WCED
Enti 415/ CPSC 405
Dr. Chad Saunders & Mea Wang
Group #4
Rohan Rajagopal, Saif Sadat, Anusha Shaikh, Syed Shaheryar, Thi Truc Ngoc Tran

1. Mission Statement

We're on a mission to make organizational intelligence accessible, immediate, and reliable. We do this by simplifying the complexity of enterprise and ESG data into clear, actionable insights that any employee can understand and act upon. We speed up ESG compliance, reduce human error, and unlock strategic value hidden inside an organization's own information by eliminating manual analysis, reducing bottlenecks in reporting, and delivering trusted data to decision-makers in seconds.

2. Business Vision

Our long-term strategy is to scale from an ESG reporting assistant into a fully integrated enterprise intelligence platform. We start by solving a focused, urgent pain: fragmented ESG reporting that's time-consuming, manual, and prone to error. Early adopters will interact with the system through a pilot-based model: using natural-language queries to test real ESG scenarios, validate results against existing workflows, and measure the time savings. Once value is proven, organizations move to a tiered subscription where our platform becomes a core system for ESG reporting, audit preparation, and sustainability planning.

Over time, the platform will move from a purely ESG tool into a multisource analytics engine: ingesting, standardizing, and analyzing various datasets on emissions, operational performance, asset efficiency, safety, financials, and supply chain indicators. This creates one coherent intelligence layer that accelerates not only compliance but also informs strategic forecasting, risk modeling, and investment decisions.

There are three pillars to our scalability:

1. Modular Data Integration - Once the ESG data is ingested, integrating new sources such as production metrics, environmental sensors, maintenance logs, and financial systems becomes incremental rather than foundational. This allows for easy deployment across other departments without rebuilding infrastructure.

2. Industry-Adaptive Models: The core engine is designed to learn domain-specific language and patterns. As we scale into new sectors like utilities, mining, manufacturing, infrastructure, and finance, the analytic modules will be configured to reflect each industry's compliance standards, KPIs, and reporting frameworks. The architecture will support multi-vertical growth without rewriting the platform.

3. Repeatable Deployment - Companies can have similar ESG frameworks: Scope 1–3 emissions, TCFD, GHG Protocol, SEC, and ISSB. We can repeat implementations across energy producers, midstream operators, financial players, and large emitters. The more ESG regulation universalizes globally, the more adoption becomes scalable-not less.

3. Business Pitch

ESG reporting teams in Western Canada's oil and gas industry have to work with emissions data fragmented across disparate production systems, metering infrastructure, SCADA logs, Excel sheets, environmental audits, and historical reports. The presence of drag-and-drop SQL tools is common; however, this is fundamentally based on users already knowing which tables to query, what relationships to join, and which metrics to look for, thus making traditional querying reactive or requiring human interpretation at each and every step.

Our platform solves the core issue: ESG reporting isn't about data retrieval; it is about interpreting information. Rather than require users to find values through SQL-style searches, our system can automatically regularize sources, identify anomalies, perform cross-checks for inconsistencies, and transform complex patterns of data into contextual insights related to ESG requirements. Instead of “write a query,” users can ask real questions: “Which facilities are at risk of exceeding methane intensity thresholds next quarter?”, “Where are my Scope 1 emissions trending above forecast?”, “Which missing reporting fields are preventing audit readiness?” This cuts down the hours of manually sifting through data, checking values, and reconciling discrepancies. It also uncovers insights that SQL fetches but never explains: trends, correlations, forecasting, uncertainty, and risks in data quality.

As energy producers come under intensifying demands for sustainability, regulatory scrutiny, and investor pressure, time-efficient and insight-driven analytics becomes mission-critical. By removing dependency on technical querying and manual interpretation, our platform accelerates ESG reporting cycles, strengthens decision confidence, and empowers companies to advance sustainability goals with the data they already have.

And that opportunity expands well beyond oil and gas. Any organization with multi-system data, ESG compliance requirements, or intensive reporting workloads stands to benefit from AI-driven intelligence-from utilities and manufacturing through to logistics, public infrastructure, finance, and insurance. But as ESG disclosure standards continue to tighten, the need for clean insights and not just raw data extraction positions this venture for rapid growth as a core engine for energy and sustainability data analytics.

4. Proposed Technology

The proposed technology integrates AI-driven natural-language querying with WCED's existing SQL Server production database. This transforms how users interact with publicly available petroleum and emissions data. At its core, the system connects a LangChain-based middleware or LlamaIndex agent to WCED's 13 million-row SQL dataset. This bridge layer enables an OpenAI GPT-4o model to translate user questions into optimized SQL queries, execute them securely, and return answers through an AI chatbot interface. For development and testing, the database is hosted in WCED's on premise VPS while computation and prototyping occur within Databricks, providing a scalable yet cost-efficient environment for experimentation.

This technology is innovative because of its ability to eliminate technical barriers between data and decision-making. Traditionally, analyzing WCED's petroleum data requires domain knowledge, SQL expertise, and manual exports from complex front-end dashboards. By introducing an LLM-powered query interface, the chatbot would slightly mirror ChatGPT but is specialized for Western Canadian energy data. Users can ask questions like "Which company improved venting reduction the most in 2024?" or "Compare Spur Petroleum and Batex production for Q3." Then we would be able to receive immediate, accurate insights without deep interpretation. This represents a disruptive leap from static reporting tools such as Tableau or Power BI toward intelligent, on-demand analytics tailored to the oil-and-gas sector.

Commercialization will occur through integration into WCED's public web portal, as a subscription-based feature. The chatbot would serve as the primary interface for both existing and new customers, with tiered access models like free trial, limited queries per month, and premium unlimited plans. This approach positions the chatbot as a scalable SaaS offering rather than a one-off project, which unlocks future revenue streams through API licensing, data-as-a-service, and potential partnerships with other provinces as WCED expands into Manitoba, Yukon, and NWT. In short, the technology not only enhances usability but also creates an entirely new commercial pathway for WCED by converting raw public data into an interactive intelligence product that is easy to grasp for the general public.

5. Competitor Analysis

Our competitive landscape spans three categories: public energy data portals, enterprise BI/analytics tools, and manual analyst-driven workflows - each widely used across Western Canada's energy sector, but each failing in ways that our platform directly resolves.

Direct competitors include open/public portals such as WCED/Petrinex, which remain the authoritative source for production and emissions data. Their strength is reliability and regulatory completeness; however, they are not designed for exploration. Users must already know which tables to open, how fields relate across systems, and how to manually reconcile inconsistencies, making them inaccessible for non-technical staff.

Enterprise BI tools (Power BI, Tableau, Looker) represent the second major direct competitor group. These platforms excel at visualization, dashboard publishing, and governed reporting, and they are well-funded with strong enterprise adoption. Their limitation is that they are fundamentally consumer tools, not discovery tools: they require pre-built models, predefined metrics, and ongoing analyst maintenance. When a new ESG question emerges, e.g., “Which facilities are likely to exceed methane intensity thresholds next quarter?”, these tools cannot generate an answer unless an analyst has already modeled that logic or manually built a one-off report.

