GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

CHIP DESIGN GERMANY

KI4BoardNet

edaWorkshop25 FORUM 2025

12. 05. - 13. 05. 2025 Messe Dresden

Chip2System
LLM-gestütztes Co-Design

Julian Withöft
julian.withoeft@tu-dortmund.de
Werner John
Philip Vargehese Modayil
Emre Ecik
Nima Ghafarian Shoaee
Bekhzodkhon Khamdamov
Domenik Helms
Ralf Brüning
Jürgen Götze

technische universität dortmund

DT/IPL Arbeitsgebiet Datentechnik Information Processing Lab

DLR

PYRAMIDE 2525 ElektronikEntwicklung KI-Anwendungen

ZUKEN®

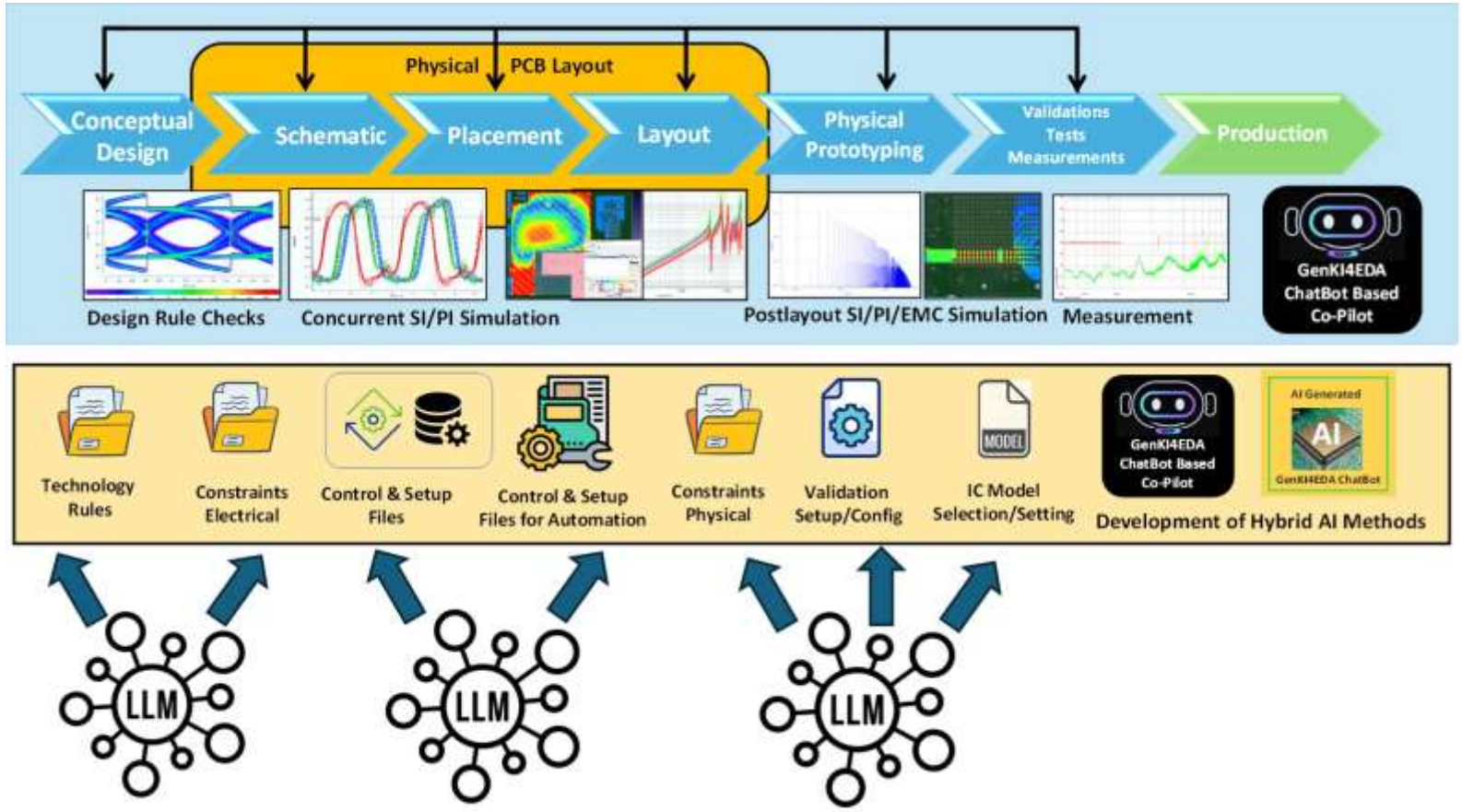
Das diesem Beitrag zugrundeliegende Vorhaben KI4BoardNet wird vom Bundesministerium für Bildung und Forschung (BMBF) (Förderprogramm Mannheim) unter dem Kennzeichen (Förderkennzeichen: 16ME0779/TUDO - 16ME0777/IDMT) gefördert. Für den Inhalt dieser Publikation sind die Autoren verantwortlich.

