The Anatomy of a Personal Health Agent

- Published in Google Research (2025) and give credits to all Authors.
- LLMs are redefining healthcare interfaces.
- A multi-agent system integrating data science, domain expertise, and coaching.
- Goal: Enable dynamic, personalized, and context-aware health interactions.

Why Do We Need Personal Health Agents?

Wearables
- Heart rate
- Sleep
- Steps

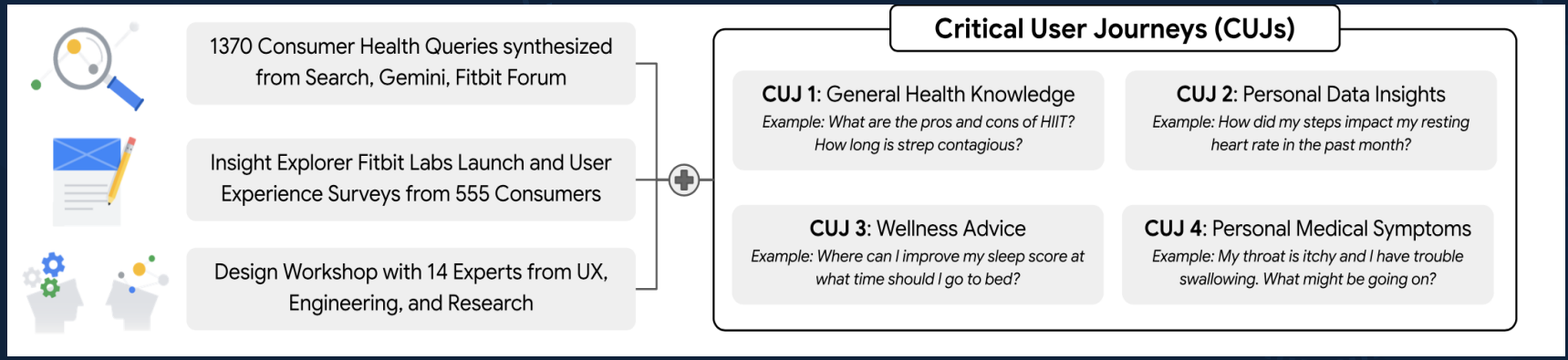
Gaps:
limited numerical reasoning, no behavioural

Personalized, context-aware health recommendations

User Research and Critical User Journeys (CUJs)

- What people actually struggle with in managing their health and how to identify ?

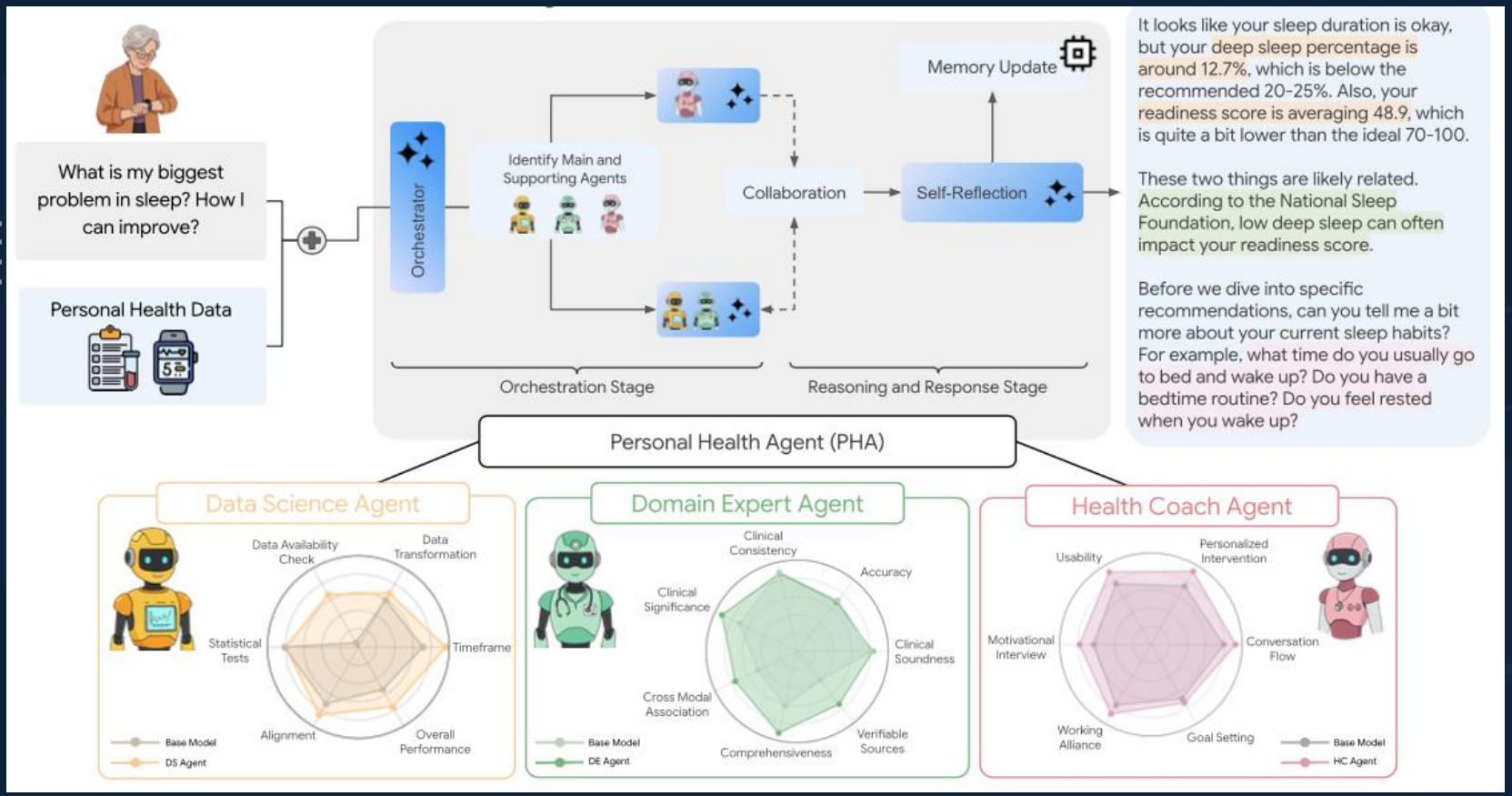
- What is CUJs ?
  They show *what users expect from an intelligent health agent* from asking questions to getting personal insights, actionable advice, and reassurance about symptoms.



