

How Qualcomm Technologies' Customer Engineering team leverages Contextual AI for resolving complex engineering support challenges

Thousands of engineers globally are more efficient and productive thanks to accurate answers from the Contextual AI Platform

About Qualcomm

Headquarters
San Diego, CA

Industry
Semiconductors

Employees
~49,000

Global Footprint
150+ offices in 30+ countries

Tens of thousands

of annual support cases covered

Thousands

of customer engineers onboarded

Millions

of pages of multimodal content ingested

Executive summary

Qualcomm, a global leader in the semiconductor industry, is renowned for its pioneering work at the intersection of mobile connectivity and artificial intelligence. With a bold vision to create a more connected and intelligent world, Qualcomm Technologies' relentless pursuit of AI-driven solutions is reshaping today's technology and laying the foundation for a smarter, more integrated future. Qualcomm Technologies Vice President of Engineering, Yogi Chiniga, and his team sought a next-generation AI solution to enhance and scale their highly specialized engineering workforce. After evaluating several RAG solutions and finding them to be inadequate, the company selected Contextual AI to help transition beyond the demo phase towards full-scale production. As of November 2024, the resulting solution has been smoothly incorporated into the daily workflows of thousands of engineers across various technology areas, resulting in increased efficiency and enhanced productivity.

Qualcomm branded products are products of Qualcomm Technologies, Inc. and/or its subsidiaries.

“Much of the work with our customers involves engineers sifting through complex, scattered, and dense internal documentation. It is critical that we help our customers resolve their issues quickly, but even straightforward questions can take hours to research. We realized that solving this challenge required more than just a basic AI assistant—it required a partner capable of understanding the technical depth and specialization of our work, and one that could streamline the process without compromising on quality.”



Yogi Chiniga

Vice President of
Engineering, Qualcomm
Technologies

The Production Challenge

The Customer Engineering organization at Qualcomm Technologies plays a vital role in supporting thousands of customers, including major companies with all aspects of product development—from design and development to debugging and integration. The team tackles a broad array of complex technical challenges, requiring deep expertise across multiple Qualcomm® technology areas. However, like many teams with highly specialized technical talent, Customer Engineering faced increasing bandwidth constraints. To address the growing demands on the team, Qualcomm Technologies leadership sought an AI solution that would help maximize the team’s productivity and deliver an excellent customer experience. They envisioned an application capable of efficiently querying and synthesizing the company’s vast technical documentation, both internal and external-facing, ultimately reducing the time needed to resolve certain customer issues.

The project’s high accuracy requirements and low tolerance for failure made it exceptionally challenging for standard RAG (retrieval-augmented generation) solutions to deliver the accuracy needed to be considered production-ready:

- 1. Massive volumes of noisy data:** For most RAG solutions, the ability to retrieve the right information degrades as the volume of data increases, especially in real-world deployments where data quality can be inconsistent. This made Qualcomm Technologies’ use case, which required handling millions of pages of highly technical documents, with thousands of new pages added each day, particularly difficult to implement.

2. Complex, multimodal content: The company's documentation included complex diagrams, technical images, nested tables, and code, spread across a diverse set of data formats (PDF, HTML, Excel, and more), complicating extraction and accurate retrieval.

Qualcomm Linux Bluetooth Guide: Overview

The following figure shows the various components of the Qualcomm connectivity chipset, including the Bluetooth subsystem and its interaction with the application processor:

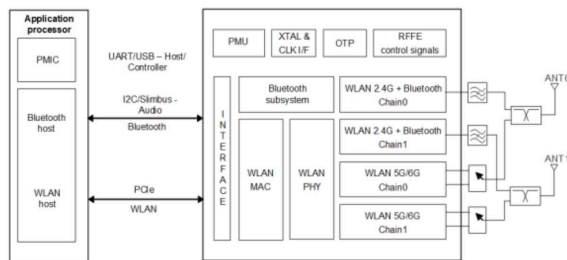


Figure 1-1 Qualcomm connectivity chipset block diagram

- NOTE**
- The Qualcomm Linux platform allows you to develop applications for QCS6490 and QCS5430.
 - The software support for the WCN6750 chipset will be provided in future releases.