Our indirect competitors are the entrenched workflows that dominate industry practice today: analysts exporting data into Excel or PowerPoint, ad-hoc SQL queries built from memory, and static dashboards that present stale KPIs without offering deeper insight. These workflows are flexible but slow, error-prone, and heavily dependent on staff expertise. They also do not scale; every new question requires new manual effort.

Across all categories, the universal blind spot is the same: none of these solutions help non-technical users understand the data or ask ad-hoc questions without knowing the underlying schema. All competitors assume the user already knows where the information lives, how to join it, and how to interpret it.

This is the precise competitive gap our solution fills.

Our AI chatbot platform transforms ESG and production analytics by combining (1) a single, cleaned, normalized WCED/Petrinex data layer in Databricks with (2) a domain-aware natural-language interface that understands energy terminology, emissions logic, regulatory frameworks, and dataset relationships. Instead of browsing tables or writing SQL, users ask real operational questions in plain language and receive accurate, contextualized insights supported by transparent lineage.

This creates a unique and defensible advantage:

- We eliminate the prerequisite expertise required by portals and BI tools.

- We reduce analyst dependence and remove manual reconciliation.

- We convert raw data into consumable insights, including anomaly detection, missing-field identification, trend analysis, and ESG-ready context.

- We offer predictable pricing models where general LLMs (ChatGPT, Gemini) become cost-volatile, schema-agnostic, and non-compliant for regulated reporting.

- We enable true self-service analytics, something no existing portal, BI system, or analyst workflow provides.

By directly addressing the most painful limitations of every competing approach and validating this through the operational realities of ESG reporting teams, data fragmentation, schema complexity, and manual analysis bottlenecks, our platform establishes a clear and differentiated position in the market. It does not simply expose data; it makes it interpretable, actionable, and accessible to any employee, which is the exact capability missing from the current competitive landscape.

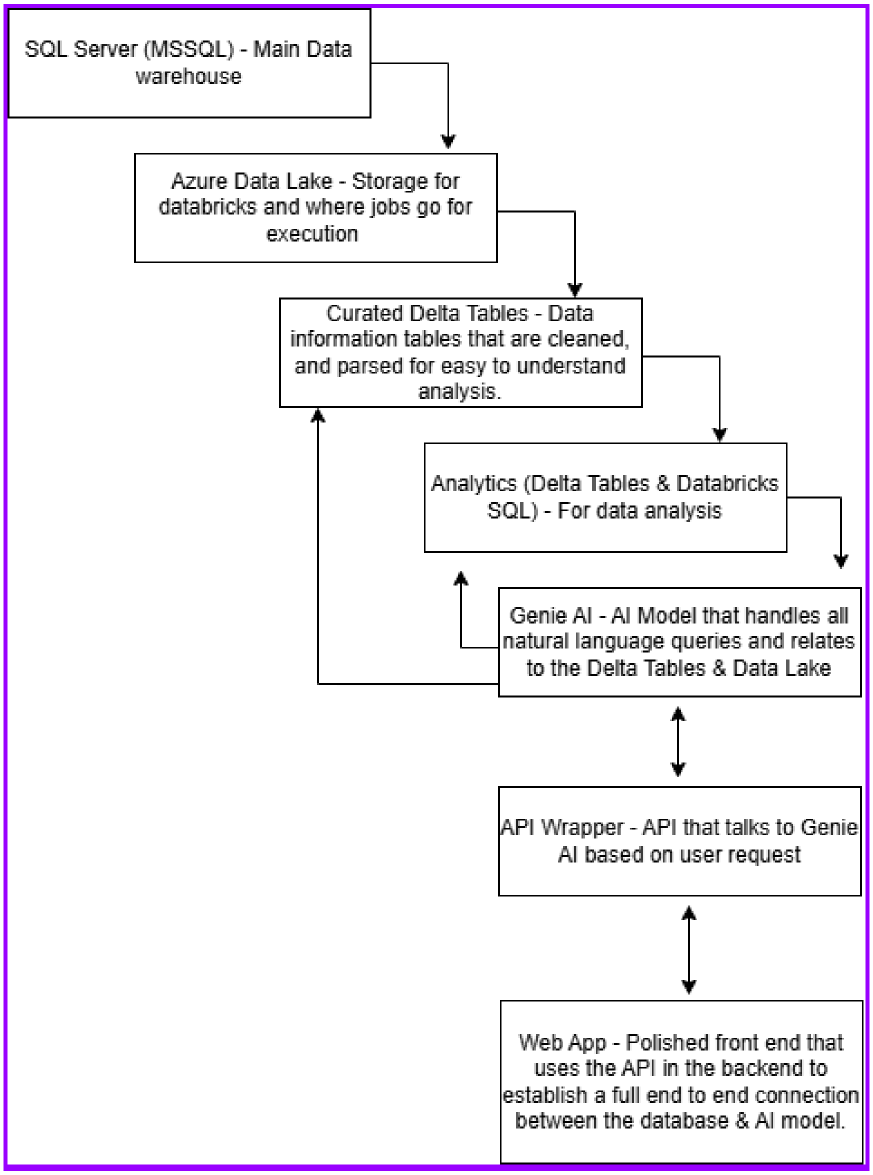
6. Technology Review

Existing technologies for accessing oil-and-gas production data primarily revolve around static reporting systems, business-intelligence dashboards, and manual SQL-based querying. Existing tools such as Petrinex, Power BI, and Tableau in the market that enable users to visualize datasets from the Alberta Energy Regulator (AER) and provincial databases usually rely on specialized knowledge and complex navigation menus[1]. However, petroleum data analytics emphasizes the growing need for accessibility and real-time insight delivery. No public patents or commercial systems currently focus on conversational retrieval of provincial-level oil-and-gas production data, leaving a clear innovation gap[1]. Current solutions are inadequate because they prioritize visualization over accessibility. While WCED's existing front-end provides accurate metrics, it forces users into rigid, multi-step workflows where users have to select the companies, export data, and manually compare production or venting values. These platforms assume technical literacy and provide limited automation for impromptu comparisons [4]. In addition, the scale of WCED's database of over 13 million rows makes static BI dashboards slow and inflexible, especially for trend analysis or company-to-company analysis [5].

Earlier attempts to improve this process have focused on data cataloguing and pre-canned analytics, not conversational interfaces. WCED's own "Petroleum Registry" prototype and similar government dashboards like BC Oil and Gas Commission DataHub demonstrate that the demand for transparency exists, but these systems remain one-directional as they present information but do not interpret it [2]. Other experiments in applying chatbots to data analytics, for example, Databricks Genie and OpenAI SQL agents, prove the technical feasibility of natural-language querying, yet they are generic and lack domain awareness of AER terminology, facility hierarchies, and venting regulations [3].

Our approach will move the field forward by combining large-language-model reasoning with domain-specific energy data. Through a LangChain or LlamaIndex middleware connected to WCED's SQL Server, the system dynamically converts user questions into precise SQL commands, executes them, and translates results back into human-readable insights. Unlike general AI assistants, it embeds the WCED data dictionary and regulatory schema within its prompt context, allowing it to understand industry-specific queries such as "boe/day" or "gas venting." Furthermore, integrating Databricks enables scalable, cloud-based execution without requiring users to host heavy infrastructure. This design shifts the paradigm from static dashboards to interactive, explainable analytics, providing the first conversational gateway to Western Canada's energy-production data and setting a foundation for future multi-provincial, ESG-oriented intelligence systems.

