Long COVID Intelligence Project — Publishable Investor Narrative

Building Biomedical Intelligence Infrastructure for Complex Disease Research

Investor Business Plan and Five-Year Commercialization Strategy



Executive Summary

The Long COVID Intelligence Project is a proposed biomedical artificial intelligence company focused on transforming fragmented scientific information into connected, evidence-linked research intelligence. Its initial focus is Long COVID and related post-infectious illnesses—areas characterized by rapidly expanding literature, heterogeneous clinical presentations, multi-system biology, and unresolved questions about mechanisms, biomarkers, and treatment.

The company's central proposition is that researchers and life-science organizations need more than a general-purpose AI assistant or conventional literature search. They need a specialized platform that can connect findings across scientific disciplines, preserve traceability to source evidence, organize biological relationships, identify knowledge gaps, and help prioritize research questions.

The platform will combine biomedical data ingestion and normalization, a curated knowledge base, a biomedical knowledge graph, evidence-grounded AI, research analytics, and a secure enterprise interface. Intended users include pharmaceutical and biotechnology companies, universities, academic medical centers, research networks, nonprofit and government research organizations, and life-science investors and consultants.

Executive Summary: Financial Outlook

The proposed business model combines recurring researcher and institutional subscriptions with enterprise licensing, API and data products, research collaborations, and scientific analytics. The initial disease focus creates depth and a differentiated knowledge foundation; the underlying architecture may subsequently be extended to other complex diseases, subject to validation of scientific need and commercial demand.

The base-case financial model projects revenue growing from $0.5 million in Year 1 to $28 million in Year 5, with cumulative five-year revenue of $50.5 million. Operating expenses are projected at $63.46 million over five years. The model anticipates annual operating profitability in Year 5, with an estimated $10.17 million operating profit. These are illustrative planning assumptions—not validated forecasts—and depend on product execution, customer acquisition, contract values, retention, market adoption, and financing.

The company's long-term value proposition is to build specialized biomedical intelligence infrastructure: proprietary data relationships, knowledge graph assets, research workflows, analytical methods, software, validated outputs, and recurring commercial relationships.

Illustrative projections

Year 1 revenue **$0.5 million**; Year 5 revenue **$28 million**; cumulative revenue **$50.5 million**; cumulative operating expenses **$63.46 million**; Year 5 operating profit **$10.17 million**.

1. The Research Problem

Biomedical knowledge is distributed across scientific publications, clinical trials, epidemiological studies, molecular datasets, genomics, proteomics, immunology, clinical registries, drug-development pipelines, and institutional research programs. Each source provides only part of the picture. Research teams often must manually identify, interpret, compare, and connect evidence across domains.

Traditional search can locate relevant papers, but it does not consistently reveal relationships between findings, assess the strength or limitations of evidence, reconcile contradictory results, or prioritize mechanisms and therapeutic hypotheses. General AI tools may provide summaries, but scientific applications require reliable source attribution, careful handling of uncertainty, transparent reasoning pathways, and specialized analytical workflows.

Long COVID illustrates the challenge. Its clinical and biological heterogeneity makes it difficult to evaluate research across symptoms, organ systems, immune and metabolic pathways, neurological and autonomic findings, potential biomarkers, and treatment approaches. Evidence evolves continuously, while the research community lacks a single integrated intelligence environment that can organize the landscape and support systematic exploration.

This fragmentation creates an opportunity to improve research efficiency, evidence discovery, hypothesis development, and scientific decision support.

2. The Proposed Platform

The Long COVID Intelligence Project will develop a specialized biomedical AI research environment with five interconnected capabilities.

