Magnol.AI Copilot: Multimodal LLMs for Conversational Insight Generation

Hui Zhang, Guangchen Ruan, Hui Xiao

Eli Lilly & Co., Indianapolis, Indiana, USA
(zhang_hui, ruan_guangchen, hui_xiao)@lilly.com

Abstract

We present **Magnol.AI Copilot**, an extension of the Magnol.AI digital biomarker platform that integrates multimodal large language models (LLMs) to transform **digital health technology (DHT) trial dashboards** into conversational systems. Copilot augments the platform with a multi-agent orchestration layer and vision-enabled LLMs that interpret visualizations, tabular summaries, and textual metadata. The system enables natural language queries and automatic generation of contextual insights, allowing researchers to interact with wearable data through dialogue rather than static inspection. A case study with an **actigraphy device** demonstrates Copilot's ability to identify nightly compliance gaps and provide contextual explanations, reducing cognitive load compared to manual dashboard review. This work presents a **novel integration** of IoMT infrastructure with multimodal LLMs, advancing digital biomarker research toward conversational and accessible DHT trial platforms.

Introduction

Digital biomarkers (dBMs) derived from wearable sensors such as actigraphy devices are increasingly central to digital health technology (DHT) trials. They enable continuous measurement of activity, sleep, and compliance, providing richer endpoints than traditional assessments (Zhang et al. 2022; Ji et al. 2023; Zhang et al. 2023c, 2024). Platforms such as **Magnol.AI** already support ingestion, quality assessment, and visualization of multimodal data streams for biomarker development.

However, interpretation of DHT trial dashboards remains challenging. Current systems present static plots and tables, requiring technical expertise and manual inspection to extract insights. This limits accessibility for clinical scientists, increases cognitive load, and slows decision-making.

We introduce **Magnol.AI Copilot**, which augments Magnol.AI dashboards with **multimodal large language models (LLMs)** and a **multi-agent orchestration framework** (Ruan, Zhang, and Xiao 2025). Copilot allows users to interact with dashboards through natural language, enabling tasks such as in-place chart explanation, cross-chart comparison, and insight generation. This work represents a **novel integration** of IoMT infrastructure with multimodal LLMs, bridging visualization and conversational AI to advance digital health technologies.

System Overview

Magnol.AI Copilot extends the Magnol.AI platform with three integrated layers:

- **Visualization Workspace.** Users interact with dashboards by selecting regions of interest, zooming, or requesting explanations of specific plots (Zhang et al. 2023a,b). Visualizations can be added to a workspace for side-by-side comparison and custom prompting. The workspace supports in-place chart explanation, multi-visualization comparison, and AI-assisted insight generation (Figure 2).
- **Multi-Agent Orchestration.** User queries are routed through an intent classifier and task guide, which determine the workflow and delegate to specialized agents (e.g., table retrieval, compliance analysis). A supervisor agent integrates results and maintains dialogue context. The orchestration architecture is shown in Figure 1.
- **Multimodal LLM Integration.** A vision-enabled LLM interprets dashboard figures, tables, and metadata, producing contextual explanations in natural language. For example: *"Compliance decreased after week 3, coinciding with reduced nightly wear-time."*

Together, these layers unify IoMT data pipelines, interactive dashboards, and multimodal conversational AI into a single platform for DHT trial biomarker research.

Implementation Details

Magnol.AI Copilot is implemented as a modular, cloud-native system that integrates IoMT data pipelines, dashboards, and conversational AI services:

- **Data Layer.** Continuous wearable data from actigraphy devices are streamed through AWS Kinesis, stored in S3 and Elasticsearch, and exposed via secure APIs for derived measures such as activity, sleep, and compliance.
- **Visualization Layer.** Dashboards are built with React and Plotly, supporting region-of-interest selection, zoom, and workspace management. These visual outputs are captured as multimodal inputs to the LLM.