System Architecture of the Personal Health Agent (PHA)

Three Specialized Agents under Orchestrator:

- Data Science Agent • Domain Expert Agent • Health Coach Agent
- All agents powered by Gemini 2.0 models



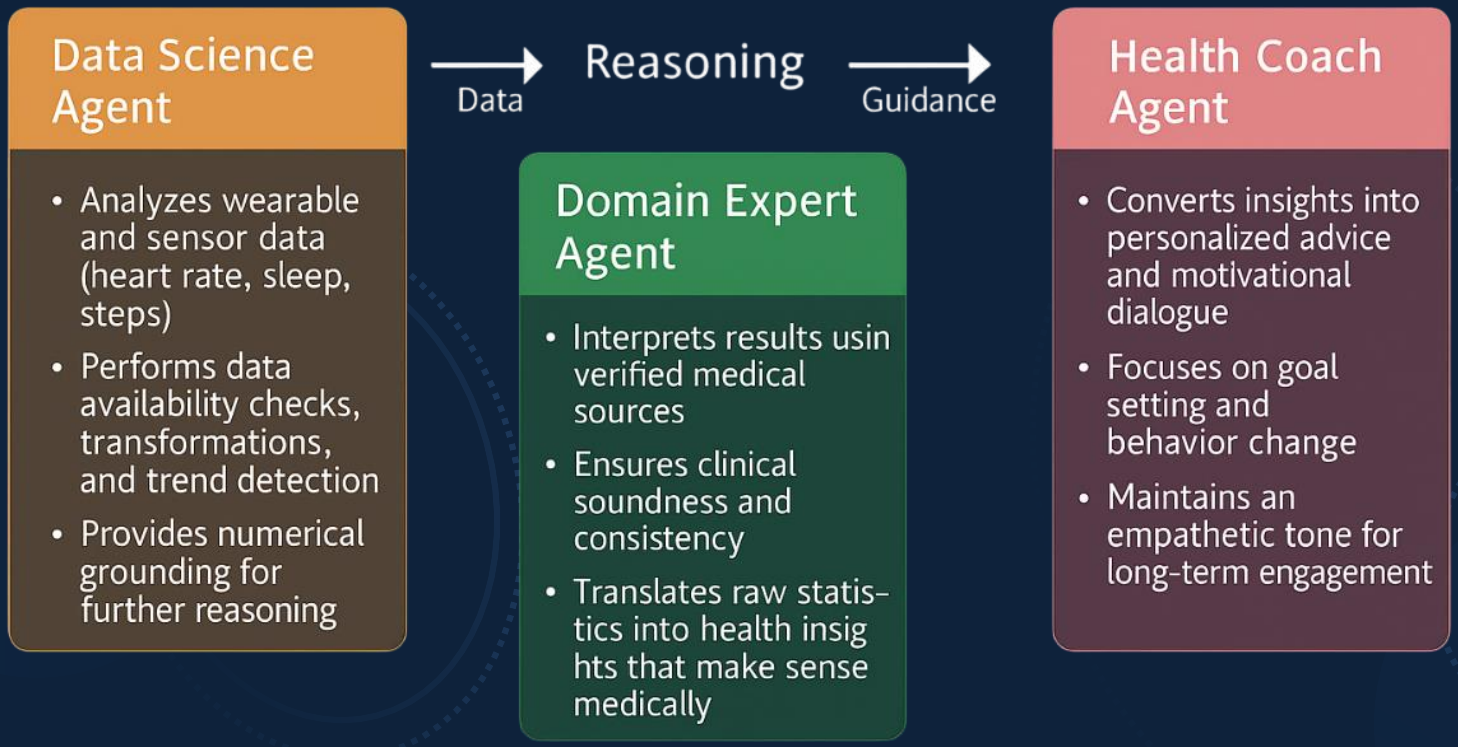
How the PHA Thinks: From Orchestration to Personalized Response

Agent Orchestration:
- The orchestrator identifies the most relevant agents.
- It manages communication and merges outputs.

Collaboration & Self-Reflection:
- Agents reason together and critique each other's answers.
- The system revises its thinking before responding.
- Memory updates help it learn over time.

Response Generation:
- Combines results into a personalized explanation.
- Communicates empathetically with actionable advice.

