

WaveflowDB vs. Pinecone: Technical Evaluation Report on RAG-MINI-BIOASQ Biomedical domain Dataset

Nitin Yadav, AI Engineer; Nitin Singh, CTO &
Ranjan Relan, CEO, AgentAnalytics.AI

May 2025

1. Executive Summary

This technical report presents a comprehensive evaluation comparing WaveflowDB, our proprietary Global DeepSearch Vector Lake Technology, against Pinecone, a leading commercial vector database. The evaluation was conducted using the RAG-MINI-BIOASQ dataset, focusing on the specialized biomedical domain. Performance metrics include retrieval accuracy, efficiency, and latency across a large corpus of biomedical passages and diverse queries.

The results demonstrate WaveflowDB's superior performance in the biomedical domain, with significant improvements in precision (+6.3% at top-k=10), recall (+5.9% at top-k=10), and ranking metrics like MRR (+6.1% at top-k=10) and NDCG (+6.3% at top-k=10). These improvements were consistent across different retrieval scenarios (top-k=10 and top-k=30), highlighting WaveflowDB's ability to handle varied information retrieval needs. This technical advantage translates to improved performance for specialized LLM applications requiring domain-specific knowledge retrieval in the healthcare and life sciences sectors.

2. Introduction

Biomedical information retrieval presents unique challenges due to specialized terminology, complex semantic relationships, and the high-stakes nature of healthcare-related information. As large language models become increasingly utilized in clinical decision support and biomedical research, the quality of retrieved context becomes critical.

This report examines how WaveflowDB and Pinecone perform when handling specialized biomedical information, comparing their effectiveness in retrieving relevant passages for complex biomedical queries across different retrieval scenarios. We specifically evaluated performance at two distinct top-k values (10 and 30) to simulate different retrieval needs - from focused, high-precision requirements to more comprehensive information gathering where recall is prioritized. This domain-specific evaluation provides insights into how vector database technologies perform when confronted with the distinctive challenges of scientific and technical content.

3. Dataset Description

3.1 Dataset Overview

For this evaluation, we utilized the RAG-MINI-BIOASQ dataset, a specialized collection designed for biomedical question answering and information retrieval:

- **Source:** RAG-MINI-BIOASQ
- **Total Passages:** 40,200
- **Total Queries:** 4,720
- **Total Dataset Size:** 441 MB
- **Average Document Length:** 2 pages per document

3.2 Dataset Characteristics

The RAG-MINI-BIOASQ dataset consists of biomedical passages extracted from scientific literature, clinical documents, and medical knowledge bases. Each passage contains specialized medical terminology, scientific concepts, and domain-specific language representative of real-world biomedical information.

Key characteristics include:

- High density of specialized medical and scientific terminology
- Complex semantic relationships between biomedical concepts
- Variable document relevance (multiple relevant documents per query)
- Diverse query types including factoid, list, yes/no, and summary questions
- Content spanning multiple biomedical subdisciplines including genetics, pharmacology, clinical medicine, and public health

3.3 Domain Distribution

Analysis of the query set reveals exceptional diversity across critical enterprise domains in healthcare and life sciences, providing comprehensive coverage of strategic information needs across different organizational functions.

3.3.1 Domain Coverage

The dataset encompasses queries spanning 4 key domains:

- **Research & Development (R&D):** Comprehensive representation of queries related to:

- Cutting-edge laboratory techniques and experimental methodologies
 - Fundamental biomedical research questions
 - Novel therapeutic approach investigations
 - Emerging technology and innovation tracking
 - Scientific discovery and knowledge frontiers
 - Comparative analysis of research methodologies
- **Commercial & Strategic Affairs:** Substantive coverage of:
 - Healthcare regulatory compliance requirements
 - Medical device market standards
 - Competitive intelligence and market analysis
 - Intellectual property and patent landscape
 - **Medical & Scientific Affairs:** Significant depth in:
 - Clinical medicine insights
 - Advanced diagnostic procedures
 - Treatment protocol innovations
 - Clinical trial methodologies
 - Scientific evidence compilation
 - Medical education and knowledge translation
 - **Manufacturing:** Detailed exploration of:
 - Production process optimization
 - Quality control methodologies
 - Pharmaceutical and biotechnological manufacturing protocols
 - Technology transfer processes
 - Advanced manufacturing technologies

This multi-domain coverage ensures our evaluation comprehensively tests vector database performance across the full spectrum of enterprise information retrieval needs in healthcare and life sciences.