7. Product Design (for technical projects)



The product architecture shown in the diagram is designed around a clear, end-to-end flow from WCED's SQL Server to a user-facing web application. Each component has a specific role in making the data usable, scalable, and ready for commercialization.

At the top of the stack, SQL Server (MSSQL) acts as the main data warehouse that stores the raw production and emissions data from WCED/Petrinex. Rather than querying this transactional system directly, data is periodically extracted and loaded into Azure Data Lake. Azure Data Lake is a cloud-based storage layer optimized for big data analytics; in our design, it is the storage backend used by Azure Databricks. It provides low-cost, scalable storage where large tables can be reorganized and accessed efficiently by analytic jobs.

Within Databricks, this raw data is transformed into curated Delta Tables. These are cleaned, standardized, and partially denormalized tables that are easier for analytics and AI models to work with. Typical jobs in this layer include parsing fields, handling missing values, and creating commonly used aggregates or views. By the time data reaches these curated Delta Tables, it is structured in a way that supports fast SQL queries and consistent business logic. On top of the curated tables, we use Databricks SQL as the analytics layer. Databricks SQL provides a SQL endpoint and query engine that can run analytical queries at scale over the Delta Tables. This is the main interface that our AI layer talks to when it needs to answer a user's question. In our prototype, we configured Databricks Genie to translate natural-language questions into SQL that runs against these curated tables. Genie is not a model we built from scratch; instead, we configured and tuned an existing Databricks capability to understand the WCED schema and map user questions to the right fields and tables.

Between Genie and the user interface, we introduce an API Wrapper. The API wrapper is a lightweight backend service that exposes a REST-style endpoint. When the user submits a question from the web app, the API wrapper receives the request, attaches metadata (such as user ID, membership tier, and query context), forwards it securely to Genie/Databricks SQL, and then returns the structured result back to the front end. It also handles authentication, logging of each query, basic rate limiting, and error handling. This is where we will enforce membership tiers (for example, limiting free users to a small number of queries per day) and capture usage analytics over time.

At the bottom of the stack, the web app provides the polished front end where users actually interact with the system. For this course, we designed and began implementing a simple browser-based application that connects to the API wrapper. The web app presents a chat-style interface for asking questions, displays tables or summary results, and surfaces basic feedback controls (such as thumbs up/down) so we can evaluate answer quality. While our current prototype is still closely tied to the Databricks environment, the product design clearly supports embedding this web app into WCED's public portal as a standalone feature.

From a commercialization perspective, this design supports growth and reliability. Azure Data Lake and Delta Tables provide a scalable data foundation that can handle larger volumes and new provinces without rebuilding the pipeline. The Genie + API wrapper combination allows us to authenticate requests, enforce query limits, and implement caching of frequent questions so that common queries can be answered quickly even under higher load. The usage data collected by the API and web app (for example, query counts per user and peak usage times) can then be used to refine membership tiers, for instance, allowing a small number of queries for a free plan and scaling up to higher query limits and advanced features for paid plans.

Overall, this product design explains how each technical component fits together: Azure Data Lake as scalable storage for analytics, curated Delta Tables and Databricks SQL as the core data and query engine, Genie AI as the natural-language interface to that engine, an API wrapper as the secure bridge, and a web app that our team has prototyped to deliver the experience directly to end users.

8. Problem / Challenges to be Addressed

The primary technical problem our product aims to solve is the difficulty of accessing and interpreting complex petroleum production data housed within WCED's SQL Server database. Despite being public, this data is highly fragmented as it is spread across millions of rows and dozens of interrelated tables representing wells, facilities, production volumes, and emissions. The technical challenge lies in enabling an AI system to query this structured data accurately through natural language without breaking schema integrity or returning misleading results. Large language models, while powerful, are not inherently SQL-aware, meaning they often generate queries that fail due to mismatched column names or misinterpreted relationships. Ensuring accuracy, scalability, and response time efficiency across 13 million records requires fine-tuning query translation pipelines, caching frequently used metrics, and maintaining a secure bridge between the chatbot and the live database, which we as a group are still working on. Additionally, handling sensitive metadata such as venting or emissions data demands robust data-governance and user-access protocols.

Beyond technical complexity, a significant challenge in technology adoption stems from the market's current reliance on manual and static reporting. The oil and gas sector is traditionally conservative with new software adoption, especially AI-based tools, due to data sensitivity and regulatory scrutiny. Many industry professionals are comfortable with legacy tools like Excel, Power BI, or the existing WCED dashboard, which creates a behavioural barrier to switching. Convincing these users to trust an AI chatbot for business-critical queries will require demonstrating reliability, transparency in data handling, and tangible productivity gains. Another adaptation challenge involves aligning with different provincial data standards. Each region defines attributes differently, meaning the chatbot must normalize datasets to ensure accurate cross-provincial comparisons, which is a critical requirement for WCED's expansion plan.

Finally, for the product to meet its business goals, several gaps must be addressed. The first is achieving consistent accuracy in natural-language querying, since incorrect outputs could undermine user confidence and harm WCED's credibility. The second is establishing commercial viability, ensuring the chatbot's cost structure, which is currently built on low-cost VPS hosting and Databricks Community Edition, scales efficiently when transitioning to enterprise customers.

The third gap is the lack of user feedback loops because WCED currently lacks a systematic way to collect user behaviour data and satisfaction metrics, which are essential for refining both AI performance and subscription pricing models. To meet long-term business objectives, the team must therefore focus on improving query precision, integrating user analytics dashboards, and developing clear usage-based pricing that reflects data volume and query frequency. Addressing these challenges will allow WCED's AI chatbot to move from a functional prototype to a market-ready, revenue-generating SaaS platform.

9. Core Innovation

Our principal innovation is the tight integration of a domain-tuned AI chatbot with an energy data warehouse on Databricks. Instead of needing to teach users about WCED/Petrinex formats or BI tools, we merely enable them to ask real questions ("show top 10 producers last quarter," "CO_2 for Company X in July 2025") and convert those requests into structured queries for us. The innovation is not "a chatbot" in itself; it's that the chatbot is pretrained on Western Canadian production/emissions schemas and situated directly on top of a filtered Databricks layer so responses come back as trusted dashboards, tables, or summaries.

This is valuable as current technology in the sector requires SQL skills, various steps in reporting, or IT support, which hinders ESG and regulatory reporting. By enabling non-technical staff to query standardized information with natural language, we reduce reporting time, minimize dependency on analysts, and make public energy data actually useful across the organization. This aligns with the shift towards self-service analytics and tighter disclosure regulations in Western Canada.

It can be done because the architecture is already defined in four layers data ingestion, storage in Databricks, NLP/LLM processing, and web/chat UI and all of those are being facilitated by current cloud platforms. The paper already defines an MVP path: start small with a slice of WCED/Petrinex, prove NL→SQL accuracy, and then expand to other datasets and provinces, showing that small experiments and prototypes could be run early on real data from BlincSoftware.