Biomedical Knowledge Engine

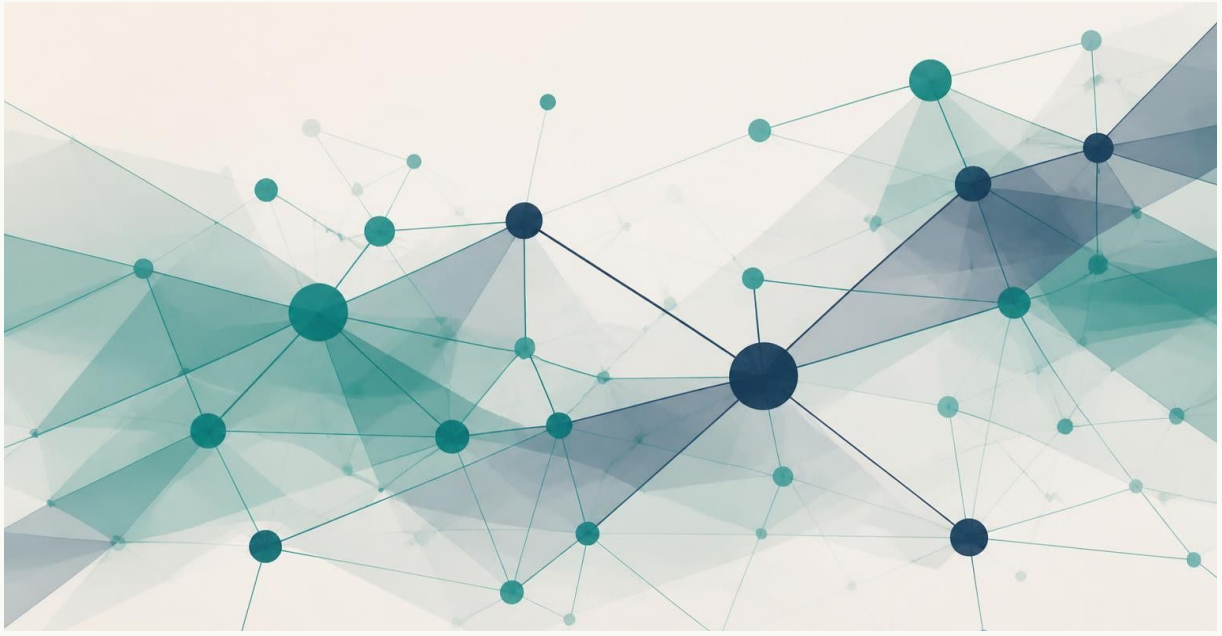
The platform will integrate scientific literature, clinical studies, appropriately licensed datasets, research registries, molecular findings, disease mechanisms, genes, proteins, pathways, biomarkers, drugs, interventions, and outcomes. Data will be normalized and indexed to support cross-source discovery, provenance, and structured relationships.

Evidence-Grounded AI Research Assistant

Researchers will be able to pose complex questions and receive synthesized responses that link back to source material. The assistant will support literature summaries, evidence comparisons, mechanistic explanations, research-gap identification, and hypothesis exploration. Outputs will clearly distinguish established findings, emerging evidence, conflicting results, and hypotheses.

Biomedical Knowledge Graph

A knowledge graph will map relationships among clinical phenotypes, symptoms, disease mechanisms, genes, proteins, immune pathways, organs, biomarkers, medications, interventions, and outcomes. This structure is intended to support discovery of relationships that may be difficult to find through keyword search alone.



2. The Proposed Platform (continued)

Research Analytics

Potential analytical modules include phenotype clustering, biomarker prioritization, pathway analysis, treatment-response exploration, drug repurposing research, disease comparisons, publication trends, and research landscape analysis. The development and release of each capability will depend on data availability, technical feasibility, and validation.

Research Dashboard and Collaboration Tools

The user interface will provide scientific search, evidence summaries, interactive knowledge maps, research alerts, publication analytics, treatment intelligence, reports, and collaboration functions. Enterprise versions may add access management, administrative controls, integrations, APIs, and customized research environments.

3. Initial Market and Expansion Strategy

The initial market is organizations and professionals conducting research into Long COVID, post-infectious illness, immune dysfunction, neurological disease, complex chronic illness, and related therapeutic development.

Primary customer groups include:

Pharmaceutical companies: scientific landscape analysis, target identification, mechanism research, drug repurposing, competitive intelligence, and therapeutic opportunity assessment.

