Fresh Food Ventures Group Bv	Food & Agriculture	North Holland, Netherlands	Operates in food ventures; exact scope unclear due to limited public data.
John Altman	Food & Agriculture	North Holland, Netherlands	A health food/snack producer; targeting healthy meals.
Double Dutch Drinks	Food & Agriculture	London, UK	Beverage-focused company; less relevant unless drinks are required.
Camps Food	Food & Agriculture	North Brabant, Netherlands	Distributor for food and beverage brands; could help connect to local markets.

Model 2: B Corp Explorer

Model 3: Structured Match

Table 5: Model 3 Recommended Companies

Company Name	Industry	Location	Summary
Appèl Holding BV	Catering & Food Services	's-Hertogenbosch, Netherlands	Large-scale Dutch caterer with expertise in institutional and event food service.
De Groene Afslag Holding B.V.	Event Catering & Plant-based Food	Laren, Netherlands	Focus on sustainable, plant-based catering; smaller size may limit scale.
Wilson Vale Catering	Contract Catering	Ashby de la Zouch, UK	Experienced in craft catering for schools/offices; emphasizes sustainability.
Monty's Bakehouse UK Ltd	In-flight Food Solutions	UK & Netherlands	Specializes in aviation food products; strong in sustainability and innovation.
De Smaakspecialist	Organic Food	North Brabant, Netherlands	Develops and distributes organic food; aligned with health and sustainability goals.

Model 4: Template-based Regional Recommendation

Table 6: Model 4 Recommended Companies

Company Name	Industry	Location	Summary
Crisp	Online Supermarket	Netherlands	B Corp certified fresh-produce e-supermarket with strong local sourcing—ideal for integrating imports.
Appèl Holding BV	Catering & Food Services	's-Hertogenbosch, Netherlands	Large, sustainable caterer with institutional & event food production capacity.
Albert Heijn Nederland	Retail Supermarket	North Holland, Netherlands	Major supermarket chain with extensive supply network for distributing imported products.
Camps Food	Food & Beverage Distribution	Breda, Netherlands	B Corp distributor for food & beverage brands in Benelux, useful for B2B logistics.
Monty's Bakehouse UK Ltd	Food Manufacturing	UK & Netherlands	Sustainable, aviation-targeted food solutions—experienced in inflight meal production.
De Smaakspecialist	Organic Food Producer	North Brabant, Netherlands	Develops/distributes organic & nutritious food; aligned with healthy sourcing.
Vanhulley	Apparel (Sustainability)	Groningen, Netherlands	Though not food, exemplifies sustainable operations—possible eco-brand partnership.
The Makers Food GmbH	Food Manufacturing	The Netherlands	Produces sustainable plant-based food products for B2B/B2C markets.
The Organic Family Ltd	Organic Pantry Supplier	UK/Netherlands	Offers organic Italian pantry staples; strong supply chain for quality ingredients.
Nutricia AMN Greater China	Medical Nutrition	China	Large nutrition manufacturer in China; potential for importing specialized food products.
LiLi Time	Restaurant / Food Services	Shanghai, China	Shanghai-based sustainable food service; possible partner for sourcing imported foods.

PERFORMANCE BENCHMARKING OF MODELS

Table 7: Performance Comparison of Different Models Across Two Test Questions

Model	Question	Response Time (s)	CPU Usage Change (%)	Memory Usage Change (MB)
Model 2	Q1	50.51	119.20	1.88
Model 2	Q2	44.56	128.80	0.00
Model 3	Q1	11.13	5.40	50.82
Model 3	Q2	9.03	12.80	4.72
Model 4	Q1	13.87	0.60	15.84
Model 4	Q2	12.49	1.10	9.95

5 DISCUSSION

5.1 Discussion of the results

The first research question asked how a hybrid computational framework could be designed to effectively identify B Corp partners. This study answers this question by proposing and validating a multi-stage, progressively enhanced system architecture. The research demonstrates that an effective framework must address several key information science challenges:

First, it requires a well-structured and classified data foundation. The initial development phase (P1) showed that replacing the original website's chaotic classification with a clear, hierarchical system is a prerequisite for any meaningful search, significantly improving the intuitiveness of the screening process. Second, an effective framework should support diverse and flexible query methods. The system evolved from providing user-friendly, multi-criteria filtering (P1) to supporting natural language input via semantic search (P2). This allows users to express complex, nuanced collaboration needs that go beyond simple keyword matching. Finally, the most effective design, validated in the final prototype (Model 4), is a synergistic hybrid architecture. This framework integrates a traditional search engine (Elasticsearch) for its precision in handling structured data like geographic location, with an advanced LLM (Gemini) for its deep semantic understanding of unstructured text. The LLM not only interprets user intent but also provides reasoned, transparent explanations for its recommendations. The importance of these design elements—particularly the focus on industry, region, and shared values—was consistently validated through user interviews.