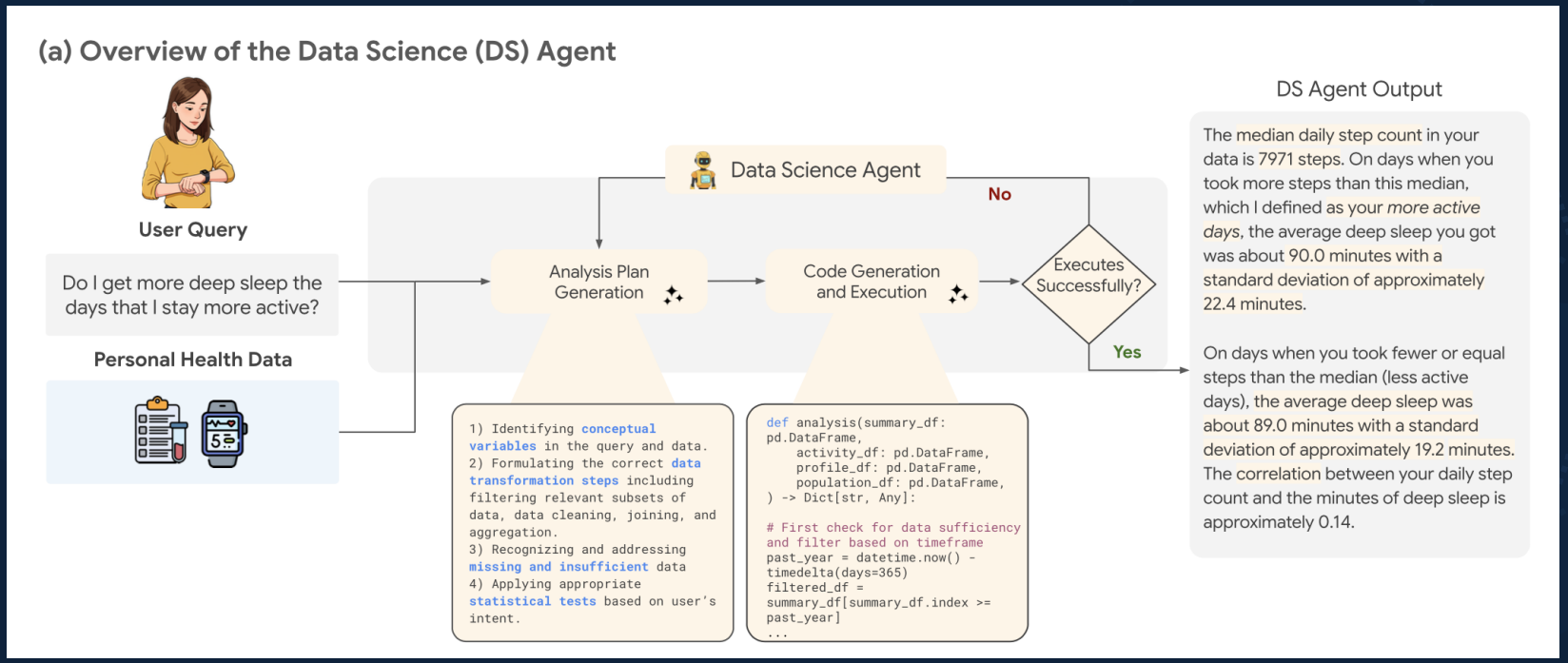
Inside the PHA: Roles of the Three Core Agents



Together, these agents transform raw data into interpretable insights and personalized actions — bridging the gap between analytics and care.

Analytical Workflow of the Data Science Agent:

- Parses natural-language health queries into structured analysis plans.
- Identifies key variables (e.g., “deep sleep” and “activity level”).
- Automatically generates and executes statistical code using data subsets.
- Produces interpretable numeric summaries.