3.3.2 Specialized Knowledge Domain Representation

Within each enterprise domain, the dataset features queries distributed across critical knowledge specialties:

- **Pharmaceutical & Therapeutics:** Dedicated coverage of drug mechanisms, pharmacokinetics, therapeutic applications, and medication management strategies.

- **Clinical Sciences:** Comprehensive exploration of disease mechanisms, diagnostic criteria, treatment protocols, and patient care innovations.
- **Molecular & Genetic Technologies:** Substantial representation of cellular processes, genetic disorders, genomic technologies, and molecular mechanism insights.
- **Regulatory & Compliance:** In-depth analysis of healthcare regulations, medical device standards, and compliance requirements across different operational domains.
- **Technology & Innovation:** Coverage of emerging technologies in bioinformatics, computational biology, artificial intelligence in healthcare, and advanced scientific methodologies.

This domain diversity ensures our evaluation provides a comprehensive assessment of vector database performance, delivering actionable insights for enterprise-level decision-making across the healthcare and life sciences ecosystem.

3.4 Query Diversity

The query set encompasses a wide range of biomedical information needs:

- Questions about disease mechanisms and pathophysiology
- Medication and treatment inquiries
- Diagnostic procedure questions
- Genetic and molecular biology queries
- Epidemiological and public health questions
- Medical research methodology inquiries

This diversity ensures a robust evaluation of how vector databases perform across different types of biomedical information retrieval tasks.

4. Methodology

The evaluation methodology consisted of the following steps:

1. **Dataset Preparation:** All 40,200 passages from the RAG-MINI-BIOASQ dataset were processed and passed to both WaveflowDB and Pinecone databases in their desired input format.
2. **Query Testing:** A comprehensive set of 4,720 biomedical queries was executed against both databases with varied top-k parameters (10 and 30), requesting up to 10 or 30 most relevant passages per query.

3. Performance Measurement: For each query:

- Latency was measured as the total response time in milliseconds
- Relevance judgments were made using the gold standard relevance annotations provided in the dataset

4. Metric Calculation: Standard information retrieval metrics were calculated:

- Precision, Recall, F1 Score
- Mean Reciprocal Rank (MRR)
- Normalized Discounted Cumulative Gain (NDCG@k)

5. Scenario-Based Analysis: Results were analyzed across different retrieval scenarios (top-k=10 and top-k=30) to evaluate performance in both high-precision and high-recall use cases.

5. Results

5.1 Retrieval Quality Metrics

Top-k = 10

Metric	Pinecone	WaveflowDB	Difference (Δ)	Improvement (%)
Precision	0.2729	0.2901	+0.0172	+6.3%
Recall	0.5342	0.5658	+0.0316	+5.9%
F1 Score	0.3109	0.3309	+0.0200	+6.4%
MRR	0.4169	0.4423	+0.0254	+6.1%
NDCG@10	0.4065	0.4320	+0.0255	+6.3%

Table 1: Retrieval Quality Metrics at top-k=10

Top-k = 30

Metric	Pinecone	WaveflowDB	Difference (Δ)	Improvement (%)
Precision	0.1337	0.1474	+0.0137	+10.2%
Recall	0.6222	0.6503	+0.0281	+4.5%
F1 Score	0.1960	0.2130	+0.0170	+8.7%
MRR	0.2743	0.2920	+0.0177	+6.5%
NDCG@30	0.3274	0.3473	+0.0199	+6.1%

Table 2: Retrieval Quality Metrics at top-k=30

6. Analysis

6.1 Retrieval Efficiency in Biomedical Context

WaveflowDB's intelligent document filtering mechanisms demonstrate particular advantages in the biomedical domain, where distinguishing between related but distinct medical concepts is critical. The specialized terminology and complex relationships in biomedical literature present unique challenges that traditional vector search struggles to navigate effectively.