Biotechnology companies: biomarker research, pipeline development, preclinical research support, disease mechanism exploration, and specialized scientific intelligence.

Academic institutions and medical centers: investigators, research departments, medical schools, translational science centers, and multidisciplinary research networks.

Healthcare and research organizations: research programs, specialty clinics, clinical research groups, and data-driven initiatives.

Government and nonprofit research entities: disease foundations, public health programs, research institutes, and grant-funded collaborations.

Life-science investment and consulting organizations: research landscape evaluation, scientific diligence, pipeline intelligence, and disease-area analysis.



3. Initial Market and Expansion Strategy (continued)

Expansion into ME/CFS, dysautonomia and POTS, autoimmune disorders, neuroinflammatory conditions, Alzheimer's disease and related disorders, post-infectious syndromes, and other complex diseases may provide additional markets. Such expansion will be pursued only after assessing scientific relevance, data requirements, competitive conditions, and customer demand.

The company should avoid claiming a specific total addressable market until market research, customer interviews, pricing tests, and credible industry data have been completed.

4. Product Portfolio and Commercial Model

Revenue will be diversified across several products, with recurring subscriptions and enterprise contracts intended to become the core.

Researcher subscriptions: Individual access with literature intelligence, AI-assisted synthesis, citations, research alerts, and advanced tools. Illustrative annual pricing ranges from $2,400 to $12,000, depending on features and usage.

Institutional licenses: University, medical-center, research-institute, and departmental licenses. Illustrative pricing ranges from $25,000 to $250,000 or more annually based on user count, scope, integrations, support, and deployment requirements.

Pharmaceutical and biotechnology enterprise licenses: Customized research intelligence, analytical workflows, enterprise access, integrations, and specialized disease-area capabilities. Illustrative annual contracts range from $100,000 to $1 million or more, subject to contract scope and demonstrated value.

API and data products: Licensing of knowledge graph access, research data products, AI services, evidence summaries, analytics, and customized data feeds. Illustrative contracts range from $25,000 to $250,000 or more per year.

Research partnerships: Collaborative projects with universities, foundations, biotechnology firms, pharmaceutical companies, and research organizations. Contract values may range from $100,000 to several million dollars, depending on scope, duration, funding, and deliverables.

Scientific intelligence and analytics: Customized research landscape reports, therapeutic intelligence, biomarker reviews, evidence synthesis, and project-based analytical services.

Pricing will be tested through pilots and customer discovery. Published price ranges are preliminary and do not imply existing customer commitments.

5. Go-to-Market Plan

The commercialization strategy will begin with a focused group of founding customers and research partners. Early pilots will validate use cases, product usability, scientific quality, willingness to pay, integration requirements, and measurable time savings.

Year 1 will emphasize design partners, prototype validation, and pilot contracts with academic research groups, biotechnology organizations, and pharmaceutical research teams. The goal is to secure approximately five founding customers or pilot engagements.

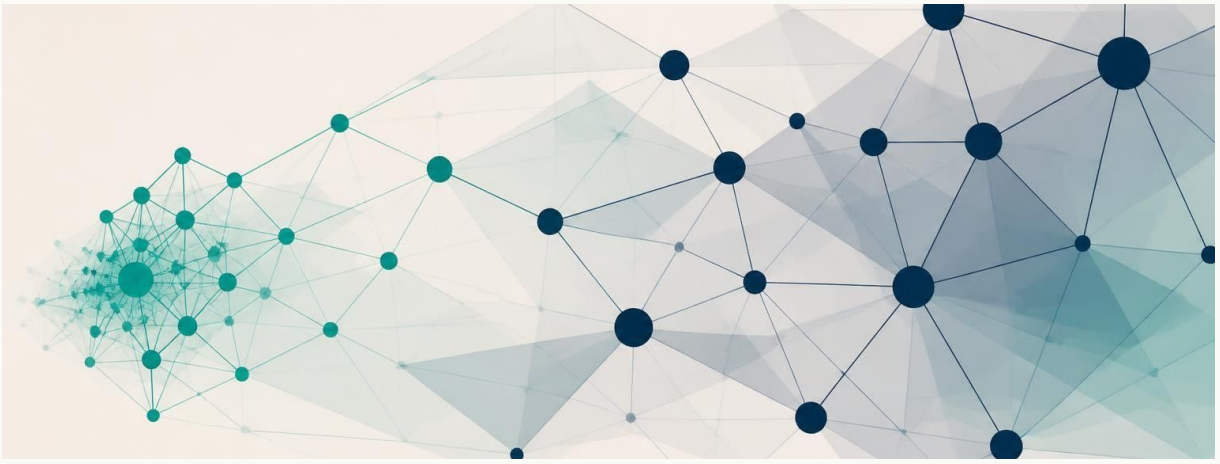
Year 2 will focus on commercial launch, conversion of pilots to paid contracts, institutional subscriptions, conferences, demonstrations, case studies, and partner channels. The planning target is approximately 20 paying customers or contracts.