Qualcomm Linux Bluetooth Guide: Verify

Run Bluetooth inquiry scan

To initiate an inquiry for nearby devices, run the following command:

```
inquiry
```

The sample output of this command appears as follows:

```
inquiry
Inquiry Started
Device Found details:
Found device Addr: 70:09:71:2c:f7:6a
Found device Name: [TV] 224 Breakroom-1
Device Type is: 1
Device Found details:
Found device Addr: a0:d0:5b:97:7e:f4
Found device Name:
Device Type is: 1
Device Found details:
Found device Addr: 20:37:a5:f2:25:78
Found device Name: iPhone 13
Device Type is: 1
Inquiry Stopped automatically
```

Cancel Bluetooth inquiry scan

To cancel an inquiry that is in progress, run the following command:

```
cancel_inquiry
```

Qualcomm documentation includes complex figures, specialized concepts, and technical terminology that make it difficult to extract and accurately retrieve the right information

3. Diversity of domains: The solution needed to span a large number of fundamentally distinct technology areas (e.g., Multimedia, Modem, Hardware, RF), each with unique terminology, documentation conventions, and internal acronyms.

"Other RAG platforms would have required us to piece together multiple disparate components, which ultimately wouldn't have scaled or delivered the accuracy our use cases demanded. With Contextual AI, we not only achieved exceptional accuracy at scale, but also found a reliable, all-in-one partner that streamlined our AI initiatives to empower our customer engineers with cutting edge technology."