10. Minimum Viable Product (MVP)

The first version of our product is a natural language (NL) chatbot interface that enables users to query the Western Canadian Energy Repository through simple conversational text. Rather than navigating complex dashboards or writing SQL queries, users can ask questions in plain English and receive clear, contextual responses. This delivers immediate value by making highly technical energy and ESG datasets accessible to non-experts such as policy analysts, ESG teams, and operational managers. We selected this feature set because it provides the fastest path to validating our core hypothesis: that conversational access significantly improves usability, increases engagement, and reveals the highest-value data needs within industry workflows.

Technically, the MVP is powered by Databricks Genie, which allows natural language prompts to query the underlying repository. This configuration tests key feasibility questions such as whether NL queries can consistently map to the correct structured data, whether retrieval-augmented generation (RAG) improves answer quality, and whether latency meets expectations for interactive use. Each response includes a feedback mechanism such as thumbs up or thumbs down, and all usage is logged, allowing us to capture the most common questions, evaluate accuracy, and identify demand patterns across different user groups. These signals form the foundation of our pilot strategy and guide the next phases of model fine-tuning, accuracy improvements, and performance hardening.

The MVP also serves as an early commercialization tool. Because it directly queries a real energy dataset, it offers a functional, demo-ready experience that can be shown to potential customers, partners, and investors. This allows us to secure letters of interest, validate real-world workflows, and demonstrate that the technology meaningfully reduces the effort required for ESG reporting and operational analysis. By gathering technical performance data, customer usage patterns, and explicit feedback, the MVP not only proves feasibility but also establishes a clear path into structured pilots and early revenue opportunities.

11. Product Scalability

From MVP to a fully scaled enterprise product, we will follow a staged, metric-driven growth path that advances functionality, tightens performance SLAs, and expands market reach while controlling costs and risk.

In the MVP (Months 0 - 3) we ship the natural language chatbot connected directly to the Western Canadian Energy Repository to validate the core hypothesis that non-expert users will prefer conversational access to complex energy data and that the repository can be reliably queried via NL prompts; this stage focuses on a lightweight production stack, RAG with a vector index of canonical records, basic provenance links on every answer, and per-response feedback (thumbs up/down) to capture fine-grained training signals and high-value questions for product discovery. Databricks is explicitly designed to enable natural-language BI and fast, self-serve data access at this layer (Source: Databricks). During the MVP, we instrument engagement (DAU/WAU on queries), precision-of-answer (human-validated correctness on sampled responses), and time-to-insight (seconds per query) so we know where to invest next.

In the Pilot phase (Months 3 - 9) we run 4 - 6 week focused departmental pilots with 2 - 3 anchor customers each (agreeing upfront on success criteria (accuracy thresholds, percentage reduction in manual reporting time, and adoption rates)) and iterate quickly on connectors (SAP, Oracle, Salesforce, common cloud object stores), security (SCIM, SSO, RBAC, VPC peering), and compliance mappings. As pilots progress, we harden performance: instrument query caching, shard vector indices for low-latency retrieval, introduce model-fine-tuning and prompt engineering to raise targeted domain accuracy (moving from “good enough” to >90% on high-value question classes), and add observability (SLOs for 95th-percentile latency, error budgets, and detailed provenance trails for audit). For full enterprise readiness, we plan to meet multi-hour bulk export SLAs, sub-second median query latency for cached results, and 99.9% availability for interactive sessions.

The Scaled Product phase (Months 9 - 24) transitions the solution into a multi-tenant, modular intelligence platform: a marketplace of connectors and vertical templates (energy, utilities, mining, finance), role-based analytic modules (ESG reporting, audit prep, risk scoring, forecasting), and packaged automation (scheduled reports, alerting, and API access). Commercially, we move customers from free/low-friction pilots to tiered subscriptions and usage pricing (seat + query-volume + connector premium) and pursue “land-and-expand” motions: secure a department win, instrument clear ROI, and let internal demand pull us across the organization. To enter new verticals, we reuse the core ingestion and entity-resolution pipeline and swap in industry-adapted ontologies and fine-tuned models—making expansion incremental rather than re-architectural. We will accelerate adoption via partnerships (analytics consultancies, major cloud marketplaces) and by publishing case studies and white papers that demonstrate real-time time savings and risk reduction.

To control costs as we scale, we adopt proven cloud cost-optimization best practices. This includes rightsizing and instance scheduling, autoscaling, using spot/spot-like capacity for non-critical workloads, savings plans/reserved capacity for predictable baseline load, and continuous cost monitoring and tagging (Source: AWS Well-Architected Framework). Operationally, we will validate unit economics at each stage (CAC payback, LTV/CAC, gross margin on subscription revenue) and refuse to scale the sales motion beyond repeatable pilot-to-enterprise conversion rates until those metrics meet predetermined thresholds. Finally, our timeline balances speed and realism: MVP (0 - 3 months) to validate demand and collect high-value queries; compact pilots (3 - 9 months, 4 - 6 week experiments per customer) to validate measurable time-savings and compliance fit; and a deliberate scale phase (9 - 24 months) to deliver multi-tenant architecture, industry verticalization, enterprise integrations, hardened SLAs, and a repeatable GTM engine that turns departmental pilots into organization-wide deployments.

12. The Team

Our project group has five members three computer science students (Shaheryar Syed, Sadat Saif, Thi Truc Ngoc Tran) and two business/ENTI students (Rohan Rajagopal, Anusha Shaikh) which gives us a nearly 50/50 split of technical build and business/commercial work, exactly what this AI data project needs. The project is sponsored by BlincSoftware and endorsed by Curtis Bodie (Sponsor Representative), with overall course oversight by Dr. Chad Saunders and Dr. Mea Wang, de facto feasibility and industry fit advisors.

Shaheryar Syed - Data & Platform Lead (Technical)
Shaheryar is a computer science student with hands-on experience in data engineering, cloud platforms, and distributed systems. He is responsible for the Databricks deployment, WCED/Petrinex data ingestion, and Delta Lake schema design. His background in building scalable data pipelines ensures that the chatbot always interacts with clean, consistent, and governed data. Shaheryar's strengths in SQL optimization, ETL development, and database normalization make him the ideal lead for the platform's core data infrastructure.

Sadat Saif - AI/Chatbot Lead (Technical)
Sadat is a computer science student specializing in AI systems, natural language processing, and software architecture. He leads the design and verification of the natural-language-to-query pipeline, including prompt engineering, query translation, and response validation. Sadat works closely with Shaheryar to map user questions to the correct fields and tables within the WCED/Petrinex dataset. His prior work with LLM agents, LangChain/LlamaIndex, and API-based AI workflows equips him to build a reliable and domain-aware chatbot interface.

Thi Truc Ngoc Tran - Product & Integration Lead (Technical/Product)
Thi is a computer science student with experience in full-stack development, UI integration, and API documentation. She is responsible for integrating the chatbot into the web interface, ensuring seamless user interaction, and validating usability with the project sponsor. Thi's experience in frontend frameworks, RESTful API design, and user-centered development allows her to translate technical features into an intuitive and stable product. She also monitors technical risks associated with sponsor requirements and deployment constraints.

Rohan Rajagopal - Business & Stakeholder Lead (Business)
Rohan comes from the ENTI/business specialization and brings experience in customer discovery, market validation, and stakeholder management. He leads sponsor communication, requirements gathering, and the development of the business model, including pricing tiers and adoption strategies for different customer sizes. Rohan's background in business analysis and

competitive research ensures that the project aligns with real buyer needs and fits BlincSoftware's "develop first, then fund" philosophy.