The second research question assessed the extent to which this hybrid approach improves the effectiveness of recommendations compared to baseline methods. The comparative analysis of the four models provides a clear answer: the final hybrid approach (Model 4) is demonstrably superior across multiple dimensions.

In terms of recommendation accuracy and relevance, Model 4 demonstrably outperformed all other models by successfully mitigating their specific weaknesses. The purely semantic approach of Model 2, for example, sometimes struggled with contextual nuance, occasionally misinterpreting keyword similarity. A notable instance was its recommendation of a company focused on "sustainable finance" when the query was for "sustainable food," simply because both contained the word "sustainable." Similarly, the LLM-only approach of Model 3, while having strong semantic capabilities, frequently failed on precise, factual constraints like geography. For instance, when asked for companies in the Netherlands, it often included UK-based companies in its recommendations, likely because its embedding space places these two nearby countries in close proximity.

Model 4's hybrid architecture directly solves these problems. The integration of Elasticsearch provides robust, precise filtering that resolves the geographical identification issues that plagued Model 3. Simultaneously, the semantic capabilities of the LLM ensure that recommendations are more contextually relevant than those from the keyword-dependent Model 1 or the sometimes-confused Model 2. This synergy was evident in both the broad and specific test queries, where Model 4 consistently provided more complete and on-target results.

Regarding perceived user utility, qualitative feedback gathered from semi-structured interviews with four B Corp stakeholders provided crucial insights. The overall reception of the system, demonstrated through an interactive session with Model 3, was highly positive.

Participants affirmed that the system directly addresses a significant real-world challenge they face: the difficulty of finding collaboration partners. They highlighted that the official B Corp website's reliance on rigid industry tags made discovering potential partners inefficient. In contrast, they praised the system's ability to accept natural language queries, which they found to be a far more intuitive and effective method for expressing their complex needs. Furthermore, users commented on the simplicity and usability of the interface, noting that the dialogue-based design was straightforward and required minimal guidance to operate. Finally, the quality of the information provided was deemed high. During the live demonstrations, the system was tested with real-life problems faced by the participants, who confirmed that the recommendations were "insightful" and provided valuable, actionable information. This feedback suggests that both Model 3 and the subsequent Model 4 are capable of delivering significant value to end-users.

At the same time, these interviews were instrumental in identifying key areas for improvement, which directly guided the development from Model 3 to Model 4. Participants expressed three primary needs that were subsequently addressed: First, the ability to provide information about their own company, in addition to their collaboration request, to achieve a better match. Second, a desire for more accurate and reliable location-based searching. Third, a minor UI suggestion to use different colors to distinguish between user and chatbot messages for better readability. These three suggestions were successfully implemented in the final iteration, Model 4.

However, the stakeholders also raised two valuable suggestions that were beyond the scope of this project's timeline. First, they requested the inclusion of company contact information. This was not implemented because the system operates on a closed, prescraped database that does not contain this data, and introducing open-source web crawling was deemed out of scope. Second, users expressed a strong interest in a multi-turn dialogue capability to iteratively refine their search queries. While this more advanced conversational feature was not successfully built due to time constraints, a template-based input structure was introduced in Model 4 as a practical compromise to help users formulate more structured and effective initial queries. Nevertheless, a true multi-turn dialogue remains a superior future goal.

While the hybrid approach (Model 4) proved most effective overall, the evaluation also revealed that the "best" implementation is highly context-dependent, highlighting several important design trade-offs. For instance, the local model (Model 2) offers superior data privacy and security. In a real-world scenario where a company wants to find partners for a confidential project, they would be hesitant to upload sensitive information, such as a new business plan or proprietary technology details, to a third-party cloud-based LLM. In such cases, a locally deployed model that processes all data offline would be the only viable and secure option, even at the cost of lower recommendation quality. Similarly, the simpler LLM integration (Model 3) demonstrated the fastest response times, making it a more suitable choice for applications where low latency is the

primary concern over pinpoint accuracy. This indicates that while the final hybrid model provides the highest quality recommendations, the optimal solution for a given organization will always depend on a careful balancing of specific contextual constraints, including security needs, latency requirements, and operational costs.

5.2 Limitations

This research, despite its findings, is subject to several limitations in both its system design and evaluation methodology.

In terms of system design, there are inherent constraints. First, the system's conversational capabilities are limited; it does not support multi-turn dialogue where context is maintained across several interactions. Each query is treated as a standalone transaction, and the system's responses are constrained to the template of providing "recommended companies." Although an earlier prototype (Model 3) attempted to answer general questions, this feature was removed in the final version (Model 4) due to inconsistent accuracy.