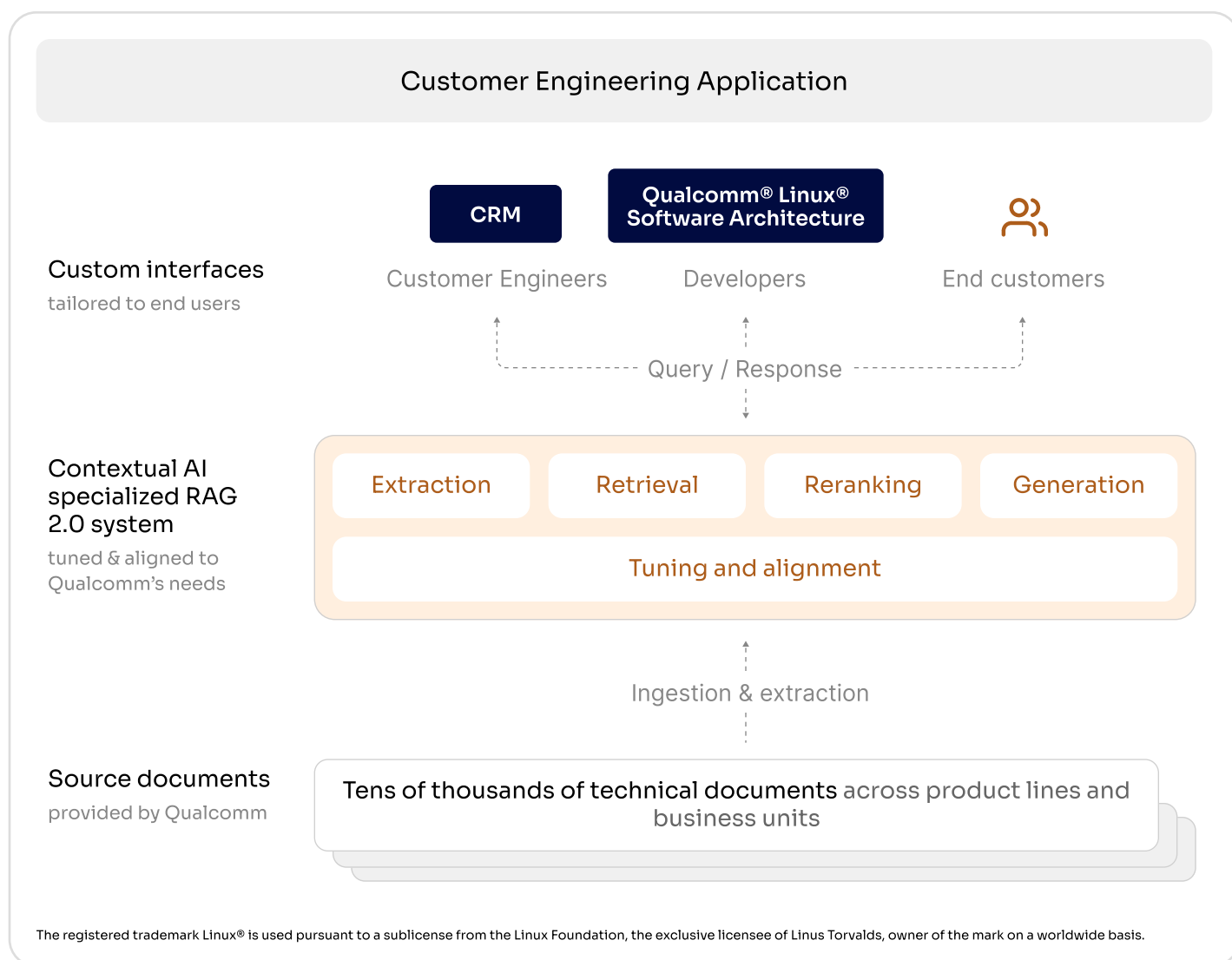


Anuja Thakur

Director of Engineering
Operations, Qualcomm
Technologies

Achieving the “production bar” with Contextual AI

To achieve the level of accuracy required for a production deployment, Qualcomm Technologies collaborated with Contextual AI, the leader in production-grade RAG. Unlike most alternative solutions, which struggled to deliver the required precision, Contextual AI's platform—fueled by their breakthrough RAG 2.0 approach—optimized the RAG pipeline as a unified system and helped specialize the system for the Customer Engineering use case. The solution enabled Qualcomm Technologies to answer highly specialized questions with production-level accuracy, and paved a clear path to a global production deployment across thousands of customer engineers.



Contextual AI's platform enabled Qualcomm Technologies to deploy highly accurate and specialized RAG applications to its customer engineers

Key capabilities	How it was put into practice
State-of-the-art retrieval and reranking	Contextual AI's mixture-of-retrievers approach and state-of-the-art reranking helped contextualize the grounded language model with the most relevant technical knowledge across complex data formats
Grounded generation	The language model was strongly grounded in Qualcomm Technologies' data to ensure it generated answers with greater accuracy, fewer hallucinations, and increased traceability
Multimodal extraction	Multimodal extraction, layout segmentation, and dynamic chunking were used to maximize contextual understanding across a large and diverse set of intricate technical documentation
Continuous ingestion	Automated and ongoing ingestion of the latest Qualcomm Technologies documentation, ensuring new knowledge was available to the application within 24 hours
Fine-tuning and alignment	Specialized and jointly optimized model components, as a part of the system, further sharpen application accuracy, identifying key signals from in-product feedback
Safety and guardrails	Alignment to meet Qualcomm standards across core safety, brand safety, sensitive topics, without tradeoffs for accuracy
End-to-end enterprise security and deployment	SSO and RBAC support, query-time entitlements, secure single-tenant deployment on Contextual AI's infrastructure, and expert support to move quickly and safely to production
Hands-on partnership and deep AI expertise	Deep engineering and product partnership between Qualcomm Technologies and Contextual AI teams to implement a specialized RAG solution for the Customer Engineering team

The screenshot shows the Contextual AI interface with a query: "Give me more information about how to debug issues related to the Qualcomm RB3 Gen 2 Vision Development Kit". The interface displays "Retrieved 8 documents" and a detailed answer titled "Debugging Issues on the Qualcomm RB3 Gen 2 Vision Development Kit". The answer includes steps for enabling SSH access and collecting Bluetooth logs. A callout box highlights a "Mainboard and interposer block diagram" from the Qualcomm RB3 Gen 2 Development Kit Guide.