Gliederung

- Motivation
- Rasa Intent-based Bot
- LLM als Fallback
- Fine-Tuning eines Domänen-spezfischen LLMs
- LLM Agents
- Anwendungsbeispiele
- Zusammenfassung und Ausblick

Co-Pilot Motivation für Chip2System Entwurfsaufgaben

- Chatbot-gestützte **Design-Assistenz (Co-Piloting) in jedem Entwurfschritt** mittels Nutzung LLM-generierter Control Sets, welche Regeln/Constraints für individuelle Design Sets enthalten
- ➢ Chatbot liefert **Design-Empfehlungen** und auch **Input** für **Entwicklungsdaten** basierend auf **Design-Kontext**



Co-Pilot Motivation für Chip2System Entwurfsaufgaben

- **Sicherstellung** eines umfassenden **IP-Schutzes**
- Ziel ist Entwicklung **hybrider KI-Methoden**: **Wissen** (**Regeln**) mit **Lerndaten** (**z.B. LLM**) kombinieren → **Unterbindung** von **Halluzinationen**.
- Implementation **multimodaler LLMs**
  - Training mit verschiedenen Datenquellen (Text; Bild/Video; Sprache).
  - Analyse von Schaltungsentwürfen direkt aus Bild- oder Simulationsdaten.
- **Chatbot-Wissensbasis:** Strukturiertes Systementwurfswissen (Vektor Datenbank + MetaDaten)
  - Chatbot Dokumentenanalyse/Verarbeitung für die **EDA-Wissensdomänen IC + PCB (EMC/SI/PI) + Bordnetz.**
  - Chatbot-Infrastruktur (EDA-Wissensdomänen IC + PCB (EMC/SI/PI) + Bordnetz + Konzeptions- und Entwicklungsprozesse).
- Integration von **Simulations-** und **AI-APIs (Agenten)**
  - EDA-Werkzeuge (Simulatoren + Analyse).
  - KI-Modelle.
  - LLM-Modelle.

External Repository
KI-Model basierend auf generischen Daten

Company Specific Repository
KI-Model basierend auf anwender-spezifischen Daten

Rasa Intent-Based Bot

- Fokus auf **sehr kleine** Wissensdomäne: Decoupling Capacitors
- Natural Language Understanding (NLU)
- **Antwort** kommt direkt **aus** einer **Datenbank**