Second, the use of an LLM introduces a "black box" effect. While the system provides explanations for each recommendation, thereby enhancing transparency, it remains difficult to fully understand why certain companies were not selected. The interpretability of the overall screening logic is therefore limited.

Third, the system was designed and tested for local or lightweight cloud deployment. Its performance under conditions of high concurrency or global-scale usage has not been evaluated, which constrains its proven applicability in large-scale operational scenarios.

At the evaluation level, the quantitative metrics are subject to certain constraints. The use of precision and recall is limited by the inherent subjectivity in the manual curation of the "ground truth" partner list. Moreover, due to the vast number of potential candidates, recall could not be measured with complete accuracy. Furthermore, the small sample size of four participants in the qualitative user interviews may not fully capture the diverse range of collaboration needs across the entire B Corp ecosystem.

Finally, two broader contextual limitations must be acknowledged. Conceptually, inter-organizational collaboration is an inherently multi-dimensional and human-centric process. Matching based on database criteria can only indicate the likelihood of a viable partnership, not guarantee it. The system should therefore be viewed as a tool to initiate and facilitate partnerships, not to definitively determine them. Technologically, for the specific B Corp case which relies on publicly available information, general-purpose models like ChatGPT with integrated web-search capabilities can now perform similar partner-finding tasks with considerable effectiveness. The unique value of the framework developed in this thesis, therefore, lies not in its application to public data, but in its potential for deployment on private, proprietary, or domain-specific knowledge bases where data privacy and specialized search logic are paramount.

5.3 Generalisation

Building upon the distinction made in the previous section, the primary generalisable contribution of this research is not a tool for public web search, but rather a robust framework for querying private and domain-specific knowledge bases. While general-purpose models excel at retrieving information from the open internet, the true value of the hybrid architecture developed in this thesis lies in its applicability to internal, curated datasets where data privacy, query specificity, and the ability to leverage proprietary data structures are paramount. This research, therefore, provides a transferable blueprint for empowering organizations to unlock the value hidden within their own data ecosystems.

The B Corp partner-matching problem served as a compelling case study to design and validate this framework. From a business matching perspective, the architectural pattern of combining structured data filtering with unstructured semantic matching is highly transferable. This model can be applied to other complex commercial scenarios where a holistic evaluation is essential. For instance, in B2B supplier discovery, a large corporation could leverage this approach to sift through thousands of potential suppliers, filtering not only by structured criteria like price and location but also performing semantic analysis on unstructured data such as technical documentation to find the most "synergistic" partner.

At the methodological level, the hybrid information retrieval architecture validated in this project, which pairs Elasticsearch with an advanced LLM, presents a powerful and effective paradigm. This model can be widely applied to any information system that needs to balance the search efficiency and recall of a traditional search engine with the deep semantic understanding and relevance provided by an LLM.

However, the successful generalisation of this framework is contingent upon two key prerequisites. First, the target domain must possess an accessible database containing both structured and unstructured information. Second, the matching problem itself must be decomposable into clear, definable dimensions, as these are essential for designing and training the system effectively.

In conclusion, although this project was instantiated through the B Corp case, its primary contribution is a transferable, data-driven blueprint for solving complex entity-matching problems within an organization's own private, networked ecosystems.

6 CONCLUSION

This thesis set out to address the challenge of facilitating strategic partnerships within large, value-driven business networks, using the B Corp ecosystem as a case study. By employing a Design Science Research approach, this project successfully designed, built, and validated a hybrid information retrieval system that demonstrates a significant advancement over existing partner discovery tools.

The primary contribution of this research is a validated architectural blueprint that effectively combines a traditional search engine (Elasticsearch) with a state-of-the-art Large Language Model (Gemini). The comparative evaluation of four distinct system models revealed that this hybrid approach is superior in terms of recommendation accuracy, relevance, and computational efficiency. It successfully balances the precision of structured filtering with the deep contextual understanding required to match entities based on nuanced, unstructured data such as shared values and mission

statements. Qualitative feedback from B Corp stakeholders further confirmed the system's practical utility and user-friendly design.

While this study has inherent limitations, including the "black box" nature of LLMs and the need for further testing at scale, it provides a strong foundation for future work. Future research could focus on enhancing the system's interactivity, expanding the dataset to include more dynamic information, and refining the recommendation logic.

Ultimately, this project demonstrates that modern AI, particularly the thoughtful integration of LLMs with established information retrieval techniques, can provide powerful solutions to complex matching problems. Although instantiated within the B Corp community, the methods and principles developed here offer a generalisable framework for any organization seeking to unlock the collaborative potential hidden within its own knowledge bases, fostering smarter, purpose-driven connections in an increasingly interconnected world.