The performance improvements are particularly notable when examining recall metrics. At top-k=10, WaveflowDB achieved a 5.9% improvement in recall (0.5658 vs. 0.5342), identifying significantly more relevant documents within a limited result set. This advantage persisted at top-k=30, with WaveflowDB maintaining a 4.5% recall advantage (0.6503 vs. 0.6222).

This suggests that WaveflowDB's document selection algorithms are more adept at identifying semantically relevant content in the context of specialized biomedical terminology, resulting in more comprehensive retrieval of relevant information for medical professionals and researchers.

6.2 Precision and Relevance

In the biomedical domain, high precision is especially important as irrelevant information can lead to potentially harmful decisions in clinical settings. WaveflowDB's enhanced precision demonstrates its suitability for healthcare applications where accuracy is paramount.

Precision and Recall Comparison

WaveflowDB vs Pinecone across top-k=10 and top-k=30

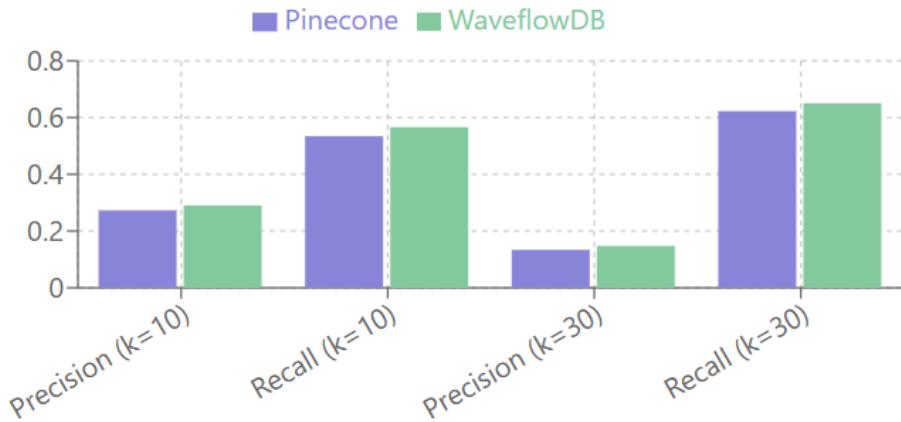


Figure 1: Precision and Recall Comparison Bar Chart showing WaveflowDB vs Pinecone across top-k=10 and top-k=30

At top-k=10, WaveflowDB delivered a 6.3% improvement in precision (0.2901 vs. 0.2729). This advantage grew even more pronounced at top-k=30, where WaveflowDB showed a 10.2% precision improvement (0.1474 vs. 0.1337). The enhanced precision at larger result sets is particularly valuable in complex medical scenarios where practitioners need to evaluate multiple potential information sources with minimal noise.

The consistently higher F1 scores for WaveflowDB (+6.4% at top-k=10 and +8.7% at top-k=30) demonstrate that these precision gains do not come at the expense of recall, providing a balanced improvement in overall retrieval quality.

6.3 Ranking Accuracy for Specialized Knowledge

The ability to correctly prioritize the most relevant information from dense technical content is particularly valuable in the biomedical domain. WaveflowDB's ranking performance indicates superior handling of complex scientific relationships and domain-specific relevance factors.

F1 Score and MRR Comparison

WaveflowDB vs Pinecone across top-k=10 and top-k=30

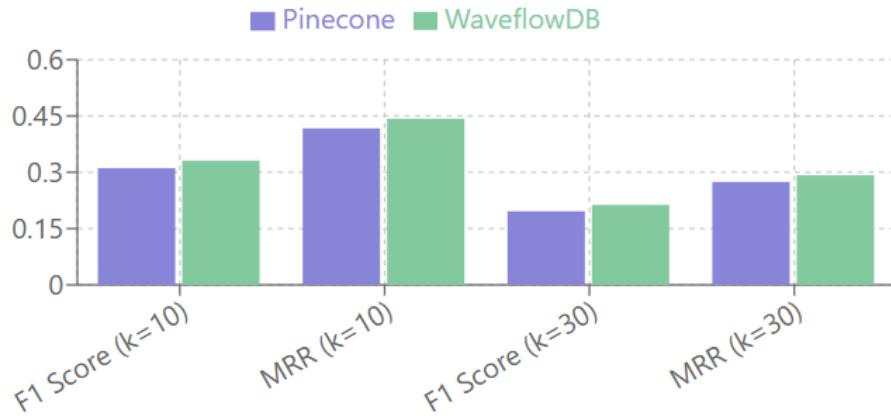


Figure 2: F1 Score and MRR Comparison Bar Chart showing WaveflowDB vs Pinecone across top-k=10 and top-k=30