Anusha Shaikh - Delivery, Operations, Finance & Business Lead (Business)
Anusha is a business student with experience in financial modeling, project operations, and organizational strategy. She owns the cost model, including VPS hosting costs, SaaS-style tiered pricing, and scalability estimates, and develops the go-to-market narrative for the final sponsor presentation. Anusha also manages deadlines, documentation control, and overall project operations. Her strengths in financial analysis and structured delivery ensure that the project remains feasible, on schedule, and investor-ready.

Talent triangle coverage
Business (finance/strategy): Rohan + Anusha work on pricing (1–3 queries/month vs. unlimited), customer needs, sponsor reports, and final presentation to Curtis Bodie.
Technical (product/engineering): Shaheryar, Sadat, and Thi together implement the Databricks layer, the AI chatbot, and the front-end integration i.e., functional working solution the sponsor asked for ("AI bot to analyze and answer questions on production/emissions data").

Domain expertise (industry/data): All five of us are working off the same WCED/Petrinex problem statement from the sponsor, but Shaheryar (data), Sadat (query logic), and Rohan (sponsor/user interviews) will validate that the fields, terms, and reports the chatbot exposes match what Western Canadian O&G users actually ask for.

In summary, the two business students make the project appealing and viable, the three computer science students make it technically viable, and the sponsor/advisors keep it grounded in real oil-and-gas reporting needs and data. As a balanced group of computer science and business students, we combine the technical depth needed to build an AI-powered analytics platform with the business insight required to position it for real-world adoption. Collectively, we have experience in data engineering, natural-language systems, full-stack development, market analysis, and financial modeling. These are skills that directly align with the core demands of this project. Our complementary strengths allow us to execute reliably across data, AI, product, and business domains, making us exceptionally well-suited to deliver a viable, scalable solution for WCED/Petrinex stakeholders.

13. Market Analysis

Our target market is Western Canadian oil and gas production and emissions data consumers who in fact need to deal with WCED/Petrinex-type data Alberta, BC, and

Saskatchewan mid-size operators, service/consulting firms producing reports for multiple clients, and public/para-public buyers (regulators, agencies, utilities) needing timely, transparent insights. This is an understandable goal since 95% of Canadian oil and much marketable gas are found in the Western Canada Sedimentary Basin, where thus data density and report agony tend to pile up.

North America currently spends approximately US$3.1B on AI in oil and gas with robust growth through 2030 – 2034. Using a conservative 10% Canadian share, that's around US$300M; as Western Canada is where most of the reporting and production takes place, we take US$240M (≈ C$325M) as the TAM i.e. all Western Canadian O&G players who would buy AI/data tools to make production/emissions data valuable.

Our SAM is the sub-segment of that market which only addresses public / regulatory / multi-jurisdiction data (WCED/Petrinex + provincial open data) and needs a low-friction solution primarily mid-sized producers, multi-client consultants, and a few public-sector consumers. If we assume 15% of the TAM for that character, then the SAM is about C$50M. This is most likely to be the group that will care about "I don't want to query this manually anymore."

Our SOM (what we can realistically win first) is an entry of pilots: e.g., 15–25 organizations (a combination of mid-sized producers and service firms) on C$15–20k/year each = C$300–500k in phase one. That is in line with the fact that a genuine sponsor (BlincSoftware) has already requested "an AI bot to parse and respond to questions on production and emission data," which is early proof of customer interest and readiness to test.

14. Value Proposition

We make Western Canadian production and emissions data actionable in seconds, not hours. By putting an AI chatbot on top of a Databricks pipeline, users can ask natural language questions and get back instant tables, summaries, or dashboards, without downloading, cleaning, or calling an analyst. That directly reduces reporting and ESG/compliance work and makes public data actually actionable.

Our initial customers are mid-size Western Canadian O&G firms, consulting firms that produce emissions/production reports for multiple clients, and regulators/partners who need quick, understandable insights. They are all interested because they are coming under increasing demand for reporting but may or may not have data/BI internal capability.

We have real interest from the sponsor, BlincSoftware, whose project brief is literally "build an AI bot to review and respond to questions on production and emission data of all oil and gas companies in Western Canada." That is a real stakeholder pointing to a pilot and a real use case, which is early proof of willingness to experiment and later pay on a SaaS level.

15. Commercialization Pathway

Our commercialization pathway follows a structured progression from MVP validation to market-ready deployment, revenue generation, and long-term profitability. We begin with a focused MVP, which is an NLP chatbot interface connected to the Western Canadian Energy Repository. This first phase provides essential technical learnings, including query accuracy, latency performance, and data-governance needs, while establishing early relationships with industry stakeholders. During this period, we secure letters of interest from energy producers, analytics partners, and ESG consulting groups who see value in faster, more accurate reporting workflows. These early commitments anchor the next stage: tightly defined departmental pilots.

As we progress from MVP to a market-ready product, we run 4 - 6 week pilots with 2 - 3 anchor customers per cohort. Each pilot measures accuracy improvements, time savings in reporting workflows, and adoption rates. This phase introduces hardened enterprise features: role-based access control, SSO/SCIM, VPC peering, query auditing, provenance tracking, and performance SLAs. We also expand connectors to SAP, Oracle, and Salesforce, enabling seamless ingestion of ESG, operational, and compliance datasets. This pilot-driven motion not only validates technical readiness but also aligns the platform with regulatory expectations, including SOC2 readiness, GHG Protocol alignment, ISSB/TCFD reporting structures, and industry-specific audit requirements. By the end of this stage, the product meets enterprise security expectations and demonstrates measurable ROI, enabling confident commercialization.

Revenue generation begins as pilot customers transition into tiered subscription plans. Our pricing blends seats, query volume, and connector premiums, creating scalable unit economics while keeping entry friction low. We focus on departmental “land and expand” motions, proving value, then expanding horizontally into operations, finance, safety, and risk groups as data modules grow. Partnerships with analytics consultancies, cloud marketplaces, and system integrators accelerate adoption by embedding our platform into existing enterprise workflows and digital-transformation initiatives.

Profitability emerges as we transition into a multi-tenant, multi-vertical enterprise intelligence platform. Our modular architecture keeps infrastructure costs predictable by enabling autoscaling, workload tiering, and vector-store optimization, while fine-tuned models and incremental ontologies allow efficient expansion into adjacent industries without re-architecting the core system. As customer lifetime value increases through cross-departmental expansion, we maintain disciplined unit economics, targeting a 3:1 LTV-to-CAC ratio, high gross margins on subscription revenue, and strong retention driven by embedded workflows and regulatory dependency. To support long-term profitability, we continue to invest in partnerships with cloud providers, ESG consultancies, and enterprise data-platform vendors whose ecosystems amplify distribution while reducing our direct sales burden.

Overall, this pathway is realistic and grounded in the mechanics of enterprise software adoption: validate the workflow pain with a simple NL interface, derisk technical and regulatory requirements during structured pilots, scale revenue through repeatable deployments, and achieve profitability through platform modularity, strong unit economics, and deep integration into customer operations. The combination of early partner interest, repeatable pilot design, enterprise-grade technical hardening, and modular expansion positions the product for sustainable growth and long-term market leadership.