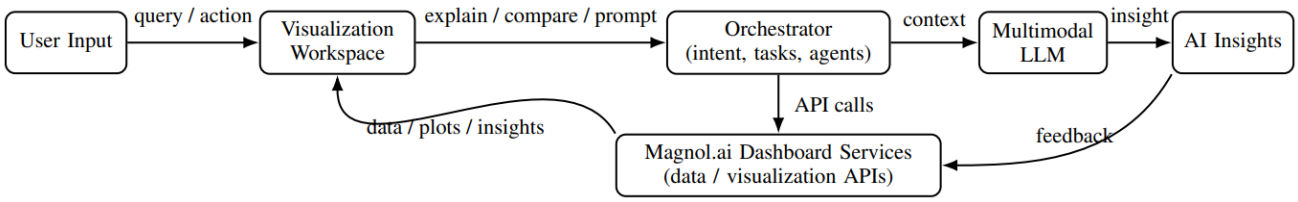

Figure 1: **Magnol.AI Copilot architecture (simplified).** User actions in the *Visualization Workspace* are routed to an *Orchestrator* (intent handling, task planning, agents), which queries *Platform Services* (data/visualization APIs) and packages context for a *Multimodal LLM*. The LLM returns *AI Insights* that can be rendered back in the workspace.

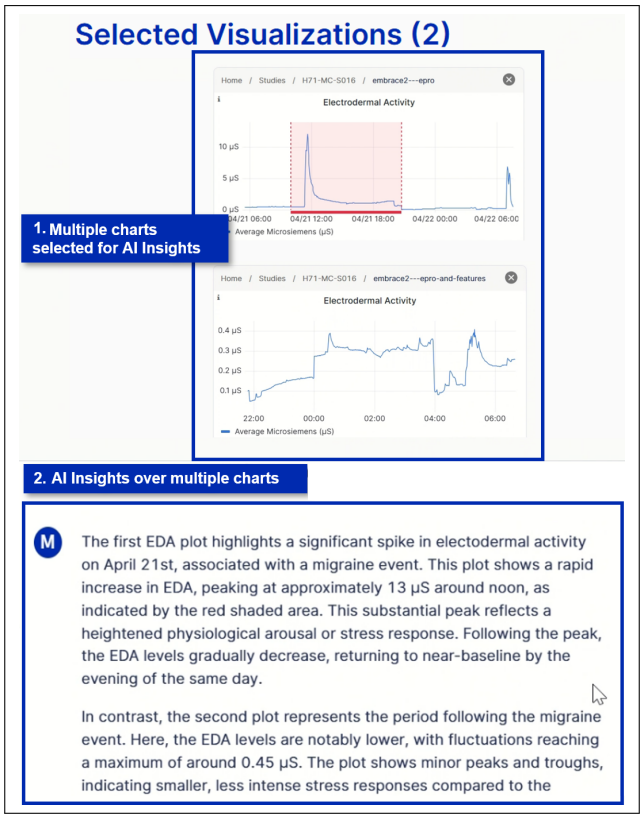

Figure 2: Magnol.AI Copilot enables interactive workspace exploration: users can select regions of interest directly from dashboard visualizations, then pin multiple charts into a workspace for side-by-side comparison. Copilot generates contextual insights from the selected visualizations, guided by custom prompts.

- **Conversational Layer.** A Python-based orchestration framework coordinates agent workflows. The intent classifier routes user queries, the task guide determines steps, and specialized agents handle retrieval and analytics. The supervisor agent integrates intermediate outputs and maintains dialogue context across turns.
- **Multimodal LLM.** A vision-capable LLM (e.g., GPT-4V or open-source equivalents) interprets figures, tables, and metadata. Prompting templates, retrieval grounding, and domain-specific instructions mitigate hallucinations and align outputs with biomarker semantics.
- **Deployment.** All services are containerized and deployed on cloud instances. The conversational interface is embedded into the Magnol.AI portal for secure web access. Typical latency for generating an insight is 3–5 seconds, and the system scales to datasets from hundreds of participants streaming multi-day wearable data.