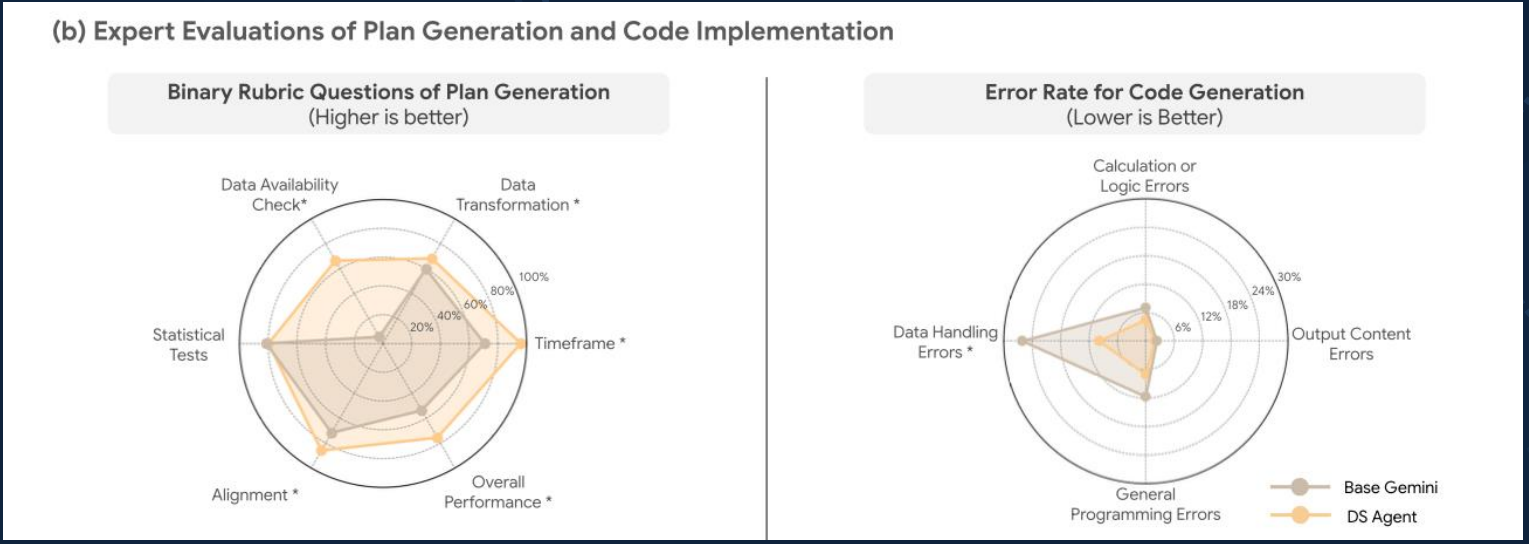
Data Science Agent - Evaluation

Evaluation Methodology:

- Experts rated the agent's *plan generation* and *code execution* against the base Gemini model.
- Metrics covered **timeframe**, **data transformation**, **data availability checks, alignment, statistical tests and error rate in code generation**

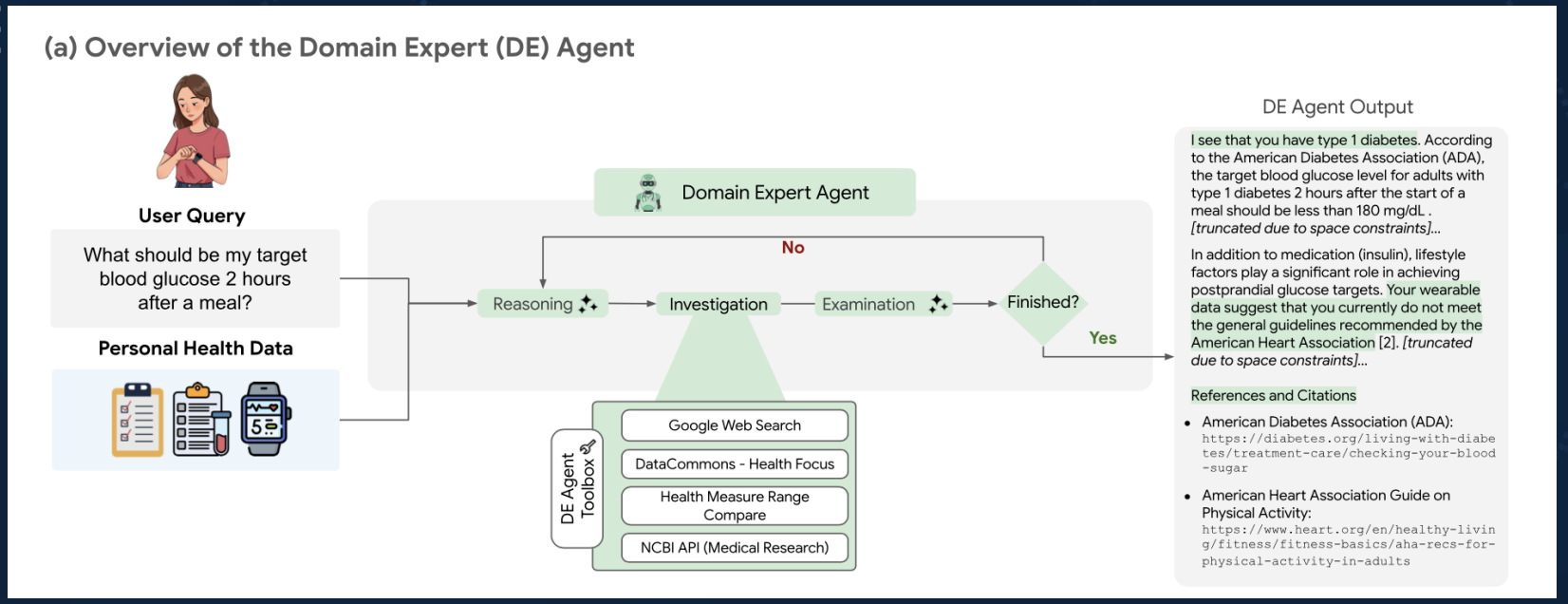
Key Findings

- The DSA consistently outperformed the base Gemini model across all six planning metrics.
- Achieved the most improvement in Timeframe and alignment.
- Reduced **data-handling and programming errors** by nearly **70 %**.
- Demonstrates strong capability in automating end-to-end statistical workflows.



Workflow of the Domain Expert Agent:

- Interprets outputs from the Data Science Agent in a clinical context.
- Uses trusted sources such as DataCommons, NCBI API, and medical web search to validate facts.
- Produces medically accurate summaries with **citations and evidence grounding**.
- Ensures responses are consistent with clinical guidelines.



Domain Expert Agent – Evaluation

Conducted **end-user and clinician-based evaluations** of single-turn question responses.