16. Firm Growth Plan

The first stage is to create a MVP using the Databricks platform to clean and train the emissions datasets. The next stage will enhance the platform's AI chatbot to create an interactive experience for the user. This will have data visualization ability for live ESG insights and updates upon analyzing these large datasets. The third stage will include a predictive analysis of data outputs to allow for informed decisions. In the long run, our venture will expand across industries to grow over time.

The cost model will primarily involve Databricks as an expense and we have set out an initial budget with BlincSoftware of about 500$. The revenue model consists of a subscription-based platform for users to enter a trial of free queries before payment is required. The basic tier consists of limited datasets, followed by the middle tier, made for companies that require automation for the ESG data outputs. Lastly the enterprise tier for large corporations with full data integration support, custom services for data, and visual analytical reports.

The staffing and operations will consist of the founders and developers to train the data under the Databricks platform with the ESG integration. Then, a business development team will handle further partnership and marketing of the venture. Staffing will continue to grow as the operations and business expand into manager roles for data analysis and the sales team. The overall expansion strategy will begin from Western Canada, and scale across industries and regions to partner with ESG-related firms. International expansion will begin once there is a foundation and companies across North America and the Middle East of GCC countries where they believe " ESG is not just a trend it's a strategic imperative. We work with investors to identify opportunities that align financial performance with positive environmental and social outcomes across the GCC and beyond." (Superdon@Gulfequity.com, 2025)

17. Financing Strategy (Pre-Commercialization)

During the pre-commercialization stage, we plan to fund development primarily through non-dilutive sources, supplemented by modest founder contributions. Our goal is to cover infrastructure, continued MVP development, and early pilot work without taking on heavy dilution too early.

We would potentially pursue the following funding sources:

1. University Innovation Grants and Student Entrepreneurship Funds
   Source: Internal grants and student innovation programs (e.g., Hunter Hub, university innovation hubs, entrepreneurship centres, or faculty-level funds supporting early-stage prototypes and pilots).
   Amount: ≈$5,000–$10,000
   Current Status: Planned, we confirmed our eligibility, and we plan on working on applications following completion of the course deliverables and final pitch.
   Role: Supports further MVP refinement, UX improvements, integration work, and structured pilot testing with early adopters.

2. Pitch Competitions and Startup Showcases
   Source: Data/AI, analytics, or ESG-focused pitch competitions and student venture competitions.
   Amount: ≈$500-2,000 in prize money and/or in-kind credits (e.g., cloud credits, mentorship programs).
   Current Status: Planned, our pitch deck and demo from this course will be adapted for external competitions in 2025.
   Role: Provides non-dilutive capital and visibility to future potential partners, while validating market interest through competitive selection.

3. Government / Innovation Grants
   Source: Provincial or federal innovation programs that support AI, data infrastructure, or ESG/analytics solutions (e.g., Alberta Innovates micro-voucher-type programs and Mitacs entrepreneurial/innovation grants aimed at early-stage tech ventures).
   Amount: ≈$7,000–$12,000.
   Current Status: To be pursued once we demonstrate a working MVP with feedback from a customer base and at least one interested pilot organization.
   Role: Funds hardening of the data pipeline, security and compliance work, and scaling

from a student project into a robust pilot-ready product.

4. Angel Capital
   Source: Individual angel investors or small pre-seed funds with a focus on AI, B2B SaaS, or data/ESG solutions.
   Amount: ≈20,000
   Current Status: In discussion, only to be raised once we secure clear traction like a paying pilot or strong enterprise interest.
   Role: Used to hire 1–2 other dedicated technical/business team members, extend our infrastructure capacity, and accelerate sales and onboarding for early enterprise customers.

Taken together, these sources are expected to provide approximately $12,500–$24,000 in non-dilutive funding with optional pre-seed capital layered on top if traction justifies it. This mix is designed to sustain ongoing development beyond the course, fund validation with potential early pilot partners and position the venture to enter commercialization with a validated product and a clear path to recurring SaaS revenue.

18. Anticipated Challenges & Risks

Our venture faces both technical and business risks as we move from an MVP into a pilot-ready SaaS product. Below, we categorize the key risks and outline how we plan to mitigate them.

Technical Risks (Feasibility & Scalability)

Technical Risk 1 – Natural-language to SQL accuracy and reliability
Our chatbot must translate natural-language questions into correct SQL across a large, multi-table dataset without breaking schema integrity or returning misleading results. If the NL-to-SQL layer misinterprets column names, joins, or filters, users could receive incorrect production or emissions numbers and lose trust in the system.

Mitigation:
We will start with a constrained subset of the schema (a smaller group of high-value tables) and expand gradually as we validate common query patterns. Each NL-to-SQL template will be backed by unit tests and regression checks on a library of “known correct” questions and answers. For early adopters, we will surface query summaries or the underlying SQL so

advanced users can verify how results were generated. Misinterpretations will be logged, reviewed, and used to iteratively improve prompts, mappings, and guardrails.

Technical Risk 2 – Performance and scalability on large datasets
Static dashboards already struggle with responsiveness on large production and emissions datasets. Adding an AI-driven query layer introduces additional compute and model-inference overhead, which can lead to slow response times or high infrastructure costs as query volume grows. As the broader NLP and conversational analytics market expands rapidly, scalability expectations will only increase (Natural language processing market, 2030; NLP market size, 2025).

Mitigation:
Our architecture uses Databricks and curated analytics tables, which allows us to optimize frequently accessed views and leverage autoscaling where appropriate. We will cache popular queries, implement sensible rate limits per user or tier, and monitor query patterns to identify heavy workloads early. Performance testing will be incorporated into our MVP-to-pilot roadmap so that we only commit to higher-cost infrastructure once we have clear usage data and a funding path that supports it.

Technical Risk 3 – Data governance, security, and regulatory alignment
Even when working primarily with public or quasi-public datasets, there are governance expectations around how data is combined, exposed, and interpreted. A poorly configured chatbot could expose inappropriate fields, enable overly broad queries, or create confusion about compliance-related metrics, especially as organizations increasingly rely on AI-augmented decision systems (AI-powered analytics, 2025).

Mitigation:
We will implement role-based access control and authentication at the application layer and clearly separate public views from any restricted or internal views. Different membership tiers will map to different permission sets, query limits, and logging requirements. As we expand to new datasets or jurisdictions, we will normalize fields carefully and apply region-specific rules so that comparisons remain accurate and aligned with local reporting standards.

Business Risks (Funding, Adoption, Competition)

Business Risk 1 – Conservative adoption and trust in AI
Target users in data-heavy, regulated industries are often conservative with new tools and may already be comfortable with spreadsheets, existing dashboards, or manual analyst workflows. If users do not trust the chatbot's answers or see it as a "nice-to-have" rather than a necessity, adoption will be slow, and churn risk will be high. This hesitancy persists even as AI-driven

decision intelligence is expected to augment or automate a large share of business decisions in the coming years (AI-powered analytics, 2025).

Mitigation:
We plan to work closely with early pilot partners to co-design workflows that clearly demonstrate time savings and reduced friction compared to current methods. Early deployments will emphasize transparency (showing how answers are generated), side-by-side comparisons with existing reports, and simple metrics such as "time saved per query/report." Positive pilot outcomes and testimonials will be used as case studies to reduce perceived risk for additional organizations.