Einfach, Strukturiert, Effizient, Kontrolliert

Wenig Flexibel, Nicht Skalierbar, Limitiert auf Datenbank

Workflow of Intent Based Model

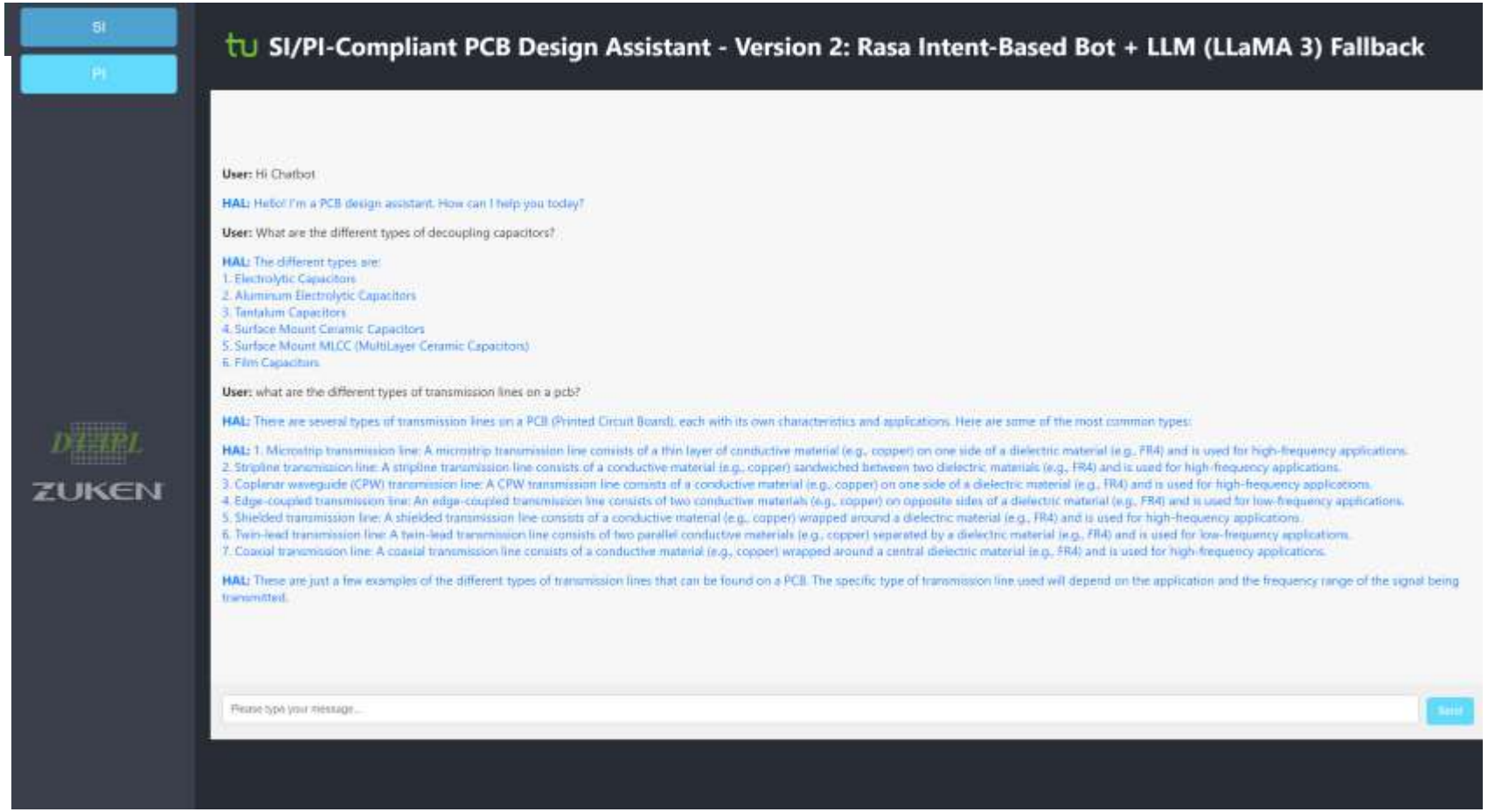
Quelle: https://rasa.com

LLM Fallback

- Wenn Rasa-Confidence < Threshold → LLM (Meta-Llama-3-8B) wird für die Response verwendet

Jede Frage kann beantwortet werden, Externe Datenquellen können via RAG eingebunden werden, LLM läuft Lokal auf dem Server

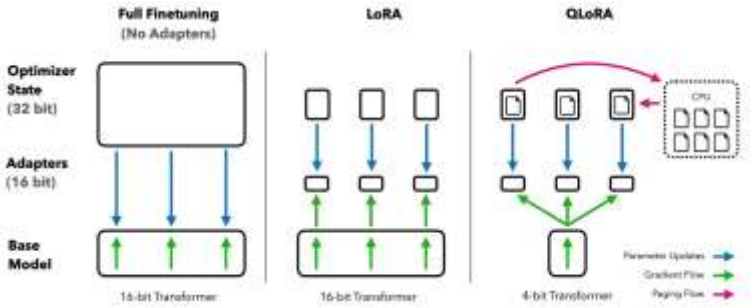
Komplexität, Halluzinationen, LLM ist kein PCB/IC-Experte



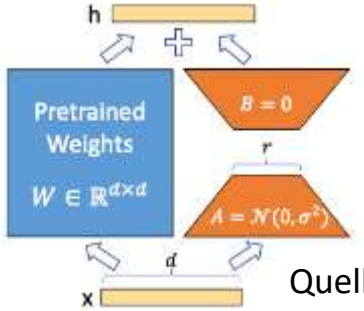
LLM Fine-Tuning

- **Supervised Fine-Tuning** → Spezialisierung des LLMs in einer Domäne
- **LLaMA-Factory** [1] vereinfacht den Fine-Tuning Prozess
- **QLoRA** wird verwendet (4-Bit Quantisierung) [2,3]

Quelle: https://github.com/hiyouga/LLaMA-Factory



Quelle: [3]



Quelle: [2]

- In [4] wurde gezeigt, dass bereits **mit 1000 Samples** eine **gute Performance** zu erreichen ist
- Unser Datensatz mit insgesamt **2247 Q&A manueller Samples** aus der **SI/PI/EMC Domäne** → **2197 Training Samples**, **50 Test Samples**

[1] Y. Zheng, R. Zhang, J. Zhang, Y. Ye, Z. Luo, Z. Feng, and Y. Ma, "LlamaFactory: Unified Efficient Fine-Tuning of 100+ Language Models," arXiv preprint arXiv:2403.13372, 2024.
[2] E.J.Hu, Y.Shen, P.Wallis, Z.Allen-Zhu, Y.Li, S.Wang, L.Wang, and W.Chen, "LoRA: Low-Rank Adaptation of Large Language Models," arXiv preprint arXiv:2106.09685, 2021.
[3] T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer, "QLoRA: Efficient Finetuning of Quantized LLMs," arXiv preprint arXiv:2305.14314, 2023.
[4] C. Zhou, P. Liu, P. Xu, S. Iyer, J. Sun, Y. Mao, X. Ma, A. Efrat, P. Yu, L. Yu, S. Zhang, G. Ghosh, M. Lewis, L. Zettlemoyer, and O. Levy, "LIMA: Less Is More for Alignment," arXiv preprint arXiv: 2305.11206, 2023.