End-User Evaluation Metrics	**Clinical-Based Evaluation Metrics**
Relevance to query	Clinical Consistency
Groundedness in data	Clinical Significance
Trustworthiness	Accuracy
Personalization & contextualization	Clinical Soundness
Defining medical terms	Comprehensiveness
Credibility of citations	Verifiable Sources
	Cross-Modal Association

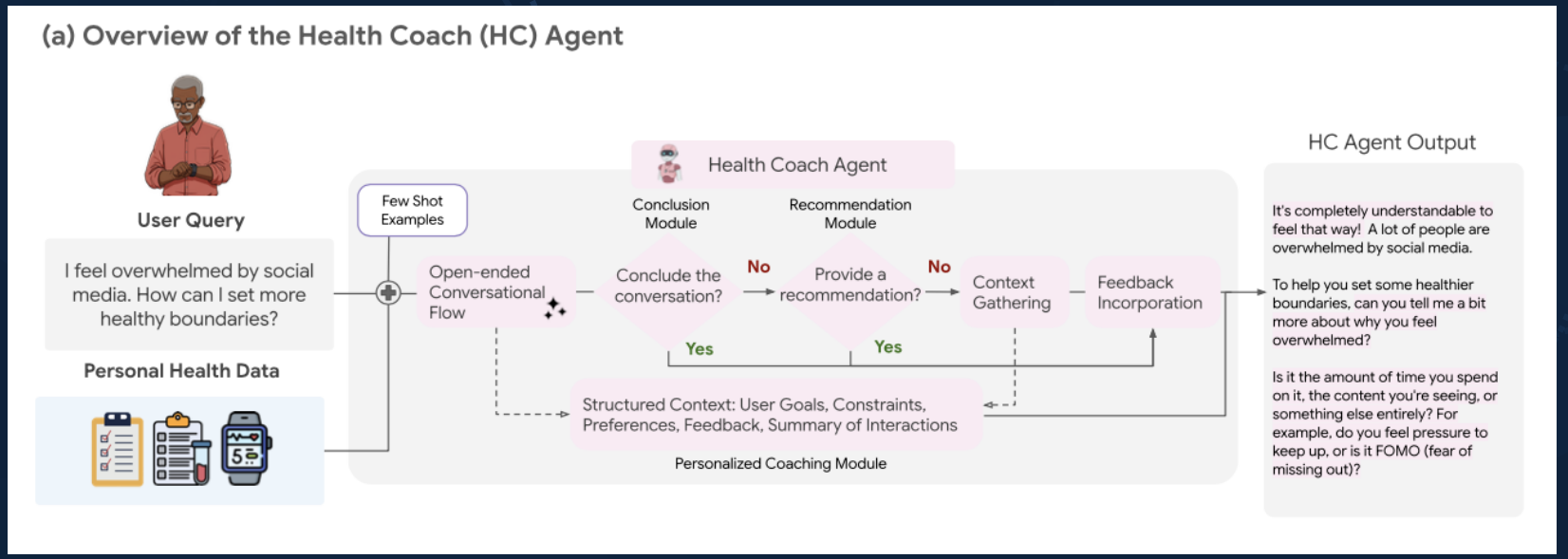
Key Findings:

- DEA scored **90–95% accuracy** on relevance, grounding, and trustworthiness.
- Achieved **>70% preference rate** in side-by-side comparisons with Gemini.
- Significantly better at explaining medical terms and providing credible citations.
- Indicates superior clinical reasoning and source verification ability.

Domain Expert Agent – Evaluation overview

Workflow of the Health Coach Agent:

- Transforms medical insights (from the Domain Expert Agent) into personalized, conversational guidance.
- Adapts its tone and phrasing to user context — encouraging motivation, goal-setting, and behavioural change.
- Integrates prior user data and memory for longitudinal tracking (sleep, stress, activity, etc.).
- Uses a reasoning loop:
  Interpret → Plan → Coach → Reflect → Adapt.