Business Risk 2 – Competitive pressure from existing tools and generic AI platforms
Our competition includes existing portals and dashboards, traditional BI tools, in-house analysts, and emerging generic conversational analytics tools from large cloud or AI providers. These alternatives are familiar to users and often backed by significant resources, especially as conversational analytics and decision-intelligence platforms become more mainstream (NLP market size, 2025; AI-powered analytics, 2025).

Mitigation:
Our strategy is to differentiate through domain specificity and workflow fit rather than trying to be a generic "ask your data" tool. We are focusing on a well-defined dataset and user group, optimizing for the exact questions they struggle to answer today and the cross-cutting comparisons that existing tools make cumbersome. We will also explore partnerships with data owners, analytics consultancies, or integration partners so that our product is embedded directly into existing environments rather than competing purely as a standalone tool.

Business Risk 3 – Funding constraints and cost structure as we scale
Our initial stack (including low-cost cloud resources and educational tiers) is suitable for an MVP but may not be sustainable as we add more users, jurisdictions, or premium features. If we underestimate infrastructure costs or fail to secure grants and early investment, we may struggle to maintain acceptable performance while keeping prices attractive. This is particularly important because many IT teams are already overextended and spend significant time on repetitive manual requests, making efficiency and cost control critical (Graf, 2025).

Mitigation:
We have outlined a pre-commercialization financing strategy that prioritizes non-dilutive funding (university grants, competitions, innovation programs) with a modest angel round only once we have clear traction. This staged approach allows us to align infrastructure spend with concrete milestones such as working pilots, letters of intent, or early revenue. We will also design pricing and tiered usage limits so that heavier usage is directly tied to higher subscription revenue, helping ensure that growth in demand is matched by growth in cash flow.

Overall, by constraining technical scope early, piloting with real users, differentiating on domain-specific value, and phasing both infrastructure commitments and financing, we aim to reduce the likelihood and impact of these technical and business risks as we move from MVP to a scalable, revenue-generating product.

19. Financing Strategy (Pre-Sales)

As we move from our MVP to the pilot stage and toward our first paying customers, we expect a period where cash outflows (infrastructure, development, and go-to-market activities) will exceed incoming revenue. Our pre-sales financing strategy focuses on extending the runway created in the pre-commercialization stage and timing equity investment so that it is tied to clear traction. We estimate that covering pre-sales cash flow needs over a 9–12 month period will require approximately $18,000–$30,000.

To bridge this gap, we plan to use the following funding sources mentioned before:

1. Carryover from Non-Dilutive Funding (Grants and Competitions)
   Source: University innovation grants, student entrepreneurship funds, and external startup competitions (e.g., Hunter Hub–type grants, university innovation hubs, and AI/data-focused pitch events).
   Amount (pre-sales allocation): ≈ $5,000–$10,000
   Role in pre-sales: A portion of the non-dilutive capital obtained during the pre-commercialization stage will be reserved to cover hard costs such as cloud infrastructure, essential development tools, and targeted pilot activities (e.g., user testing sessions, small travel/presentation costs). This carryover ensures continuity between MVP completion and initial commercialization.

2. Government / Innovation Grants
   Source: Provincial and federal innovation programs that support AI, data infrastructure, and ESG/analytics solutions (e.g., Alberta Innovates–style micro-voucher programs and Mitacs-style entrepreneurial/innovation grants).
   Amount (pre-sales allocation): ≈ $7,000–$12,000
   Role in pre-sales: These funds will be directed toward "ready-to-commercialize" work: strengthening the data pipeline, improving security and compliance, and hardening the platform to meet enterprise reliability expectations. By using grants for this phase, we reduce the amount of equity capital required to reach commercialization.

3. Angel Capital (Milestone-Triggered)
   Source: Individual angel investors or small pre-seed funds with a focus on AI, B2B SaaS, or data/ESG solutions.

Amount (target): ≈ $20,000
Current status: In discussion, only to be raised once we secure clear traction (e.g., at least one strongly engaged pilot organization, letters of intent, or early paid proof-of-concept work).
Role in pre-sales: Angel capital will be used as a flexible buffer to cover any remaining pre-sales shortfall, including 1–2 part-time or contract technical/business hires to accelerate improvements and onboarding, additional infrastructure capacity if usage grows faster than expected, and basic sales and customer success activities to convert pilot users into paying customers. This round is intentionally kept small and milestone-based to avoid unnecessary dilution.

4. Founder Contributions and Deferred Compensation
Source: Founding team (top-up cash and deferred/part-time compensation).
Amount (buffer): ≈ $2,000–$3,000
Role in pre-sales: If there are modest timing gaps between grant disbursements, angel funding, and expenses, founders will bridge them through small additional personal contributions or by deferring compensation, like maintaining part-time work while building the venture. This provides an additional safety buffer without forcing the company into a larger or premature funding round.

Therefore, our pre-sales financing capacity is expected to be approximately $12,000–$22,000 in non-dilutive capital like grants and competitions allocated toward pre-sales activities, plus approximately $20,000 in angel capital if milestones are met, and approximately $2,000–$3,000 in founder buffer contributions if required. This yields a potential pre-sales financing capacity of roughly $34,000–$40,000, which is sufficient to cover essential infrastructure and tooling, complete the work required to make the product pilot-ready, and maintain operations until we secure our first paid pilots and transition toward recurring subscription revenue.

20. Financial Projections

Pro Forma Summary (CAD)

Year	Revenue	Total Costs	Approx. Operating Result
0	$0	$30,000	-$30,000 (covered by grants/equity)
1	$78,000	$60,000	+$18,000

2	$330,000	$200,000	+$130,000
3	$516,000	$320,000	+$196,000

This graph presents a high-level pro forma view of our venture from the pre-revenue stage (Year 0) through three years post-revenue (Years 1–3). All projections are in CAD and are based on conservative assumptions about pricing, adoption, costs, and funding. Below, we outline the key assumptions underlying these projections and describe the expected evolution of our revenues, costs, and funding sources over this period.

1. **Assumptions**

Our business model is a tiered B2B SaaS subscription sold at the organizational level, not by individual seat. We assume three main tiers: a Starter plan at approximately $12,000 per year, a Professional plan at approximately $18,000 per year, and an Enterprise plan at approximately $30,000 per year. For simplicity, we assume limited discounting in the first three years so that early adopters pay close to list price in exchange for direct involvement in pilots and feedback. On the adoption side, Year 0 is dedicated to MVP development and pilot preparation, with no paid subscriptions. In Year 1, we assume five paying organizations (three on Professional and two on Starter). In Year 2, we assume eighteen paying organizations (ten Professional, five Starter, three Enterprise). In Year 3, we assume twenty-five paying organizations (twelve Professional, five Starter, eight Enterprise). Cloud and infrastructure costs (our COGS) are assumed to stabilize around 15–20% of revenue at scale. We also assume lean operations initially, with founders performing most development, sales, and admin work and only gradual outsourcing or hiring as revenue grows.

On the financing side, development in Year 0 and part of Year 1 is supported by a mix of government and university grants, pitch competition winnings, a small angel round, and modest owner equity contributions, consistent with the pre-commercialization and pre-sales financing strategies already outlined in Sections 17 and 19.