Year 3 will expand enterprise sales, institutional licensing, APIs, and specialized analytics, with an illustrative target of approximately 50 customers or contracts.

Year 4 will develop strategic accounts, pharmaceutical partnerships, international opportunities, and additional disease modules, targeting approximately 100 customers or contracts.

Year 5 will seek a mature subscription and enterprise business with approximately 160 customer relationships. An illustrative mix includes 70 individual or small research subscriptions, 45 academic or institutional accounts, 30 pharmaceutical or biotech enterprises, 10 API/data customers, and five strategic research partnerships. These counts are assumptions and should be refined as actual pricing, contract size, and sales-cycle data become available.

The customer acquisition approach will combine scientific credibility, direct enterprise sales, academic partnerships, conference participation, demonstrations, pilot programs, publications, referral channels, and targeted content marketing. Enterprise sales cycles may be long; pilots and clear value demonstrations will be important.



6. Technology and NVIDIA AI Infrastructure

The platform requires computational capacity for data processing, model evaluation, fine-tuning where appropriate, retrieval, inference, analytics, storage, and secure deployment.

The recommended architecture is hybrid. Cloud GPU resources will support flexible development and variable demand. Reserved capacity can be used for predictable workloads. Dedicated NVIDIA GPU infrastructure—including DGX-class systems where justified—may be deployed selectively when performance, utilization, security, control, or economics support it.

Infrastructure planning includes GPU compute, training and inference capacity, storage, high-speed networking, data pipelines, orchestration, monitoring, licensing, cybersecurity, backup, disaster recovery, and model operations.



The five-year planning allowance for AI and infrastructure is **$12.2 million**: $1.8 million in Year 1, $2.0 million in Year 2, $2.4 million in Year 3, $2.8 million in Year 4, and $3.2 million in Year 5. This is not a vendor quotation. Actual costs will depend on workload benchmarks, architecture, hardware purchases, cloud pricing, utilization, data volumes, and model choices.

The company should benchmark workloads and utilization before committing to substantial dedicated hardware. A staged hybrid strategy preserves flexibility and reduces the risk of premature capital expenditure.

7. Organization and Staffing

The company is expected to grow from approximately 13 employees in Year 1 to 34 employees in Year 5. Staffing will include executive and product leadership, AI/ML engineering, data science, biomedical science, clinical informatics, data engineering and interoperability, cybersecurity, MLOps and platform reliability, implementation, customer success, quality assurance, technical support, and commercial partnerships.

The financial model assumes blended base compensation of approximately $150,000 per employee, with benefits and payroll burden estimated at 30%, or approximately $195,000 in average loaded personnel cost. Actual compensation will vary materially by role, seniority, geography, and recruiting market.

Headcount planning is **13, 18, 24, 29, and 34 employees** in Years 1 through 5, respectively. Hiring will be staged against product milestones, customer traction, funding, and workload.

Leadership capabilities should include commercial strategy, technology architecture, biomedical research, AI/ML, enterprise sales, finance, security, and governance. A scientific advisory board should include expertise relevant to Long COVID, immunology, neurology, infectious disease, genomics, bioinformatics, clinical research, AI, and drug development.