Health Coach Agent – Evaluation overview

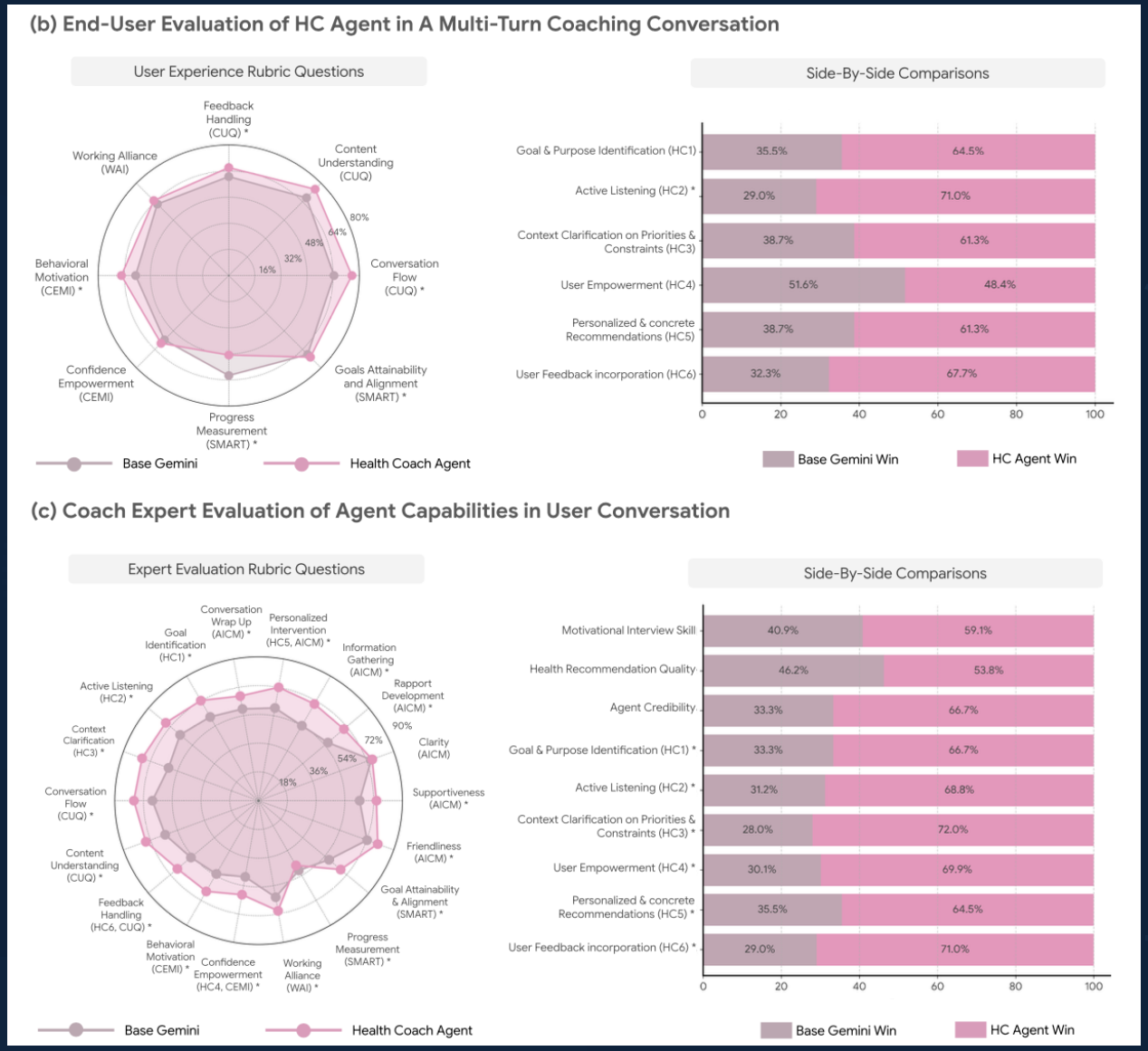
Evaluation Setup:

- Human-subject study comparing **Health Coach Agent (pink)** vs **Base Gemini 2.0 model**.
- Users rated single-turn and multi-turn coaching interactions.
- Evaluation dimensions:
  Usability & Conversation Flow
  Personalization & Goal Setting
  Motivational Interview Quality
  Working Alliance / Trust Formation

Key Findings (from paper):

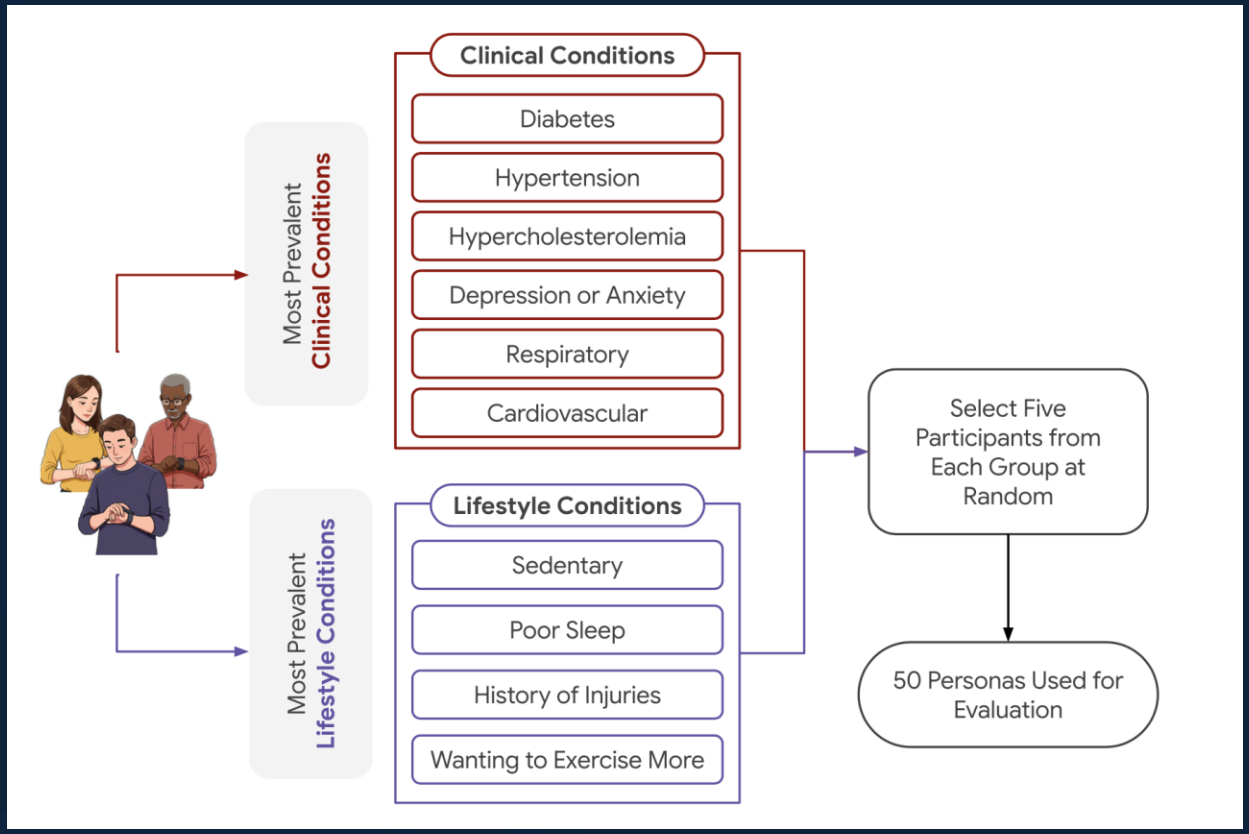
- HCA outperformed Gemini across all four dimensions.
- **+25 % gain** in "Conversation Flow & Empathy."
- **+30 % improvement** in Goal Setting and Personalized Feedback.
- Users reported the agent felt *supportive, human-like,* and *motivational.*
- Demonstrated potential for **long-term behaviour change coaching.**