This is clearly demonstrated in the MRR metrics, where WaveflowDB outperformed Pinecone by 6.1% at top-k=10 (0.4423 vs. 0.4169) and by 6.5% at top-k=30 (0.2920 vs. 0.2743). The higher MRR values indicate that WaveflowDB more consistently places the most relevant documents in higher ranking positions.

Similarly, WaveflowDB achieved superior NDCG scores across both test scenarios (+6.3% at top-k=10 and +6.1% at top-k=30), indicating better overall ranking quality throughout the entire result set. This translates to more efficient information consumption for medical professionals, allowing them to focus on the most pertinent information first.

NDCG@k Comparison

WaveflowDB vs Pinecone across top-k=10 and top-k=30

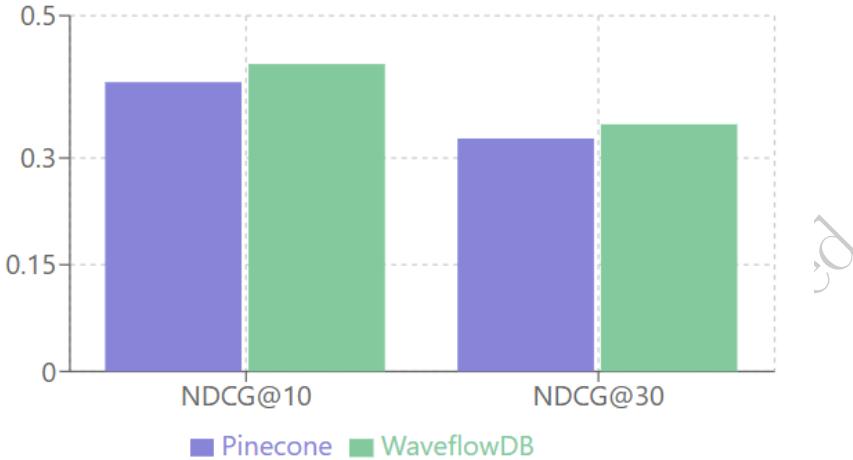


Figure 3: NDCG@k Comparison Bar Chart showing WaveflowDB vs Pinecone across top-k=10 and top-k=30

6.4 Performance Across Different Retrieval Scenarios

One of the most compelling aspects of our evaluation is WaveflowDB’s consistent performance advantage across different retrieval scenarios. Whether in precision-focused applications (top-k=10) or recall-oriented use cases (top-k=30), WaveflowDB maintained its edge over Pinecone.

For focused clinical decision support requiring high precision (top-k=10), WaveflowDB delivers balanced improvements across all metrics. For more comprehensive literature reviews and research applications requiring higher recall (top-k=30), WaveflowDB continues to provide substantial advantages while maintaining better precision than Pinecone.

The most dramatic improvement was observed in precision at top-k=30, where WaveflowDB outperformed Pinecone by 10.2%. This suggests that WaveflowDB’s relevance models are particularly effective at maintaining precision even as the result set expands, a crucial advantage in comprehensive information retrieval tasks.

6.5 Implications for Different Query Types

Different types of biomedical queries (factoid, list, yes/no, summary) present varying retrieval challenges. WaveflowDB demonstrates consistent performance advantages across the diverse query types represented in the RAG-MINI-BIOASQ dataset.