Evaluation der Fine-Tuned LLMs

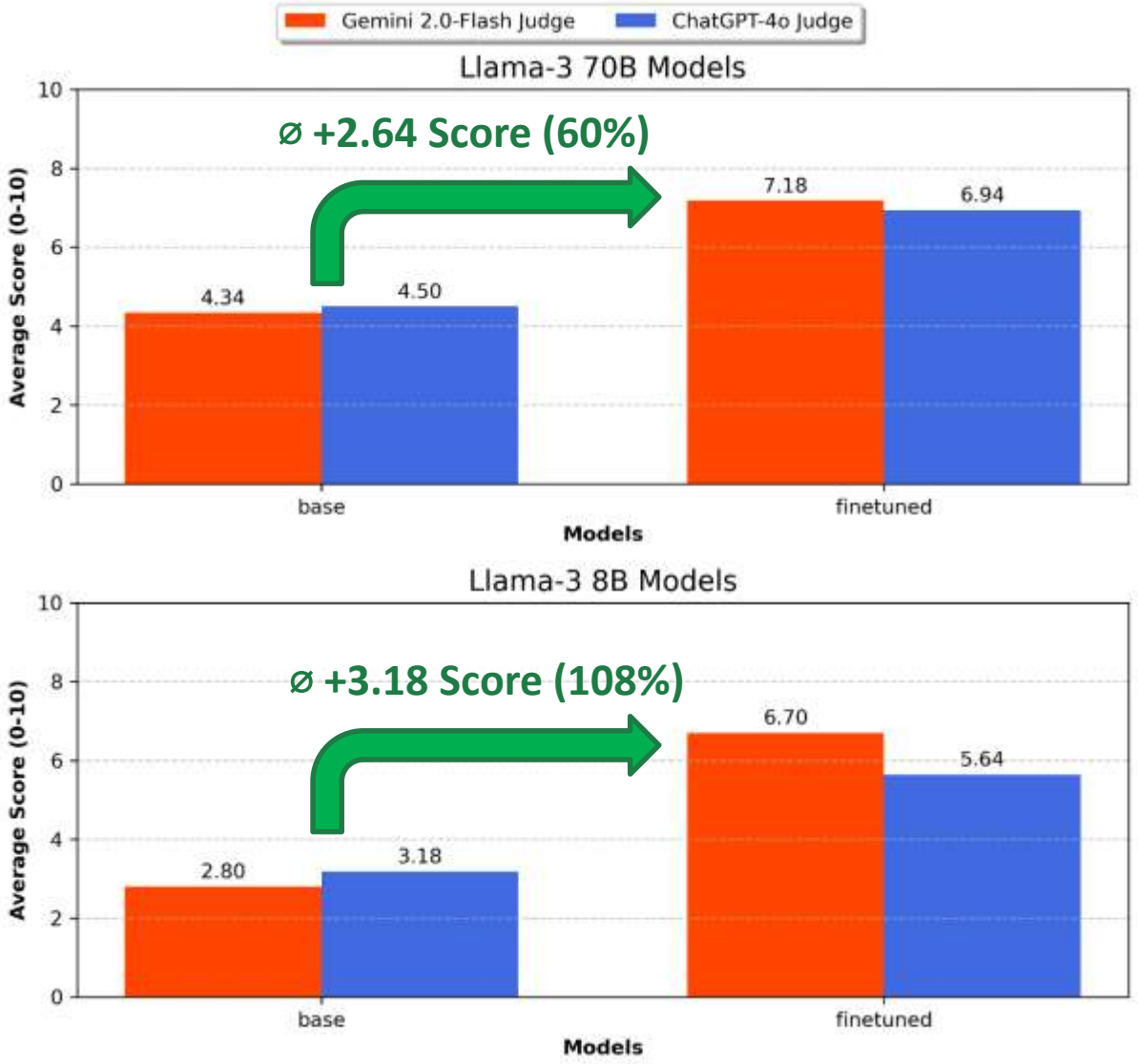
- **LLM-as-a-Judge** [5] Prinzip wird verwendet, um die Antworten der LLMs auf Basis des Testdatensatzes zu bewerten:

```
[Instruction]
Please act as an impartial judge and evaluate the quality of the
response provided by an AI assistant to the user question
displayed below. Your evaluation should consider correctness
and helpfulness. You will be given a reference answer and the
assistant's answer. Begin your evaluation by comparing the
assistant's answer with the reference answer. Identify and
correct any mistakes. Be as objective as possible. After providing
your explanation, you must rate the response on a scale of 1 to 10 by
strictly following format: "[[rating]]", for example: "Rating: [[5]]".

[Question]
{question}

[The Start of Reference Answer]
{ref_answer_1}
[The End of Reference Answer]

[The Start of Assistant's Answer]
{answer}
[The End of Assistant's Answer]
```



- **Fine-Tuning** führt eindeutig zu **besseren Antworten** der **LLMs**

[5] L. Zheng, W.-L. Chiang, Y. Sheng, S. Zhuang, Z. Wu, Y. Zhuang, Z. Lin, Z. Li, D. Li, E. P. Xing, H. Zhang, J. E. Gonzalez, and I. Stoica, "Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena," arXiv preprint arXiv:2306.05685, 2023.