7 ACKNOWLEDGMENT

I would like to thank Dr. de Boer and Dr. van Aart who made this work possible, their guidance and advice carried me through all the stages of writing my thesis. I would also thanks to the stakeholders who spent time participating the interview.

I would also thanks to my working team, thanks for keeps tracking and supporting.

8 APPENDIX

project code: https://github.com/yitong521/agent4/settings

demo video: https://youtu.be/tcowqxA41p4

REFERENCES

[1] B Lab. [n.d.]. B Corporation. https://www.bcorporation.net/en-us/. Retrieved January 31, 2025.

[2] B Lab Europe. [n.d.]. B Corp Global Climate Summit. https://bcorporation.eu/event/b-corp-global-climate-summit/. Retrieved January 31, 2025.

[3] Ke Cao and Joel Gehman. 2021. Certified B corporations and benefit corporations.

[4] Lujie Chen et al. 2017. Supply chain collaboration for sustainability: A literature review and future research agenda. *International Journal of Production Economics* 194 (2017), 73–87.

[5] H. Cotacallapa, N. Saboya, P. C. Rodrigues, R. Salas, and J. L. López-Gonzales. 2024. A flat-hierarchical approach based on machine learning model for e-commerce product classification. *IEEE Access* (2024).

[6] E. Diez-Busto, L. Sanchez-Ruiz, and A. Fernandez-Laviada. 2021. The B Corp movement: A systematic literature review. *Sustainability* 13, 5 (2021), 2508.

[7] I. Escher and P. Brzustewicz. 2020. Inter-organizational collaboration on projects supporting sustainable development goals: the company perspective. *Sustainability* 12, 12 (2020), 4969.

[8] Joel Gehman, Matthew G. Grimes, and Ke Cao. 2019. Why we care about certified B corporations: From valuing growth to certifying values practices. *Academy of Management Discoveries* 5, 1 (2019), 97–101.

[9] V. Gupta, H. Karnick, A. Bansal, and P. Jhala. 2016. Product classification in e-commerce using distributional semantics. *arXiv preprint arXiv:1606.06083* (2016).

[10] He-X. de Rijke M. Chua T. S. Lei, W. 2020. Conversational recommendation: Formulation, methods, and evaluation. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (2020), 2425–2428.

[11] J. Li, Z. Dou, Y. Zhu, X. Zuo, and J. R. Wen. 2020. Deep cross-platform product matching in e-commerce. *Information Retrieval Journal* 23 (2020), 136–158.

[12] J. Majava, V. Isoherranen, and P. Kess. 2013. Business collaboration concepts and implications for companies. *International Journal of Synergy and Research* 2, 1-2 (2013).

[13] Velegrakis-Y. Koutrika G. Papadimitriou, D. 2018. Modeling and Exploiting Goal and Action Associations for Recommendations. *EDBT* (2018), 409–420.

[14] Horvitz E. Giles C. L. Pennock, D. M. 2000. Social choice theory and recommender systems: Analysis of the axiomatic foundations of collaborative filtering. *AAAI/IAAI* 30 (2000), 729–734.

[15] Christina Scandelius and Geraldine Cohen. 2016. Sustainability program brands: Platforms for collaboration and co-creation. *Industrial Marketing Management* 57 (2016), 166–176.

[16] L. Schmidt, J. Sjöström, and A. B. Antonsson. 2015. Successful collaboration between occupational health service providers and client companies: Key factors. *Work* 51, 2 (2015), 229–237.

[17] The Big Issue. [n.d.]. B Corp month: Is certification about competition or collaboration? https://www.bigissue.com/life/ethical-shopping/b-corp-month-certification-competition-collaboration/. The Big Issue, March 2, 2023. Retrieved January 31, 2025.

[18] Juhani Ukko et al. 2022. The importance of sustainability engagement in small businesses supplier collaboration. *Sustainable Development* 30, 1 (2022), 1–9.

[19] Hevner A. Maedche A Vom Brocke, J. 2020. Introduction to design science research. *Design science research* 12, 12 (2020), 1–'3.

[20] Zheng Z. Qiu Z. Wang H. Gu H. Shen T. ... Chen E. Wu, L. 2024. A survey on large language models for recommendation. *World Wide Web* 27, 5 (2024), 60.

[21] Xie R. Hou Y. Zhao X. Lin L. Wen J. R. Zhang, J. 2023. Recommendation as instruction following: A large language model empowered recommendation approach. *ACM Transactions on Information Systems* (2023).