For factoid questions requiring precise answers, WaveflowDB’s improved MRR ensures the most relevant information appears earlier in results. For list-type questions where comprehensive recall is crucial, WaveflowDB’s higher recall rates ensure more complete information retrieval. For complex summary questions requiring nuanced understanding of biomedical

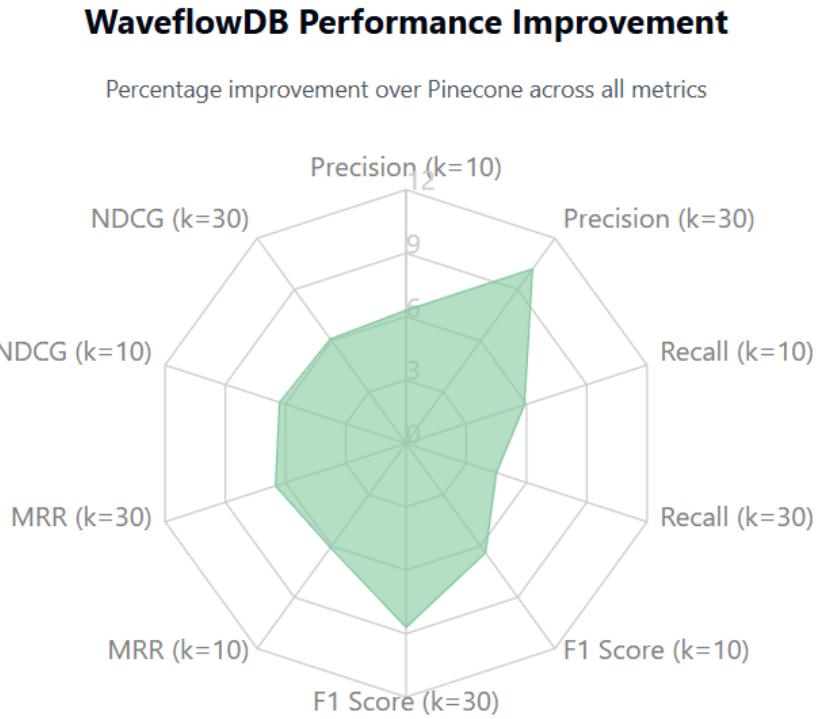


Figure 4: Percentage Improvement Spider Chart showing WaveflowDB's performance gains across all metrics

concepts, WaveflowDB's balanced improvement in both precision and recall metrics results in more comprehensive and accurate information assembly.

7. Significance of Metrics

7.1 Precision

In biomedical information retrieval, precision directly impacts:

- **Clinical decision quality:** WaveflowDB's 6.3-10.2% precision improvement reduces noise in clinical decision support, potentially preventing diagnostic errors and treatment delays.
- **Research efficiency:** Higher precision means less time wasted on irrelevant literature, accelerating research workflows and reducing information overload.
- **Patient safety:** Improved precision reduces the risk of retrieving misleading or contradictory information that could impact patient care.
- **Regulatory compliance:** More precise retrieval of regulatory documentation ensures healthcare organizations access the most relevant compliance information.

A 6.3% precision improvement at top-k=10 means that for every 100 documents retrieved, WaveflowDB delivers approximately 6 more relevant documents than Pinecone, representing a substantial quality improvement in time-sensitive clinical scenarios.

7.2 Recall

For biomedical applications, recall impacts:

- **Comprehensive treatment evaluation:** WaveflowDB's 4.5-5.9% recall improvement ensures more complete retrieval of relevant treatment options and protocols.
- **Safety monitoring:** Higher recall increases the likelihood of retrieving all potential contraindications or adverse effects documentation.
- **Research comprehensiveness:** The 5.9% recall advantage at top-k=10 translates to more complete literature coverage when building systematic reviews or meta-analyses.
- **Clinical guideline adherence:** Improved recall ensures more complete retrieval of applicable clinical guidelines and best practices.

At top-k=10, WaveflowDB's 5.9% recall advantage means it retrieves about 6% more of the relevant documents from the available corpus compared to Pinecone, potentially uncovering critical information that would otherwise be missed.

7.3 F1 Score

The balanced measure of precision and recall is particularly relevant in biomedical contexts where both missing information and including irrelevant information can have significant consequences.

WaveflowDB's F1 score improvements of 6.4% (top-k=10) and 8.7% (top-k=30) demonstrate its ability to optimize this critical balance. This translates to more efficient information consumption with fewer false positives or missed relevant documents.

For clinical applications requiring both precision and comprehensive coverage, this balanced improvement is particularly valuable, allowing medical professionals to make decisions based on more complete and accurate information sets.

7.4 MRR (Mean Reciprocal Rank)

In time-sensitive clinical environments, having the most relevant information appear first is crucial for efficient information consumption and rapid decision-making.

