Slide 1:

Hook:  
Imagine a vast ocean of scientific data, but it's all in a language you don't understand and there's no one to help you navigate it. This is the reality for thousands of researchers, students, and policymakers struggling to access over 6 million oceanographic profiles collected by autonomous ARGO floats since 1999. This information is trapped in complex, technical files like NetCDF, making it incredibly difficult for anyone who isn't a data scientist to use. This knowledge gap hinders research, delays policy decisions, and prevents the public from understanding the health of our oceans.

Let me go through some facts and numbers:

* ARGO program has collected 6+ million oceanographic profiles.
* Only 15-20% of data is actively analyzed due to high technical barriers.
* Specialized formats and access complexity create a 90%+ dropout rate among non-expert users.

Slide 2:

That's why our team, DataWaves, is proud to present FloatChat. FloatChat is an AI-powered chatbot that acts as your personal ocean data expert. It's about democratizing access to this crucial scientific information by making it as easy as having a conversation.

Here’s how it works:

\* You ask a question in plain English, like "Show me salinity near India".

\* FloatChat translates your question into a powerful data query.

\* It fetches the exact data you need from a simplified database.

\* It then generates easy-to-understand visualizations, such as maps, graphs, and charts.

Slide 3:

## Demonstrate Technical Innovation

our RAG-based approach aligns with cutting-edge applications in scientific domains. **Emphasize the technical differentiation**:

* RAG pipeline reduces AI hallucinations by 75% through factual grounding
* Multi-modal LLM integration enables complex query understanding
* Real-time data visualization generation from natural language inputs

FloatChat’s technical foundation blends advanced AI, structured data management, and interactive visualization to create an intuitive conversational interface for ocean data exploration.  
We begin by ingesting raw Argo NetCDF files and converting them into structured formats such as PostgreSQL and Parquet for efficient querying. This data is augmented with metadata embeddings stored in vector databases like FAISS or Chroma, enabling rapid retrieval based on semantic similarity.  
Our AI backend uses Retrieval-Augmented Generation with large language models such as LLaMA or GPT, which are fine-tuned with oceanographic domain data to translate natural language questions into precise SQL queries via the Model Context Protocol.  
On the front end, we provide an interactive dashboard built with Streamlit and visualization libraries like Plotly and Leaflet, allowing users to explore mapped Argo float trajectories, depth-time profiles, and comparative trend graphs.  
This end-to-end pipeline ensures scalability, responsiveness, and usability for both expert and non-expert users  
  
Slide 4:  
FloatChat is eminently feasible thanks to its reliance on mature, open-source technologies and proven scientific datasets.  
We leverage well-documented data formats like NetCDF and existing vector search and LLM frameworks to accelerate development and minimize risk. The Indian Ocean Argo dataset provides a focused yet rich environment for our Proof-of-Concept, ensuring manageable scale and clear demonstration of capabilities.  
Potential challenges include the large data volume and the need for fine-tuned AI accuracy when interpreting complex queries. We address these through scalable incremental data processing, robust prompt engineering, and a chat interface designed to guide users through ambiguous inputs with clarifying questions.  
Our architecture allows incremental expansion to additional float types, biogeochemical sensors, and satellite observations, making FloatChat future-proof and adaptable

Slide 5:

The impact of FloatChat is significant and multi-faceted.

* **For Researchers & Students:** It will unlock research and learning by saving them hours of complex programming, allowing them to quickly test hypotheses and find specific data and reduces 70% time in data discovery
* **For Policymakers:** It will inform policy and strategy by making it easy to visualize crucial environmental trends like rising ocean temperatures. This enables data-driven decisions.
* **For the Public:** It engages public understanding by making ocean science accessible and understandable, raising awareness about the health of our oceans.
* **And it is also cost saving as** $50K+ will saved per research project on data analysis

Our solution is highly feasible. The key technologies like PostgreSQL, vector databases, LLMs, and web frameworks are mature, well-documented, and readily available

slide 6: ( Explaining references an giving closing line)

Imagine a world where the vast, invaluable ocean data powering climate understanding and marine stewardship is no longer locked away in technical silos, but accessible to every curious mind—scientists, policymakers, students, and citizens alike. This is the world FloatChat brings to life.  
By harnessing cutting-edge AI, sophisticated data engineering, and intuitive visual storytelling, FloatChat transforms one of humanity’s most complex datasets into a fluent conversation. We do not just provide answers; we enable discovery, learning, and informed decisions at unprecedented speed and scale.  
The oceans cover over 70% of our planet — their health is the health of our future. With FloatChat, we step boldly towards democratising ocean knowledge, amplifying global climate action, advancing research, and nurturing the next generation of ocean advocates.  
Choose FloatChat not only as a tool but as a critical movement towards smarter, accessible, and impactful ocean science. Together, let’s chart a course to a sustainable future—one question, one insight, one conversation at a time.  
Thank you for envisioning this journey with us. We are DataWaves, and we invite you to join us in making waves that matter