Health Coach Agent – Evaluation overview



WEAR-ME Dataset – Real-World Validation

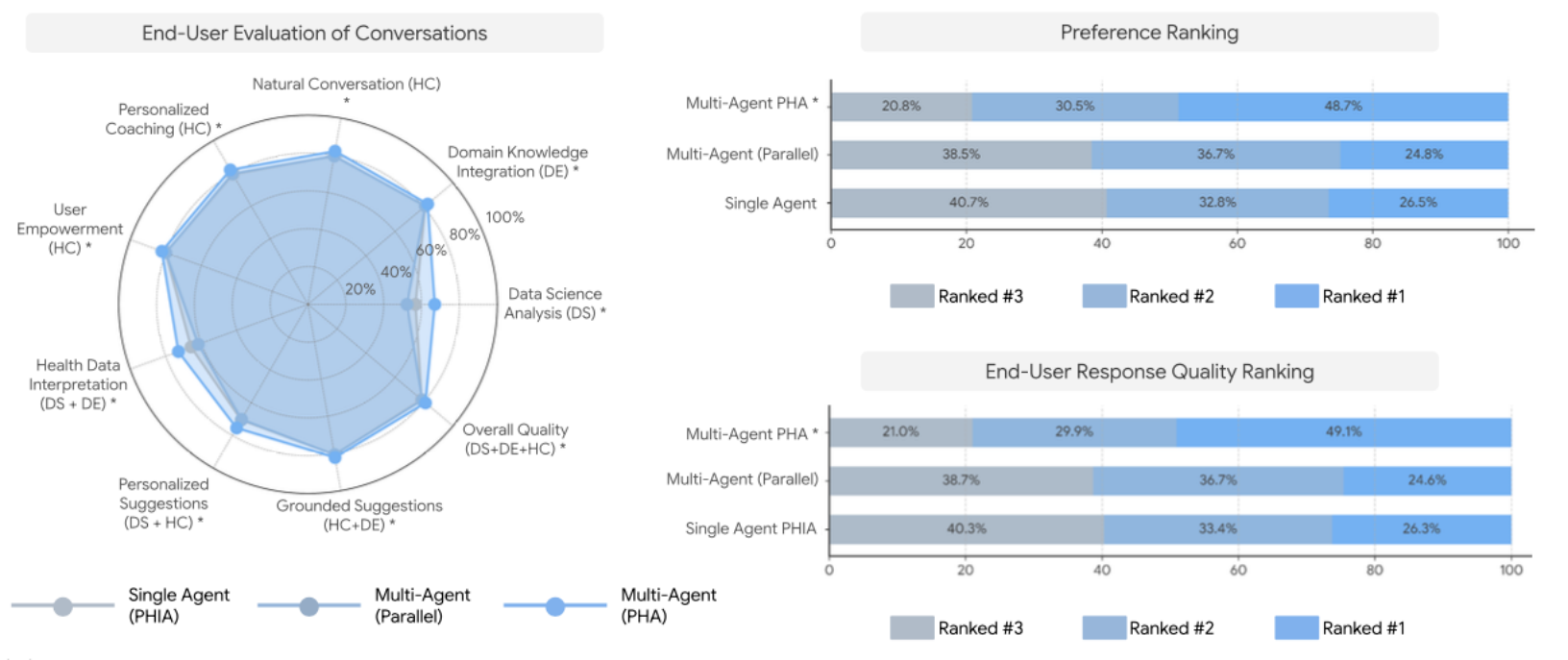
- N=1,165 participants, Fitbit + blood + surveys.
- 10 common health profiles × 5 individuals.