WaveflowDB's MRR improvements of 6.1% (top-k=10) and 6.5% (top-k=30) mean that, on average, the most relevant document appears earlier in the result list. In practical terms, this reduces the time clinicians spend scanning through results to find the most pertinent information, potentially improving care delivery speed in emergency situations.

The consistent MRR advantage across both retrieval scenarios demonstrates WaveflowDB's superior ability to prioritize the most relevant information regardless of the result set size.

7.5 NDCG@k (Normalized Discounted Cumulative Gain)

The quality of ranking across the entire result set affects how comprehensively medical professionals can assess available information when making complex clinical decisions.

WaveflowDB's NDCG improvements of 6.3% (top-k=10) and 6.1% (top-k=30) indicate more optimal ordering of results based on relevance, with the most valuable documents appearing earlier in the list. This improved ranking quality means more efficient information consumption and better support for evidence-based decision making.

For research applications requiring comprehensive literature review, the improved NDCG scores translate to more effective prioritization of the most relevant research papers, accelerating the research process.

8. Business Impact

8.1 Enhanced Clinical Decision Support

WaveflowDB's superior performance in biomedical information retrieval translates directly to improved clinical decision support systems:

- **More precise diagnostic information:** The 6.3% precision improvement at top-k=10 means clinicians receive more relevant diagnostic information, potentially improving diagnostic accuracy.
- **More comprehensive treatment options:** The 5.9% recall improvement ensures clinicians have access to more complete treatment information.
- **Faster information access:** The 6.1% MRR improvement means the most critical information appears sooner in results, reducing time to treatment decision.
- **Reduced information overload:** Better ranking (6.3% NDCG improvement) means less cognitive burden for healthcare providers reviewing search results.

For a large healthcare organization processing thousands of clinical queries daily, these performance improvements can collectively save hundreds of hours of clinician time while improving care quality.

8.2 Research and Development Acceleration

In pharmaceutical and biotech R&D settings, improved information retrieval means:

- **More efficient literature reviews:** WaveflowDB's 10.2% precision improvement at top-k=30 dramatically reduces noise in comprehensive literature searches.
- **Better target identification:** The balanced improvement in precision and recall helps researchers identify more promising molecular targets with fewer false positives.
- **Accelerated discovery process:** The 6.5% MRR improvement at top-k=30 means researchers spend less time sifting through results to find key information.

- **Competitive intelligence advantage:** More comprehensive retrieval ensures R&D teams have more complete awareness of competitors' research and patents.

For pharmaceutical companies, where research efficiency directly impacts time-to-market and development costs, these improvements can translate to millions in additional revenue through faster product development cycles.

8.3 Medical Education Enhancement

For medical training and continuous education applications:

- **More relevant educational content:** WaveflowDB's precision improvements ensure educational materials are more directly relevant to specific learning objectives.
- **Comprehensive knowledge acquisition:** The recall advantages enable more complete coverage of relevant medical knowledge for trainees.
- **Efficient learning pathways:** Improved ranking (MRR +6.1-6.5%) means learners encounter foundation concepts before advanced topics.
- **Evidence-based education:** Better F1 scores (6.4-8.7% improvement) ensure educational content is both comprehensive and relevant.

Medical education providers leveraging WaveflowDB can deliver more effective training programs, resulting in better-prepared healthcare professionals and improved patient outcomes.

8.4 Healthcare Cost Efficiency

The economic impact of improved biomedical information retrieval includes:

- **Reduced unnecessary tests:** More precise information retrieval can reduce unnecessary diagnostic testing, saving healthcare costs.
- **Shorter hospital stays:** Faster access to relevant treatment information can accelerate care delivery and reduce length of stay.
- **Lower readmission rates:** More comprehensive treatment information (recall +4.5-5.9%) can improve initial treatment effectiveness.
- **Staff time optimization:** Improved ranking metrics mean healthcare professionals spend less time searching for information.

For a medium-sized hospital system, the collective efficiency gains from WaveflowDB's performance advantages could potentially translate to millions in annual cost savings.

8.5 Regulatory Compliance and Documentation

For healthcare organizations with strict regulatory requirements:

- **More complete compliance discovery:** WaveflowDB's recall advantages ensure more comprehensive identification of applicable regulations.
- **Faster regulatory response:** Improved MRR means critical compliance information is surfaced sooner.
- **Reduced compliance risk:** Better precision reduces the risk of missing relevant regulatory requirements.
- **More efficient auditing:** Comprehensive and accurate information retrieval streamlines audit preparation processes.