Table 2-1 Kit components

Core kit	Vision kit
- Qualcomm RB3 Gen 2 Development Kit - Interposer with SAM - QD5490 - QD5490 - Power supply (12 V) - USB A to USB-C cable - Test equipment - Pin tool to help access QFP sockets	- Qualcomm RB3 Gen 2 Vision kit - Qualcomm Vision mainboard board 4000T camera - QD5490 tracking camera

Mainboard and interposer block diagram

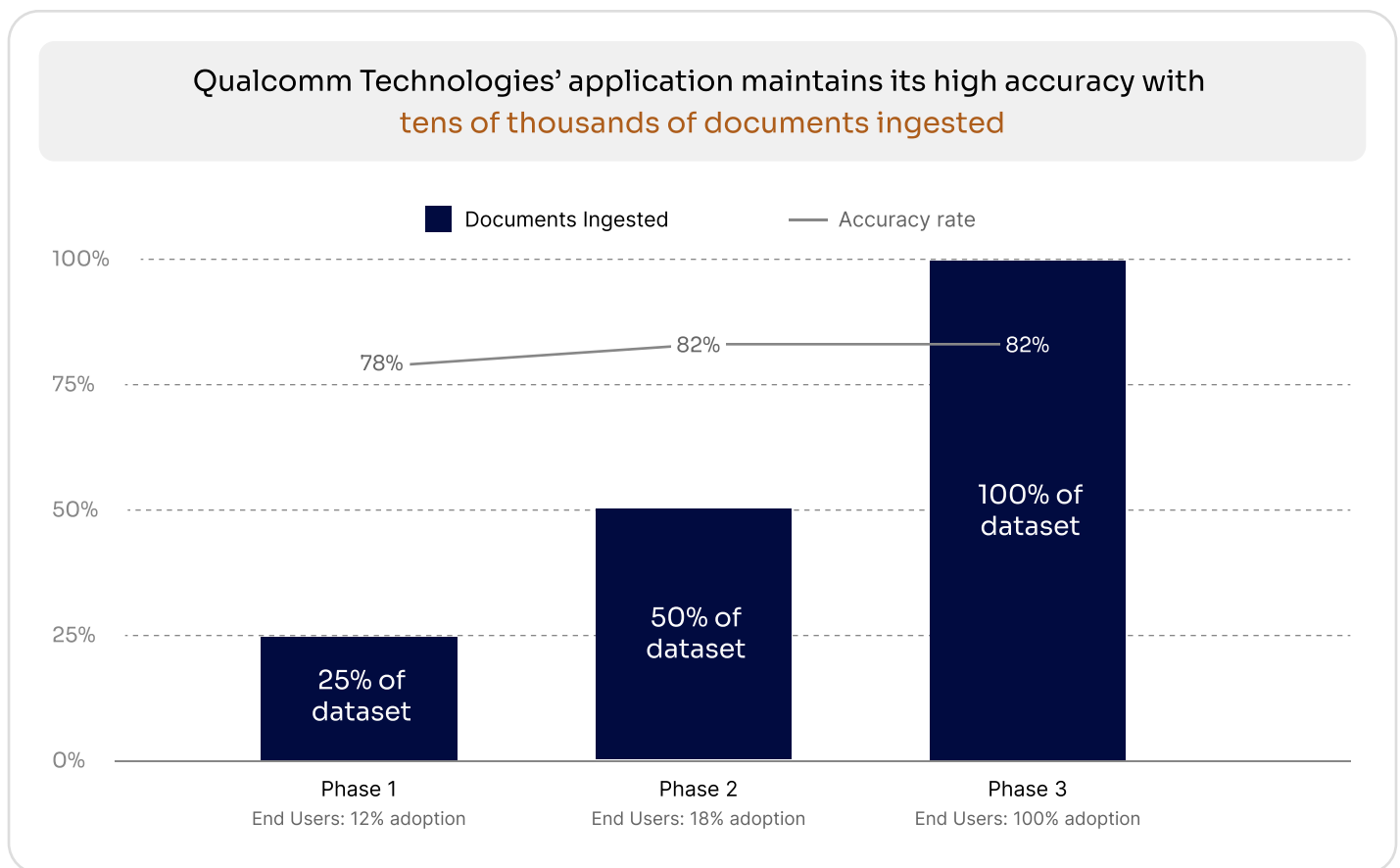
The following block diagram shows the RB3 Gen 2 interposer and interposer connections are shown in each view. Some connections are on the interposer board and some are on the mainboard. Reviewing the placement of these connections is useful in configuring and debugging them easily and effectively.