LLM Agent Framework

Input Prompt

SFT LLM + RAG

Structured prompt with requirements and knowledge augmentation with RAG

CoT Orchestrator

Testing

RAG

Vision Agent

AI Modules

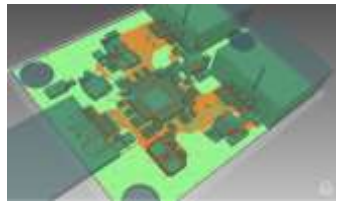
Simulation

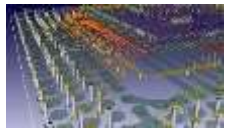
Packaging Expert LLM Agent

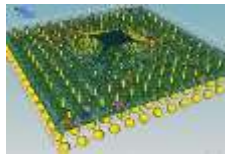
Requirement Verification

Packaging Design Proposal

Testing/ Verification





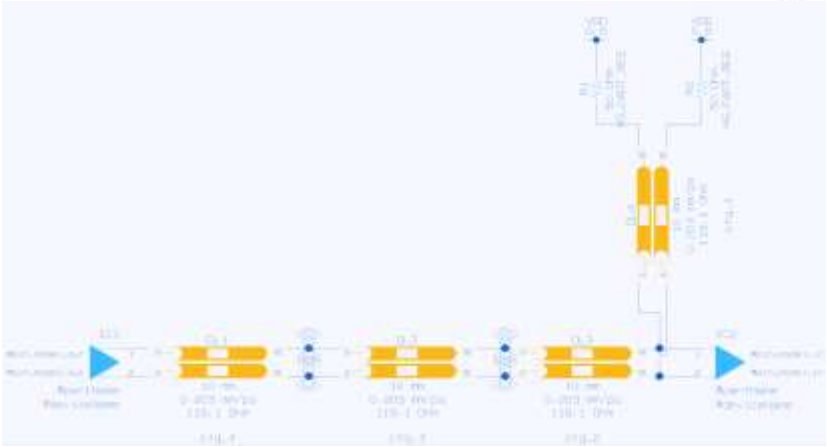


Hetero-Systemintegration

Quellen: https://flaticon.com
https://www.freepik.com

LLM Agent – Exemplarisches Anwendungsszenario

- Stackup-Recommendation for Differential Pair Point-to-Point (P2P DDR4-Interconnect)
- Agentic LLM Workflow (Exemplarisch):
  - User Input: Datasheet TI ; Screenshot Schematic + User Prompt: I am designing a PCB and I am planning to use the AM64x IC (see attached datasheet). I have to route the DQS DDR4 signals as a differential pair point-to-point interconnect (see screenshot). Can you recommend a layer stackup for my design and on which layer to route my differential DQS signals specifically? Also what is the maximum transmission line length that is allowed and where should the vias be placed? Finally, what is the max tolerance of the transmission line parameters (i.e. width, spacing, etc.)?
  - SFT LLM + RAG scans datasheet reformulates the requirements & RAG scans datasheet for additional information (i.e. Skew thresholds (here 0.4 ps), Eye mask, etc.)
  - CoT Orchestrator <Think> : I have a differential DDR4 point-to-point interconnect. I need to specify layer stackup and answer design-specific questions. First I need more information regarding DDR4 → Invoke Tool: RAG Agent
  - RAG Agent: I need further information on DDR4 so I will search for further information... Found JEDEC-DDR4 specification on local server → Scan and add to vector database.
  - <Think> : I need to evaluate the schematic that the user provided to check what is currently implemented → Invoke Vision Agent
  - MM-LLM Vision Agent: Analyzing Image.. Extracting Parameter Values.. Validating that it is a Differential Point-to-Point Topology.. Finished.. → Give back information to orchestrator
  - <Think> : I need to optimize the vias placement, transmission line lengths and tolerances regarding transmission line parameters.. Search for AI Modules for this design task.. Found AI module for specific interconnect.. → Invoke AI Module Agent
  - AI Module Agent: Searching for matching AI modules.. Found AI prediction module based on ANN model.. Running GA optimization.. Optimization finished..
  - <Think> : I need to verify the design proposal of the AI module with some simulations. → Invoke Simulation Agent
  - Simulation Agent: Generate *.sidf file.. Starting simulations via Zuken eCADSTAR.. Evaluating simulations.. Violations found – need to adjust vias and tolerances slightly.. Make adjustments to design proposal.. Finalize design proposal..
  - <Think> : I have found a design proposal. Do I need to check anything else? If yes, go back into loop, if no, give design proposal back to user