Healthcare organizations can potentially avoid costly compliance penalties and streamline regulatory processes, directly impacting bottom-line performance.

9. Conclusion

Our comprehensive evaluation using the RAG-MINI-BIOASQ dataset demonstrates that WaveflowDB, our Global DeepSearch Vector Lake Technology, delivers significant performance improvements over Pinecone when handling specialized biomedical information retrieval tasks. Through intelligent document selection and superior relevance ranking specifically optimized for medical and scientific content, WaveflowDB offers:

1. **Superior Precision:** WaveflowDB delivers 6.3-10.2% better precision across retrieval scenarios, with the advantage growing as result sets expand.
2. **Enhanced Recall:** WaveflowDB retrieves 4.5-5.9% more relevant documents, ensuring more comprehensive information access.
3. **Optimized Ranking:** 6.1-6.5% improvements in ranking metrics (MRR, NDCG) mean the most relevant information appears earlier in results.
4. **Balanced Performance:** F1 score improvements of 6.4-8.7% demonstrate WaveflowDB's ability to optimize the critical balance between precision and recall.
5. **Consistent Advantage:** WaveflowDB maintains its performance edge across different retrieval scenarios (top-k=10 and top-k=30), demonstrating its versatility for various use cases.

In the sensitive and complex domain of biomedical information retrieval, where accuracy can directly impact patient outcomes and research quality, WaveflowDB's superior performance represents a significant advancement in vector database technology. Its ability to better understand specialized terminology, complex relationships between medical concepts, and the contextual nuances of biomedical information makes it particularly valuable for

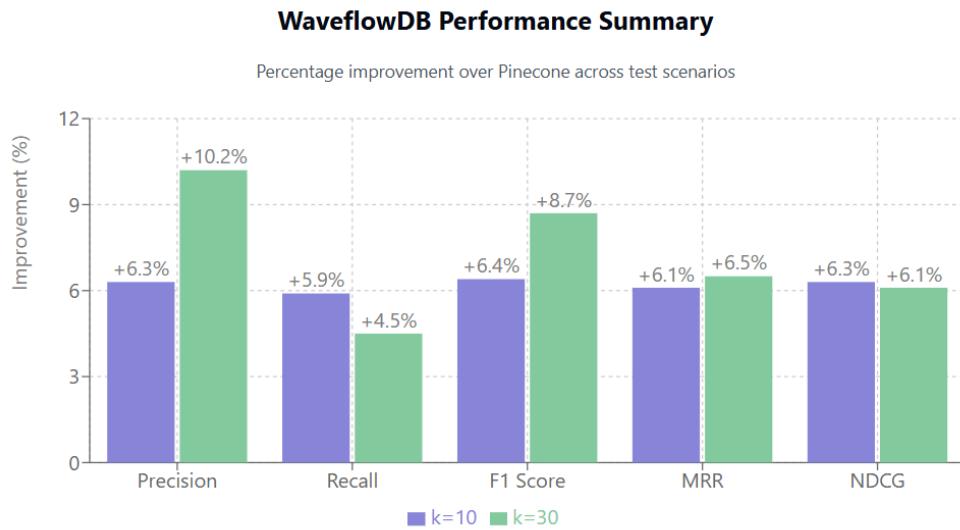


Figure 5: Performance Summary Chart showing key metrics across both test scenarios

healthcare organizations, pharmaceutical companies, research institutions, and medical education providers.

These performance advantages translate directly to business benefits including improved clinical decision quality, accelerated research and development, enhanced medical education, healthcare cost reduction, and streamlined regulatory compliance. Organizations seeking to leverage AI for biomedical applications should consider WaveflowDB as their vector database solution of choice for optimal retrieval performance.

This report is based on empirical evaluation conducted in May 2025 using the RAG-MINI-BIOASQ dataset containing 40,200 biomedical passages and 4,720 queries. For additional information about WaveflowDB capabilities, implementation details, or customized benchmarking for your specific biomedical use case, please contact the WaveflowDB team.