8. Five-Year Revenue Forecast

The base-case model projects annual revenue of $0.5 million in Year 1, $2 million in Year 2, $6 million in Year 3, $14 million in Year 4, and $28 million in Year 5. Cumulative revenue is $50.5 million.

Projected revenue by stream (millions):

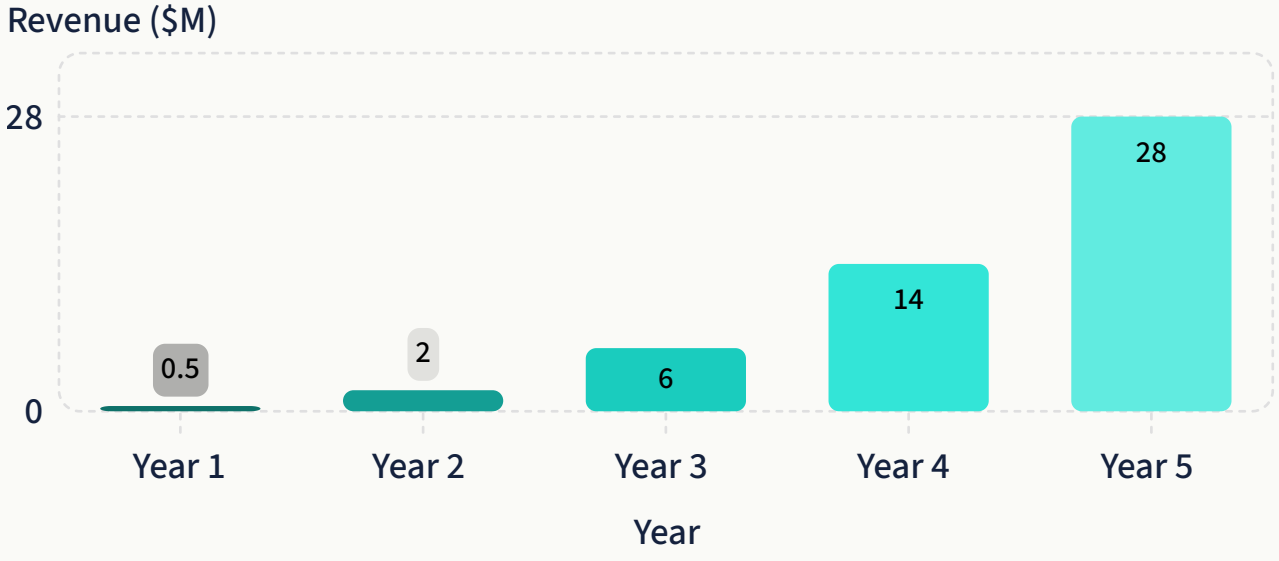
Year 1: subscriptions $0.15; enterprise licensing $0.10; API/data $0.05; research partnerships $0.15; consulting/analytics $0.05; total $0.50.
Year 2: subscriptions $0.80; enterprise licensing $0.50; API/data $0.25; research partnerships $0.35; consulting/analytics $0.10; total $2.00.
Year 3: subscriptions $2.50; enterprise licensing $1.80; API/data $0.80; research partnerships $0.70; consulting/analytics $0.20; total $6.00.
Year 4: subscriptions $6.00; enterprise licensing $4.50; API/data $2.00; research partnerships $1.20; consulting/analytics $0.30; total $14.00.
Year 5: subscriptions $12.00; enterprise licensing $9.00; API/data $4.00; research partnerships $2.50; consulting/analytics $0.50; total $28.00.

The forecast represents an ambitious growth scenario. It assumes successful product development, conversion of pilots, enterprise adoption, customer retention, pricing acceptance, and an expanding pipeline. No revenue is currently implied to be contracted or guaranteed. A downside and upside case should be developed alongside this base case before fundraising.