Contextual AI's specialized applications achieved high accuracy across a wide variety of deeply technical questions and documents, while providing fine-grained attributions in the source documentation

Delivering “ROI of AI” for Qualcomm Technologies’ Customer Engineering

Many companies struggle with moving beyond the “demo stage” of AI to a real-world, high-value production setting, especially if the underlying data sources are vast and complex, as was the case with Qualcomm Technologies. Thanks to Contextual AI, Qualcomm Technologies was able to achieve:

- 1. High accuracy at scale:** The solution exceeded production-grade accuracy thresholds across all technology areas. With the RAG components jointly optimized as a single system and further specialized using the platform’s fine-tuning and alignment engine, Qualcomm Technologies maintained production-level accuracy even as the number of users and the volume of underlying data increased dramatically.



- 2. Rapid deployment to production:** Developing high-quality RAG applications is a complex undertaking, and standard RAG techniques were not up to par for the complexity and criticality of this use case. Thanks to hands-on collaboration with Contextual AI and its cutting edge technology, Qualcomm Technologies was able to move to production quickly, enabling customer engineers to increase their efficiency and productivity.

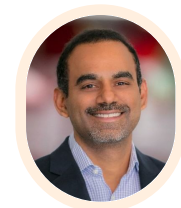
3. **Broad internal adoption:** Qualcomm Technologies embedded its application directly into existing systems for handling customer requests, expediting user adoption and seamlessly integrating with the existing workflows for customer engineers. To date, thousands of end users have been onboarded to the application.

What's next for Qualcomm Technologies and Contextual AI

While Qualcomm Technologies' first Contextual AI application was designed for internal usage, the team plans to provide a similar application directly to customers soon. By enabling customers to use the Contextual AI solution to directly interact with Qualcomm's rich and complex documentation, the Customer Engineering organization expects to significantly reduce the volume of documentation-related support. Qualcomm Technologies and Contextual AI are also actively exploring agentic RAG workflows, formulating initial search queries automatically across diverse data sources.

Beyond Customer Engineering, Qualcomm Technologies is considering using the Contextual AI Platform for a variety of additional complex and high-value use cases that require clearing challenging accuracy thresholds with the type of specialization that only a RAG 2.0 approach can provide.

"Contextual AI has been an invaluable partner in our AI journey. Their expertise is not only accelerating our progress, but also driving meaningful impact across our internal teams and significantly enhancing our customers' experience. We see immense potential in expanding our collaboration beyond Customer Engineering. While our initial use case has shown great promise, there are numerous other business units that stand to benefit from the comprehensive, end-to-end AI platform that Contextual AI has built."



Yogi Chiniga

Vice President of
Engineering, Qualcomm
Technologies