Case Study Evaluation

We piloted **Magnol.AI Copilot** on a digital health technology (DHT) study using actigraphy and electrodermal activity (EDA) sensors. The outputs shown in Figure 2 are drawn directly from this dataset. In this example, a user selects portions of the continuously ingested EDA signal and compares how the signal behaves before and after an identified event of interest. Copilot supports this workflow by allowing the user to highlight regions of the signal, add them to the workspace, and request explanations or comparisons. The multimodal LLM then generates contextual insights, supporting the formulation of an initial hypothesis of how sensor signals may serve as indicators of the target event. **Magnol.AI Copilot reduced the effort required for engineers and scientists to interpret dashboards** compared to manual inspection. Typical query-to-insight latency was under five seconds, demonstrating feasibility for interactive trial monitoring and exploratory biomarker discovery.

Related Work and Conclusion

Digital health platforms such as Medidata Sensor Cloud, Koneksa, and Evidation support ingestion and visualization of wearable data but lack conversational or multimodal capabilities. Prior work on **Magnol.AI** focused on ingestion pipelines and data quality assessment for digital biomarker research. In parallel, multimodal large language models (LLMs) and retrieval-augmented generation have been applied in domains such as data analysis and explainability, but not in digital health technology (DHT) trial dashboards.

Magnol.AI Copilot is a novel integration of IoMT infrastructure, visualization systems, and multimodal LLMs. By enabling natural language interaction with dashboards, it reduces the effort required for engineers and scientists to interpret complex multimodal datasets.

References

Ji, J.; Venderley, J.; Zhang, H.; Lei, M.; Ruan, G.; Patel, N.; Chung, Y.-M.; Giesting, R.; and Miller, L. 2023. Assessing nocturnal scratch with actigraphy in atopic dermatitis patients. *NPJ Digital Medicine*, 6(1): 72.

Ruan, G.; Zhang, H.; and Xiao, H. 2025. Magnol.AI Copilot: Multimodal LLMs for Natural, Accessible Data Interaction. In *2025 IEEE/CVF International Conference on Computer Vision Workshops (To appear)*.

Zhang, H.; Giesting, R.; Miller, L.; Ruan, G.; Patel, N.; Ji, J.; Zhang, T.; and Yang, Y. L. 2023a. A Framework for Digital Data Quality Assessment in Digital Biomarker Research. In *The Ninth International Conference on Big Data, Small Data, Linked Data and Open Data (ALLDATA 2023)*.

Zhang, H.; Giesting, R.; Ruan, G.; Miller, L.; Patel, N.; Battioui, C.; Ji, J.; Zhong, M.; Kaczorek, A. D.; Zhang, T.; et al. 2023b. Analyzing and Reporting Wearable Sensor Data Quality in Digital Biomarker Research. 72–86.

Zhang, H.; Ji, J.; Ruan, G.; Patel, N.; Giesting, R.; Miller, L.; Yang, Y. L.; and Yang, J. 2022. Digital Data Platform for Connected Clinical Trials. In *2022 IEEE International Conference on Big Data (Big Data)*, 2446–2453.

Zhang, H.; Ruan, G.; Giesting, R.; Miller, L.; Patel, N.; Ji, J.; Yang, Y. L.; and Yang, J. 2023c. Engineering large wearable sensor data towards digital measures. In *2023 IEEE 8th International Conference on Big Data Analytics*, 14–23. IEEE.

Zhang, H.; Ruan, G.; Hartich, R.; Patel, N.; Zhong, M.; Kaczorek, A.; Miller, L.; Giesting, R.; Li, W.; Zhang, B.; et al. 2024. Magnol.Ai—an Internet of Medical Things (IoMT) Platform for Digital Health Research. In *2024 IEEE International Conference on Big Data (BigData)*, 5103–5112. IEEE.