Illustrative projected revenue by stream (millions)

Year	Subscriptions	Enterprise licensing	API/data	Research partnerships	Consulting/analytics	Total
Year 1	0.15	0.10	0.05	0.15	0.05	0.50
Year 2	0.80	0.50	0.25	0.35	0.10	2.00
Year 3	2.50	1.80	0.80	0.70	0.20	6.00
Year 4	6.00	4.50	2.00	1.20	0.30	14.00
Year 5	12.00	9.00	4.00	2.50	0.50	28.00

Illustrative projected revenue by year (millions)



9. Five-Year Operating Budget

The operating budget totals approximately **$63.46 million** over five years.

Personnel salaries are estimated at $17.70 million; benefits and payroll taxes at $5.31 million; NVIDIA AI infrastructure at $12.20 million; software tools and SaaS at $3.25 million; data acquisition and research partnerships at $3.60 million; implementation and deployment at $4.05 million; support operations at $1.75 million; cybersecurity, privacy, and compliance at $3.00 million; general and administrative costs at $4.00 million; sales, marketing, partnerships, and travel at $5.10 million; and contingency and operating reserve at $3.50 million.

Annual operating expenses are projected at $7.535 million in Year 1, $10.010 million in Year 2, $12.780 million in Year 3, $15.305 million in Year 4, and $17.830 million in Year 5.

Personnel includes employee compensation and benefits. Implementation, support, software, infrastructure, data, compliance, G&A, and commercialization categories represent non-payroll or separately budgeted operating costs as modeled. The budget should be refined to prevent overlap, distinguish capital expenditures from operating expenses, and establish accounting treatment for equipment, software development, and cloud costs.

10. Operating Results and Path to Profitability

The base-case operating forecast is:

Year 1: revenue $0.50 million; expenses $7.535 million; operating loss $7.035 million.

Year 2: revenue $2.00 million; expenses $10.010 million; operating loss $8.010 million.

Year 3: revenue $6.00 million; expenses $12.780 million; operating loss $6.780 million.

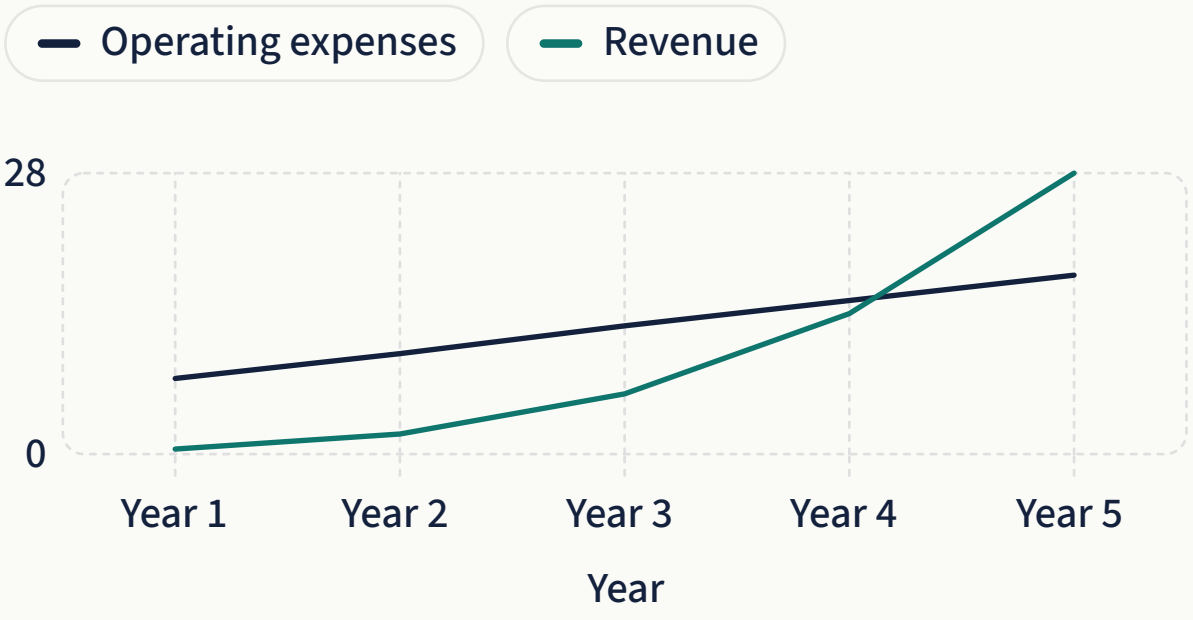
Year 4: revenue $14.00 million; expenses $15.305 million; operating loss $1.305 million.