Application Note

AM64x\AM243x DDR Board Design and Layout Guidelines

Texas Instruments

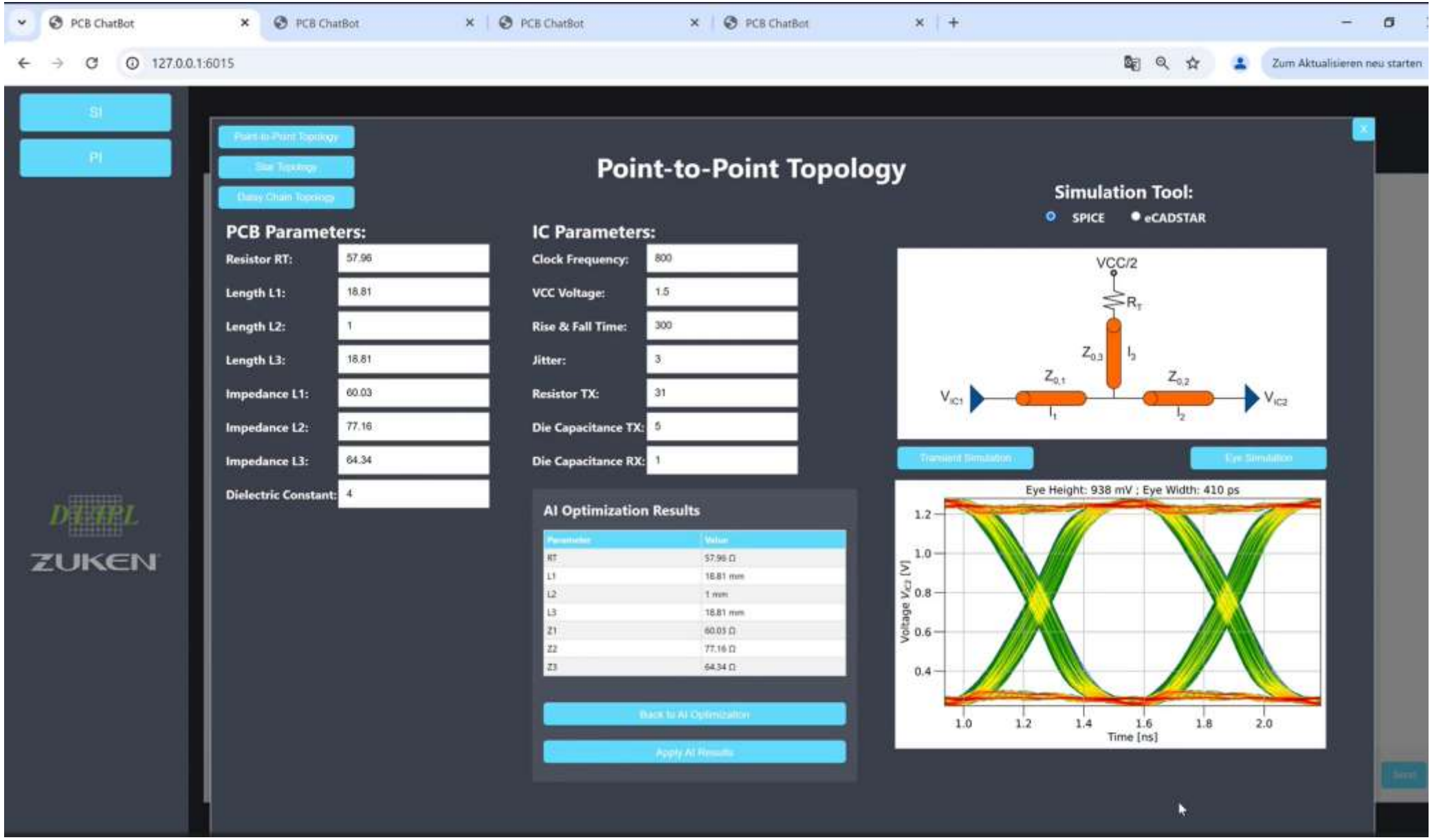
ABSTRACT

The goal of this document is to describe how to make the AM64x/AM243x DDR system implementation straightforward for all designers. The requirements have been distilled down to a set of layout and routing rules that allow designers to successfully implement a robust design for the topologies TI supports.