PHA – Evaluation

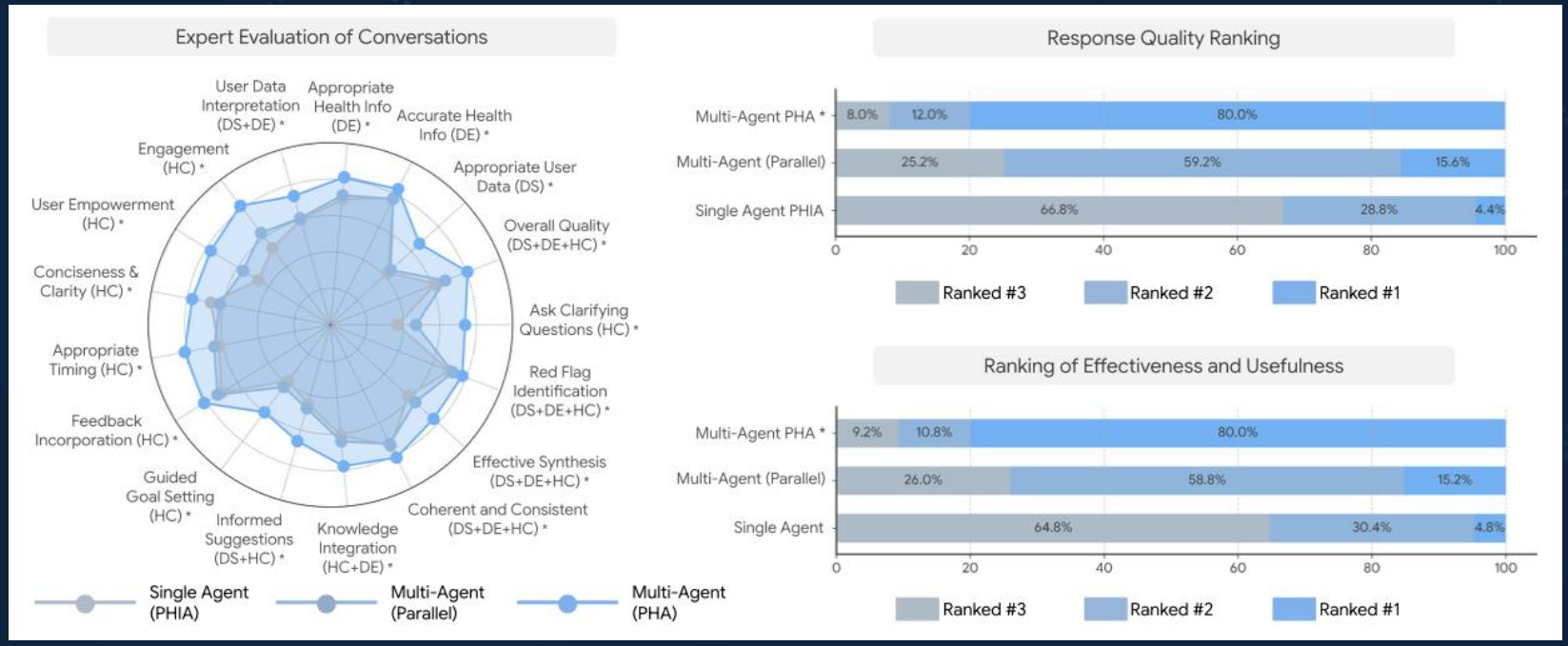
End-User Evaluation Results:
Compared 3 modes → Single Agent vs Multi-Agent (Parallel) vs Multi-Agent (PHA).
Multi-Agent PHA ranked #1 across all metrics:

- *Personalized Coaching, Data Science Analysis, and Domain Knowledge Integration.*
- Users rated PHA responses as more coherent, grounded, and motivational.



PHA – Evaluation

- Clinician and AI researcher panel assessed conversation quality.
- PHA scored highest on **accuracy, engagement, clarity, and goal feedback.**
- Notably, experts cited **effective synthesis across agents and realistic behavioural recommendations.**
- Confirms that multi-agent coordination significantly enhances both medical soundness and user trust.



Discussion — Insights and Implications of the Personal Health Agent (PHA)

1. Multi-Agent Collaboration → Improved Reasoning

• Combining specialized agents (DS + DE + HC) yields deeper analysis and context-aware responses.

• Orchestrator ensures that quantitative data, clinical logic, and behavioural feedback stay consistent.

2. Human Centred AI Coaching

• The Health Coach Agent transforms numeric and medical insights into *empathetic, motivational dialogue*.

• Encourages sustained engagement — users perceive the agent as supportive, not robotic.

3. Trust, Transparency & Accountability

• Citing medical sources (ADA, AHA) increases user trust.

• Clear separation of reasoning roles → explainable outputs for researchers and clinicians.

4. Limitations & Future Directions

• Current evaluation on short-term interactions; needs longitudinal studies.

• Real-time wearable integration and personalization memory still developing.

• Potential expansion to **mental-health coaching** and **multi-modal input (voice + sensor + text)**.

Conclusion & Future Vision - Toward Personalized AI Health Companions

Key Takeaways:

Introduced the **Personal Health Agent (PHA)** a multi-agent framework combining:

1. **Data Science Agent (quantitative reasoning)**
2. **Domain Expert Agent (clinical knowledge)**
3. **Health Coach Agent (behavioural guidance)**

- Demonstrated that **multi-agent orchestration > single-agent reasoning** — improving accuracy, empathy, and trust.
- Validated through **user and expert evaluations**, showing higher satisfaction, transparency, and engagement.

Looking Forward:

- Expand toward **long-term adaptive coaching** and **multi-modal inputs (voice, biometric, text)**.
- Develop **personal memory and privacy-aware on-device agents**.
- Bridge research → clinical applications in preventive and mental health.

Thank You