BIG-BRAIN: A UNIFIED AI-POWERED PERSONAL ASSISTANT FOR INTELLIGENT KNOWLEDGE MANAGEMENT AND DIGITAL CONTENT ORGANIZATION

USER MANUAL

1. OVERVIEW

Big-Brain is a **multi-modal AI assistant** that empowers users to capture, organize, and retrieve knowledge from any webpage. The system uses a **Chrome Extension** to collect both **text and images**, processes the data into **semantic vector embeddings**, stores them in a **Milvus vector database**, and responds to **natural language queries** via an **LLM-integrated frontend** (with voice support).

2. SYSTEM REQUIREMENTS

2.2 Hardware

- **Processor:** Intel i5 or higher (3.0 GHz+)
- **RAM:** 16 GB
- **Storage:** 500 GB SSD
- **Network:** High-speed internet
- **Display:** 15.6" monitor or better

2.3 Software

- **OS:** Windows 10 / Ubuntu 20.04+
- **Frontend:** HTML/CSS/JS with Streamlit UI
- **Backend:** Python (Flask)
- **Database:** Milvus (vector database)
- **IDE:** VS Code or PyCharm
- **Embeddings:** HuggingFace Transformers (e.g., all-MiniLM-L6-v2)
- **Browser:** Chrome (for extension)

3. FOLDER STRUCTURE

pgsql

CopyEdit

big-brain/

```
│
├── chrome_extension/
│   ├── manifest.json
│   ├── background.js
│   ├── content.js
│   ├── popup.html / .css / .js
│
├── server/
│   ├── server.py
│   └── embedding_utils.py
│
├── frontend/
│   ├── index.html
│   ├── style.css
│   └── script.js
│
└── milvus/
    └── Collection initialization and schema setup
```

4. SETUP INSTRUCTIONS

4.1 Milvus Vector DB Setup

- Install Milvus: https://milvus.io/docs/install_standalone-docker.md
- Start Milvus on localhost:19530

4.2 Backend Server

Run the Flask backend with embedding integration:

pip install flask pymilvus langchain_huggingface sentence-transformers

python server.py

This creates a Milvus collection named text_embeddings, with support for:

- Text vectorization (384-dim)
- Metadata storage
- Cosine similarity indexing

4.3 Chrome Extension

1. Open chrome://extensions/
2. Enable **Developer Mode**
3. Click **Load Unpacked**, and select the chrome_extension folder
4. Use the extension to select and save webpage text or images

4.4 Frontend

Open frontend/index.html in a browser. It allows:

- Text-based or voice queries
- Result display (text, images, source links)
- Extension download

5. HOW THE SYSTEM WORKS

MODULES

5.1 Data Collection

- Users select webpage content (text or images) using the Chrome Extension.
- Extension sends selected content and metadata (URL) to Flask backend.

5.2 Preprocessing & Chunking

- Cleans raw content (removes headers/footers).
- Splits text into 1000–1500 token chunks for embedding.

5.3 Embedding & Vectorization

- Text: all-MiniLM-L6-v2 from HuggingFace
- Images: Encoded as base64 and optionally embedded
- All data is stored in Milvus for fast similarity search

5.4 Query Handling

- User types or speaks a question (e.g., “What is AI hallucination?”)
- Query is embedded

- Top-k relevant chunks are retrieved
- LLM generates response based on embeddings
- If no match found → web search fallback

5.5 Evaluation

- Retrieval Accuracy: F1 = 92%, Exact Match = 73%
- User Feedback Loop for system improvement
- Qualitative + Quantitative metrics

6. VOICE INTEGRATION

Speech-to-text is supported via Web Speech API (webkitSpeechRecognition) in script.js. On microphone click:

- Captures user voice
- Converts to query
- Submits query to backend

7. DEPLOYMENT INSTRUCTIONS

1. **Run Milvus locally** (Docker or bare-metal)
2. **Start server.py**
3. **Load the Chrome Extension**
4. **Open index.html** from frontend folder
5. Ask questions via text/voice, view results, and test retrievals

8. TROUBLESHOOTING

Issue	Cause	Solution
Milvus connection fails	Port or host not accessible	Ensure Docker container is running on localhost:19530
Embeddings not stored	Server not running	Start server.py
Images not retrieved	Format mismatch	Use supported formats (.jpg/.png) and check base64 encoding
No voice capture	Browser restriction	Use Chrome with microphone permissions enabled

9. EXTENDING THE SYSTEM

- Add PDF/YouTube processing with LangChain integrations
- Multi-language support (add langdetect + translate)
- Integration with mobile via responsive frontend
- Plug-in support for Notion, Google Docs, and Drive

FINAL OUTPUT

- **Downloadable Chrome Extension**
- **Smart voice-enabled assistant UI**
- **Real-time, memory-powered Q&A**
- **Storage and retrieval of personal knowledge**
- **Visual + audio response capabilities**