Year 5: revenue $28.00 million; expenses $17.830 million; operating profit $10.170 million.

Across the five-year period, cumulative projected revenue is $50.5 million, cumulative operating expenses are $63.46 million, and cumulative operating deficit is approximately $12.96 million.

The model targets annual operating profitability in Year 5. It does not include interest expense, income taxes, financing costs, depreciation, capitalization adjustments, or detailed GAAP treatment. Cash requirements may differ from operating losses because of capital expenditures, working capital, timing of revenue, payment terms, and financing structure. A detailed cash-flow forecast is essential before investment decisions.

Illustrative revenue vs operating expenses (millions)



11. Capital Strategy

An illustrative initial capital target of **$25–30 million** is proposed, raised in stages and tied to development and commercial milestones.

Phase 1: approximately $8–10 million to establish the founding team, architecture, initial data integration, AI infrastructure, prototype, validation, and pilot programs. Target runway is approximately 12–18 months, subject to actual spending and timing.

Phase 2: approximately $10–12 million to expand the platform, scientific and engineering teams, enterprise products, integrations, customer acquisition, and commercial partnerships.

Phase 3: approximately $7–10 million to scale enterprise sales, APIs, data products, disease modules, partnerships, and broader commercialization. This phase may be reduced or avoided if revenues, grants, strategic investment, or customer prepayments materially improve cash generation.

An illustrative allocation of $30 million is: technology development $7 million; personnel and scientific development $8 million; AI infrastructure $5 million; data acquisition and research integration $2.5 million; commercialization $2.5 million; security, legal, and compliance $1 million; and working capital and reserve $4 million.

Financing may include preferred equity, convertible instruments, strategic investment, corporate venture capital, grants, research partnerships, and non-dilutive funding. Structure and terms should be determined with professional legal and financial advice and should reflect whether the entity is a commercial company, nonprofit, or hybrid.

12. Development Milestones

Year 1 — Foundation: Establish company and governance; recruit founding team; develop architecture; integrate initial sources; create knowledge base and prototype; establish security and data policies; validate initial use cases; secure pilot customers.

Year 2 — Product Launch: Launch commercial subscriptions; develop institutional offerings; expand dashboards and knowledge graph; implement API; convert pilots; establish research partnerships and customer support.

Year 3 — Expansion: Develop advanced analytics, biomarker and therapeutic research modules; expand enterprise deployments; integrate more sources; validate new disease-area opportunities; build commercial capacity.

Year 4 — Scale: Expand enterprise contracts and pharmaceutical partnerships; develop international opportunities; increase product breadth; strengthen operations; move toward break-even.

Year 5 — Commercial Maturity: Expand recurring revenue; improve retention; pursue profitability; scale enterprise and data products; extend disease coverage where validated; evaluate strategic partnerships and future growth opportunities.

Milestones should be measured by product readiness, data quality, validated workflows, customer usage, pilot conversion, recurring revenue, retention, gross margin, sales-cycle duration, operating efficiency, and financing runway.

13. Intellectual Property and Defensibility

Potential intellectual property assets include biomedical knowledge graph architecture, proprietary data integration methods, AI workflows, specialized models, annotations, analytical methods, research software, APIs, datasets, and validated research processes.

The company should protect its assets through appropriate copyrights, trade secrets, contractual rights, trademarks, patents where justified, licensing agreements, and access controls. Data agreements must define rights to ingest, analyze, retain, use, redistribute, and commercialize source material and derivative outputs.

Defensibility is expected to develop over time through the quality and breadth of curated knowledge, specialized workflows, validated analytical capabilities, user adoption, proprietary annotations, integration relationships, and accumulated customer feedback—not simply through use of a particular AI model or GPU platform.