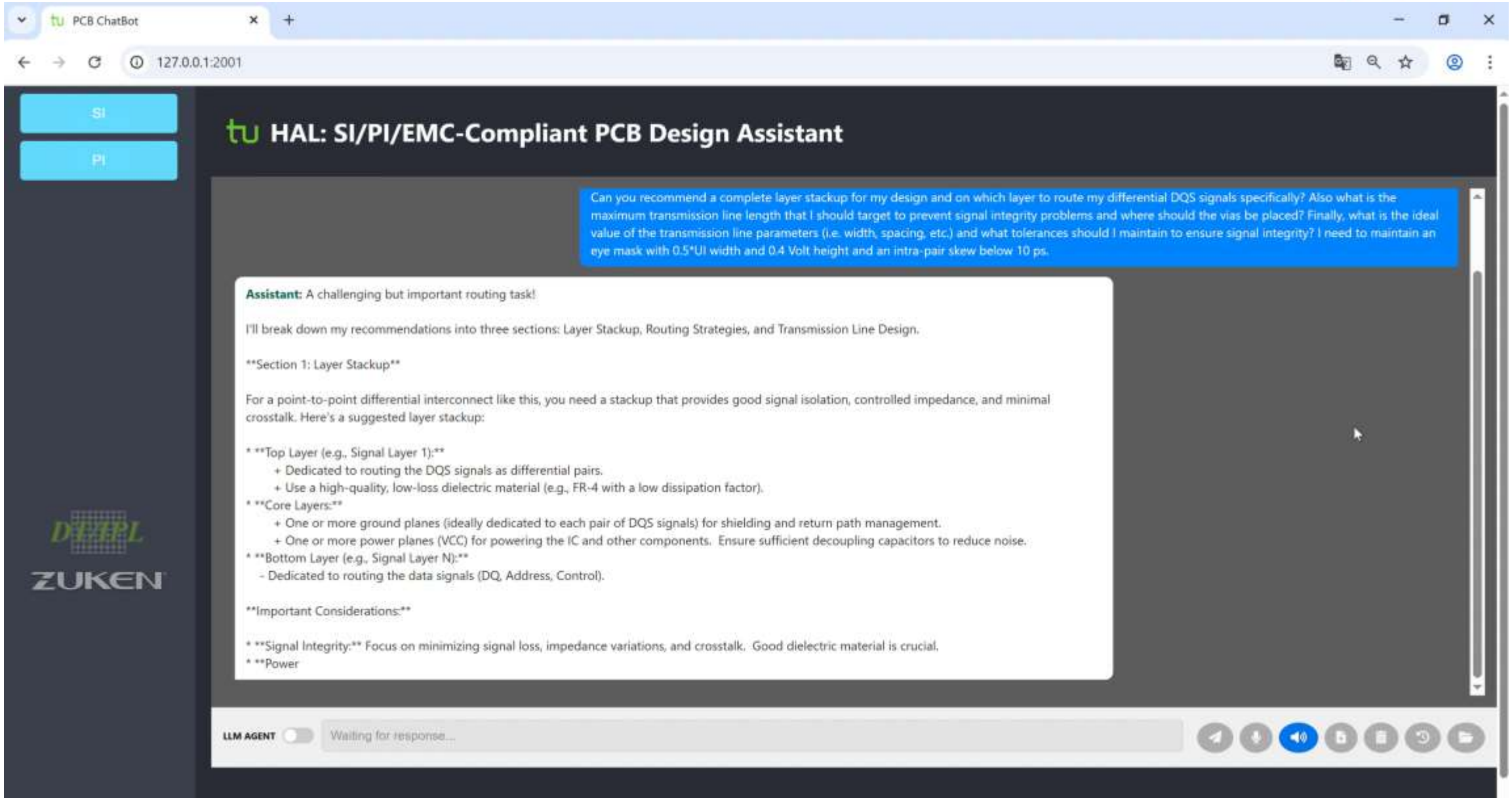
Table of Contents

1 Overview
1.1 Board Designs Supported
1.2 General Board Layout Guidelines
1.3 PCB Stack-Up
1.4 Bypass Capacitors
1.5 Velocity Compensation
2 DDR4 Board Design and Layout Guidance
2.1 DDR4 Introduction
2.2 DDR4 Device Implementations Supported
2.3 DDR4 Interface Schematics
2.4 Compatible JEDEC DDR4 Devices
2.5 Placement
2.6 DDR4 Keepout Region
2.7 VTT
2.8 Net Classes
2.9 DDR4 Signal Termination
2.10 VREF Routing
2.11 VTT
2.12 POD Interconnect
2.13 CK and ADDR_CTRL Topologies and Routing Guidance
2.14 Data Group Topologies and Routing Guidance
2.15 CK and ADDR_CTRL Routing Specification
2.16 Data Group Routing Specification
2.17 Bit Swapping
3 LPDDR4 Board Design and Layout Guidance
3.1 LPDDR4 Introduction
3.2 LPDDR4 Device Implementations Supported
3.3 LPDDR4 Interface Schematics
3.4 Compatible JEDEC LPDDR4 Devices
3.5 Placement
3.6 LPDDR4 Keepout Region
3.7 Net Classes
3.8 LPDDR4 Signal Termination
3.9 LPDDR4 VREF Routing
3.10 LPDDR4 VTT
3.11 CK and ADDR_CTRL Topologies
3.12 Data Group Topologies
3.13 CK and ADDR_CTRL Routing Specification
3.14 Data Group Routing Specification
3.15 Channel, Byte, and Bit Swapping
4 Revision History

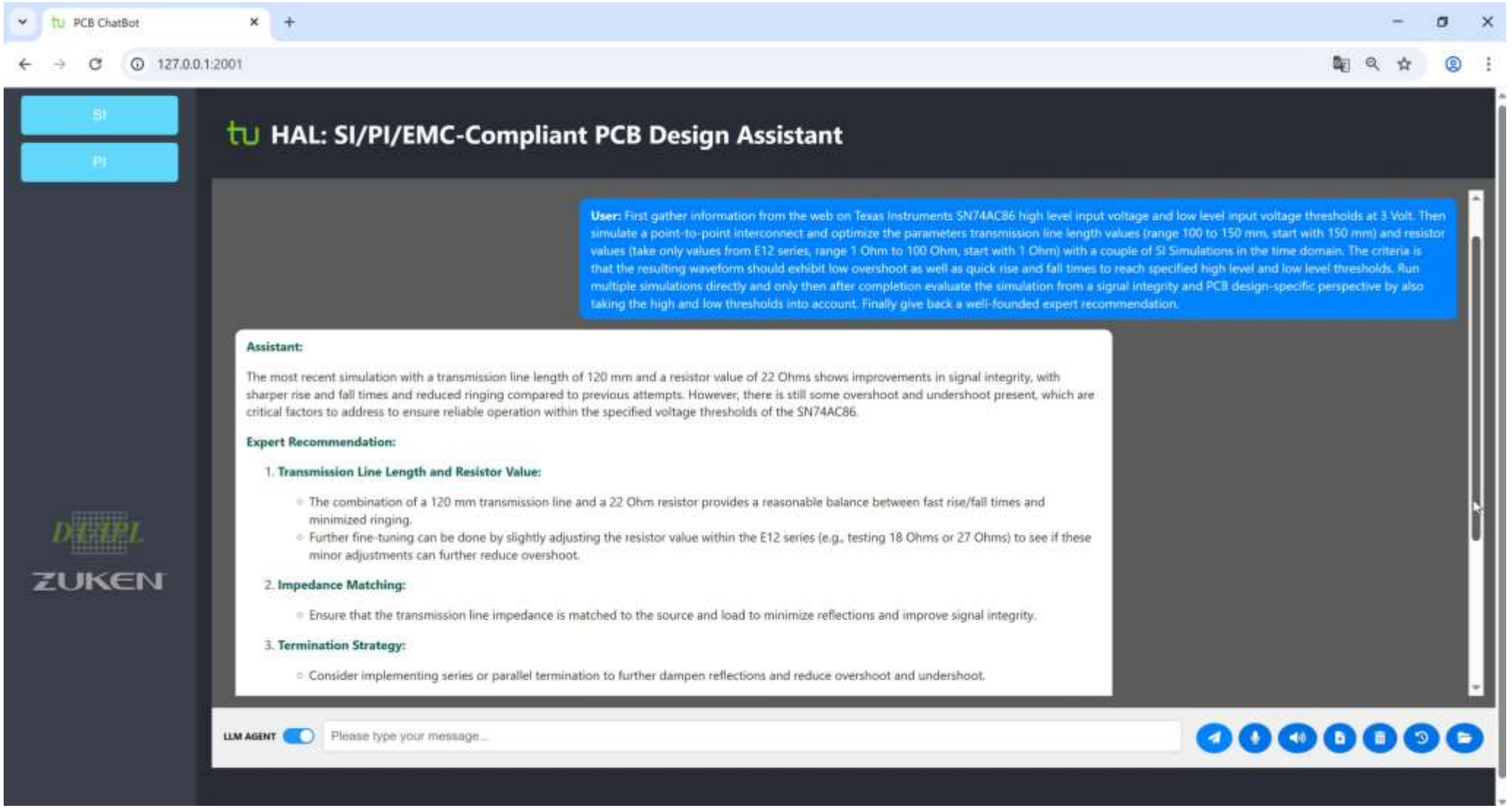
Anwendungsbeispiele #1 – Rasa SI Design Support