2. **Revenue Forecast**

Under these assumptions, Year 0 has no subscription revenue. In Year 1, revenue is driven by three Professional customers and two Starter customers. This yields approximately $54,000 from Professional plans (3 × $18,000) and $24,000 from Starter plans (2 × $12,000), for total Year 1 revenue of roughly $78,000. In Year 2, we assume ten Professional, five Starter, and three Enterprise customers. This equates to approximately $180,000 from Professional plans (10 × $18,000), $60,000 from Starter plans (5 × $12,000), and $90,000 from Enterprise plans (3 × $30,000), for a total of approximately $330,000 in revenue. In Year 3, we assume twelve Professional, five Starter, and eight Enterprise customers. This corresponds to approximately $216,000 from Professional plans (12 × $18,000), $60,000 from Starter plans (5 × $12,000), and $240,000 from Enterprise plans (8 × $30,000), for total revenue of approximately $516,000.

These revenue levels are consistent with reaching a realistic SOM of 15–25 organizations on multi-thousand-dollar annual subscriptions.

3. Development Costs and Operating Costs

Development costs include founder and early team compensation (in cash or equivalent), contracted development work, and product design. In Year 0, we estimate total development-related spending at approximately $20,000, reflecting intensive build and iteration of the MVP. In Year 1, as we harden the product and respond to pilot feedback, development costs increase slightly to around $25,000. In Years 2 and 3, development costs grow to approximately $70,000 and $110,000 respectively, as we add new features, vertical-specific modules, and more robust analytics.

Operating costs include sales and marketing, general and administrative expenses (e.g., legal, accounting, basic tools), and cloud/infrastructure (COGS). In Year 0, we estimate cloud and infrastructure costs at roughly $6,000, with operating costs (legal, minimal marketing, travel, admin) at approximately $4,000, for total Year 0 cash outflows of around $30,000. In Year 1, with pilots and first paying customers, we estimate cloud/infrastructure at approximately $10,000, sales and marketing at roughly $15,000, and G&A at roughly $10,000, for total Year 1 operating costs of about $60,000.

As we scale, we expect Year 2 operating costs to rise to about $200,000, made up of approximately $70,000 in development and product, $50,000 in cloud/infrastructure (reflecting higher usage), $50,000 in sales and marketing (more demos, travel, and content), and $30,000 in G&A (professional services, tools, part-time operations support). By Year 3, we project total operating costs of approximately $320,000, with about $110,000 for development and product, $80,000 for cloud/infrastructure, $80,000 for sales and marketing (including at least one dedicated sales or customer success role), and $50,000 for G&A (legal, HR, finance, and general overhead).

4. **Government Grants, Outside Investment, and Owners' Equity**

To support these costs, we assume that government and university innovation grants contribute approximately \$15,000–\$18,000 in Year 0 and an additional \$5,000–\$7,000 in Year 1, for a total of around \$20,000–\$25,000 in non-dilutive funding. We also assume \$3,000–\$5,000 in pitch competition winnings and in-kind support (for example, cloud credits and mentorship).

Outside investment consists of a small, milestone-triggered angel round of approximately \$20,000 in Year 1, raised only after we secure clear traction (such as engaged pilots, letters of intent, or an early paid proof of concept). Owner equity consists of founder contributions totalling approximately \$7,000–\$10,000 in cash across Years 0 and 1, plus additional sweat equity in the form of unpaid or deferred compensation.

Hence, in Year 0, the venture has no revenue and approximately \$30,000 in costs. This is largely covered by grants, competition funding, and owner equity, resulting in a roughly break-even position or a small cash deficit that can be absorbed by founders. In Year 1, revenue of approximately \$78,000 against operating costs of about \$60,000 and an additional \$20,000–\$27,000 in grants and angel capital yields positive cash flow and a modest operating profit, while also extending the runway. By Year 2, with projected revenue of approximately \$330,000 and operating costs of around \$200,000, the business becomes clearly profitable, generating an operating surplus of roughly \$130,000. In Year 3, revenue of approximately \$516,000 against costs of approximately \$320,000 yields an operating profit on the order of \$196,000, providing sufficient margin to reinvest in product, sales, and expansion. Under these assumptions, the venture relies primarily on non-dilutive funding and modest owner equity to reach initial commercialization and achieves sustainable profitability by Year 2–3 post-revenue.

21. Use of AI (Reflection)

Throughout this project, AI tools played a supportive role across both the technical and research/writing components of our work. On the technical side, AI functioned as an interactive assistant that helped clarify unfamiliar computer science and business concepts, validate our understanding of system architecture, and guide our approach when we were unsure about design decisions. Instead of replacing our work, the AI served as an on-demand tutor, explaining terms, offering examples, and helping us compare different technical strategies so we could make informed choices.

In the research and writing stages, AI provided substantial value as a collaborative drafting tool. We supplied our raw ideas, early notes, and rough outlines, and the AI helped transform them into structured, professional-quality drafts. This allowed us to focus on the actual content while using AI to improve clarity, coherence, and organization. It also helped summarize large sections of information, refine explanations, and generate alternative angles when we needed to answer specific questions in the assignment. When we were unsure how to present our points, the AI provided a clear framework or template we could follow and customize.

Overall, AI significantly changed our workflow by making it more iterative, efficient, and exploratory. Instead of spending long periods stuck on definition gaps or formatting challenges, we were able to develop ideas more quickly and refine them through back-and-forth conversations with the AI. This experience taught us how to use AI not as a shortcut, but as a thinking partner and a tool that enhances our understanding, improves the quality of our written communication, and supports complex problem-solving without removing the need for our own judgment and decision-making.

We used the following AI tools:

- ChatGPT: Interactive tool to learn from and help format our ideas
- Grammarly: Improve our writing

22. References

1. Superdon@Gulfequity.com. (2025, April 4). *The rise of ESG investing in the Middle East: What investors need to know - gulf equity partners*. Gulf Equity Partners -. http://gulfequity.com/the-rise-of-esg-investing-in-the-middle-east-what-investors-need-to-know/#:~:text=However%2C%20the%20introduction%20of%20ESG,across%20the%20GCC%20and%20beyond.
2. *Estimated crude oil production by province*. Open Government Portal. (n.d.). https://open.canada.ca/data/en/dataset/f83f252c-bb16-439d-8441-5a4157b924fd/resource/9b141d44-4a7c-4994-bd03-60985d83ec29
3. Grand View Research. https://www.grandviewresearch.com/industry-analysis/natural-language-processing-market-report
4. Future Market Insights. (2025, March 26). NLP market size, share & forecast 2025–2035. https://www.futuremarketinsights.com/reports/natural-language-processing-nlp-market
5. AI-powered analytics: How decision-making will transform by 2027. (2025, June 25). Sisense. https://www.sisense.com/reports/gartner-ai-powered-analytics/
6. Natural language processing market | industry report, 2030. (n.d.). Amazon AWS. https://docs.aws.amazon.com/pdfs/wellarchitected/latest/cost-optimization-pillar/wellarchitected-cost-optimization-pillar.pdf?utm_source=chatgpt.com

7. Graf, A. (2025, April 28). Calculating the time wasted on repetitive manual tasks. TeamDynamix. https://www.teamdynamix.com/blog/calculating-the-time-wasted-on-repetitive-manual-tasks/