14. Data Governance, Security, and Scientific Integrity

The platform will emphasize source provenance, scientific attribution, quality control, security, privacy, access management, encryption, audit logs, and appropriate licensing.

Initial products can focus on public scientific information and appropriately licensed research data. Any future use of identifiable patient information or protected health information will require suitable legal, technical, contractual, ethical, and regulatory safeguards.

AI outputs must clearly distinguish source-supported conclusions from interpretations and hypotheses. The platform is intended to support research and scientific decision-making; it should not present unvalidated AI-generated hypotheses as established evidence or clinical recommendations.

Scientific oversight, expert review, documented validation, and reproducible analytical workflows will be central to credibility.

15. Principal Risks and Mitigation

Scientific and technical risk: Models may generate inaccurate or weak hypotheses, and data may be incomplete or inconsistent. Mitigation includes source-grounded outputs, expert review, validation, quality measures, and clear uncertainty reporting.

Infrastructure risk: GPU costs, hardware obsolescence, and utilization may exceed assumptions. Mitigation includes hybrid cloud infrastructure, workload benchmarking, staged commitments, vendor flexibility, and capacity monitoring.

Commercial risk: Enterprise adoption and sales cycles may be slower than forecast. Mitigation includes pilot programs, customer discovery, clear value propositions, diversified segments, and staged hiring.

Data rights and integration risk: Licensing, normalization, and access may be difficult or expensive. Mitigation includes prioritizing accessible and licensed sources, due diligence, and scalable ingestion systems.

Capital risk: Funding may be insufficient before commercialization or revenue may fall short. Mitigation includes milestone-based financing, contingency planning, non-dilutive opportunities, and disciplined expenditure.

Competitive risk: General AI providers and specialized research platforms may compete. Mitigation includes deep disease-area intelligence, proprietary relationships, validated workflows, evidence transparency, and customer integration.

16. Management and Governance

The leadership structure should combine commercial execution with scientific and technical credibility.

The Chief Executive Officer will lead corporate strategy, fundraising, partnerships, and commercialization.	The Chief Technology Officer will oversee architecture, engineering, platform reliability, and infrastructure.
Scientific leadership will direct biomedical research strategy, validation, and research partnerships.	AI/ML leadership will develop model and analytics capabilities.
Commercial leadership will oversee sales, licensing, and enterprise relationships.	Finance leadership will manage budgeting, capital planning, accounting, and investor reporting.

A multidisciplinary scientific advisory board will support scientific priorities, research quality, validation, and strategic partnerships. Governance should include appropriate oversight of financial controls, data rights, security, scientific integrity, conflicts of interest, and investment reporting.

17. Investment Thesis and Conclusion

The Long COVID Intelligence Project proposes a specialized biomedical AI platform addressing a meaningful research challenge: the fragmentation of scientific evidence and the difficulty of connecting findings across complex biological systems.

The commercial model combines recurring subscriptions, institutional licensing, enterprise contracts, API/data products, research collaborations, and scientific analytics. The architecture is intended to scale beyond Long COVID as new disease modules are validated and market opportunities demonstrated.

The five-year base-case model projects $50.5 million cumulative revenue, $63.46 million cumulative operating expenses, and annual operating profitability in Year 5. These assumptions are ambitious and must be validated through customer discovery, pilot performance, market pricing, vendor quotes, product development milestones, and a more detailed financial model.

The investment opportunity is to build enduring biomedical intelligence infrastructure: a proprietary knowledge base, connected research graph, evidence-grounded AI workflows, specialized analytics, enterprise products, and recurring commercial relationships.

The objective is not to promise that AI will solve Long COVID. It is to build a credible, scalable technology and research business that helps investigators and life-science organizations navigate the evidence more efficiently, explore scientific relationships, prioritize questions, and accelerate the research process.

Long COVID Intelligence Project — Building a new generation of AI-assisted biomedical research intelligence.

Financial projections and pricing in this plan are illustrative assumptions for strategic planning. They are not audited forecasts, commitments, or guarantees of results. Commercialization requires further market validation, customer discovery, technical assessment, data-rights review, and detailed financial diligence.