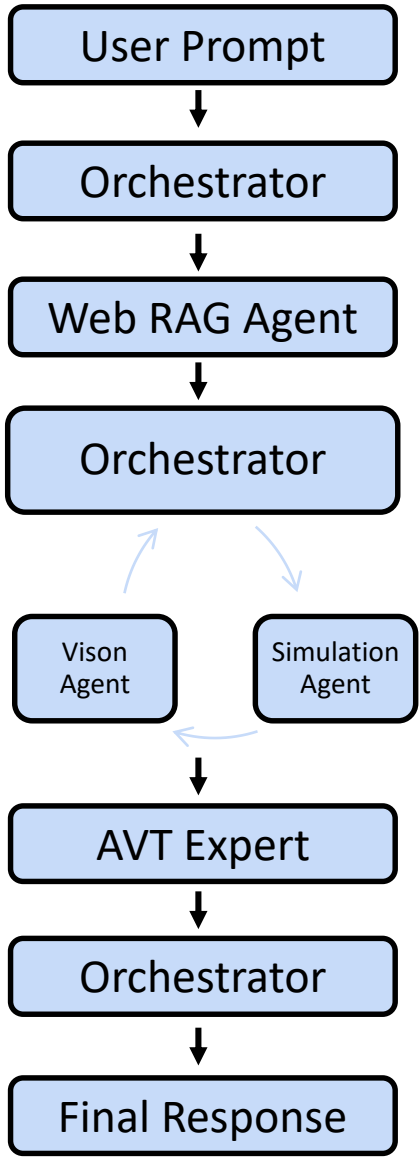


Anwendungsbeispiel #2 – Fine-Tuned LLM Chat



Anwendungsbeispiel #3 – LLM Agent Task





Zusammenfassung und Ausblick

- **Regelbasierte Aufgaben** mittels **KI-Verfahren** automatisieren (**LLM/Generative KI**) → **Entscheidender Vorteil** im Vergleich zu **konventionellen EDA-Methoden** (Regelüberprüfung mittels Simulation und geometrischen Verfahren)
- Dieser Ansatz trägt dazu bei, **manuell-gesteuerte Entwurfsschritte** zu **minimieren** und so zu erreichen, dass die Entwickler (Ingenieure) sich auf Innovationen und damit komplexe Problemlösungen konzentrieren können.
- **LLM-gestützte Co-Pilots** besitzen das Potenzial **EDA Prozesse effizienter** zu **gestalten** und den **Entwickler** bei ausgewählten **Entwurfsaufgaben** zu **unterstützen** → AVT/Packaging-Trends: PCB - Advanced Packaging - Chiplet - IC-Substrate – EDA AddOns
- **Fine-Tuning** ermöglicht die **Spezialisierung** von **LLMs** in einer **spezifischen Domäne** → Hier: **SI/PI/EMC** im **Leiterplattendesign/AVT**

- **Zukünftig** werden **insbesondere LLM Agents** und deren **Integration** in **EDA/CAD-Tools** relevant sein, um bestimmte Aufgaben mit **Tools** (Simulation, Optimierung, Requirements, RAG, KI-Module …) innerhalb eines Loops durchzuführen und den Entwickler so noch besser unterstützen zu können.
- Forschungsansatz: *Hybride KI-Methoden* ⇔ **Wissen (Regeln) mit Lerndaten (z.B. LLM) kombinieren** (Entwicklung ChatBot *Co-Pilot Umgebung +* GenKI-Framework).
- Fokus auf **kleine**, **lokale open-weight LLMs** mit **Fine-Tuning** → **Evaluation** weiterer LLMs (**Mistral**, …)