

What Makes a Trusted Rural Physician in Hawai'i? A Qualitative Pilot Study for AI in Medical Education

Bryn Higuchi² ; Jill Morimoto² ; Michele Favreau Haight¹ ; K. Kanoho Hosoda¹ ;
John A. Burns School of Medicine¹ , University of Hawai'i Manoa²

Background

- According to the 2025 Hawai'i Physician Workforce Assessment Project, there is an estimated Statewide Physician Shortage of 23%.¹
- Limited studies have characterized the attributes of successful rural physicians in Hawai'i.^{2,3}

Research Questions

- What Are the Attributes of a Trusted Physician in Rural Hawai'i?
- How do AI and human coding skills compare?

Methods

Qualitative analyses of four rural physician and community member interviews.

- Open Coding: Identify themes
- Axial Coding: Collate themes into larger categories

Emergent Categories

- Empathy** - the ability to understand, share, and vicariously experience the emotions and perspectives of another person.
- Open Mindedness** - being receptive to new ideas, arguments, and perspectives that differ from your own.
- Longevity** - a physician practicing in a rural community for a long duration of time.
- Communication Skills** - the abilities used to give and receive information clearly, accurately, and empathically.
- Rural Environment** - being from and understanding a geographic area that is located outside of large towns and cities, such as the Hawaiian islands besides Oahu.
- Cultural Humility** - a lifelong process of self-reflection and self-critique of others' cultures, but starting with an examination of her/his own beliefs and cultural identities.

Quantitative Analysis Inter-Rater Reliability

- Open Coding w/ Human Coder 2 and Gemini 3.6 Flash
- Inter-Rater Reliability : Fleiss's Kappa & Cohen's Kappa Analysis

Results

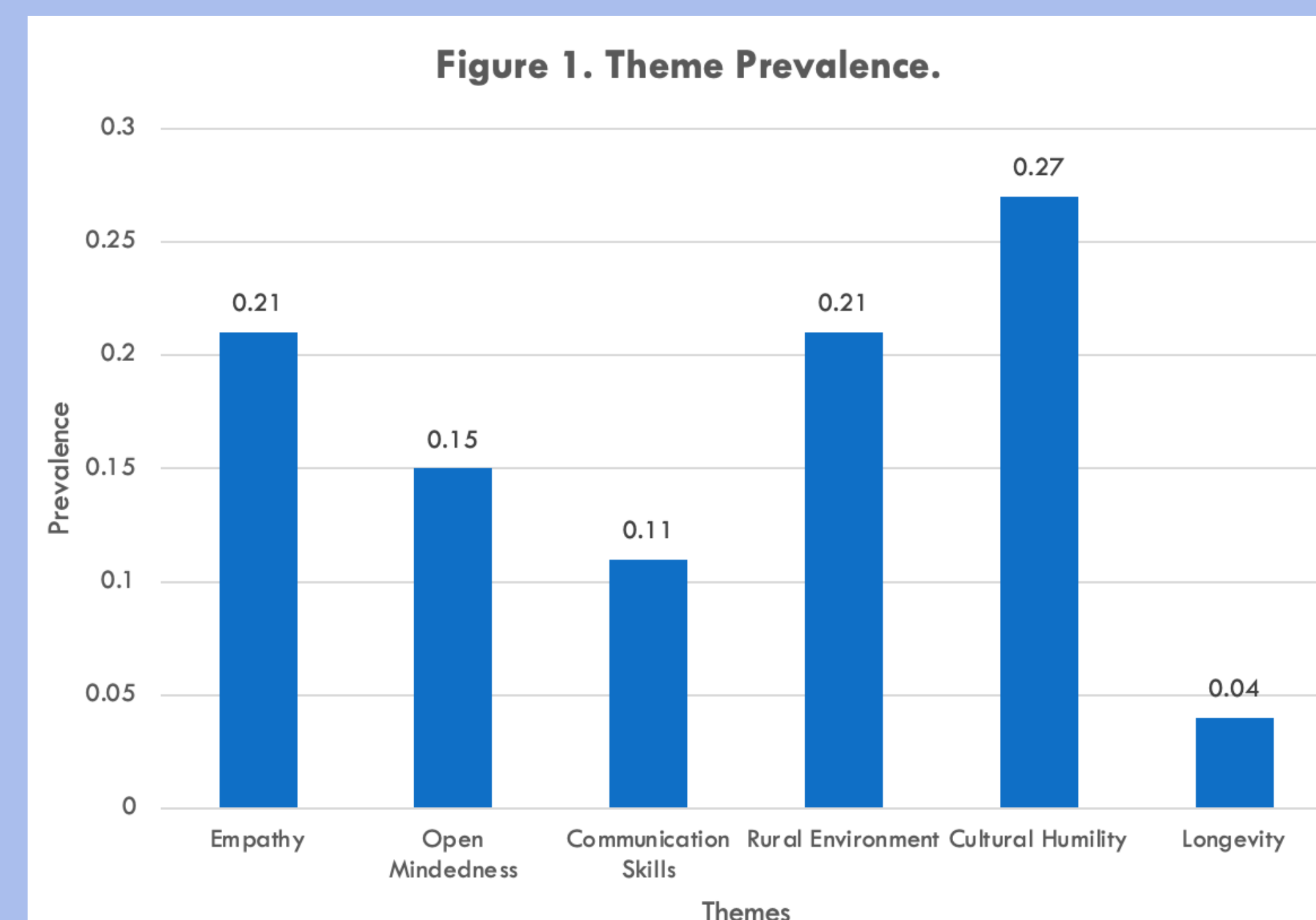


Figure 1. Theme Prevalence. Cultural humility was the most prevalent theme (0.27), followed by empathy and rural environment (0.21 each). Longevity was least prevalent (0.04).

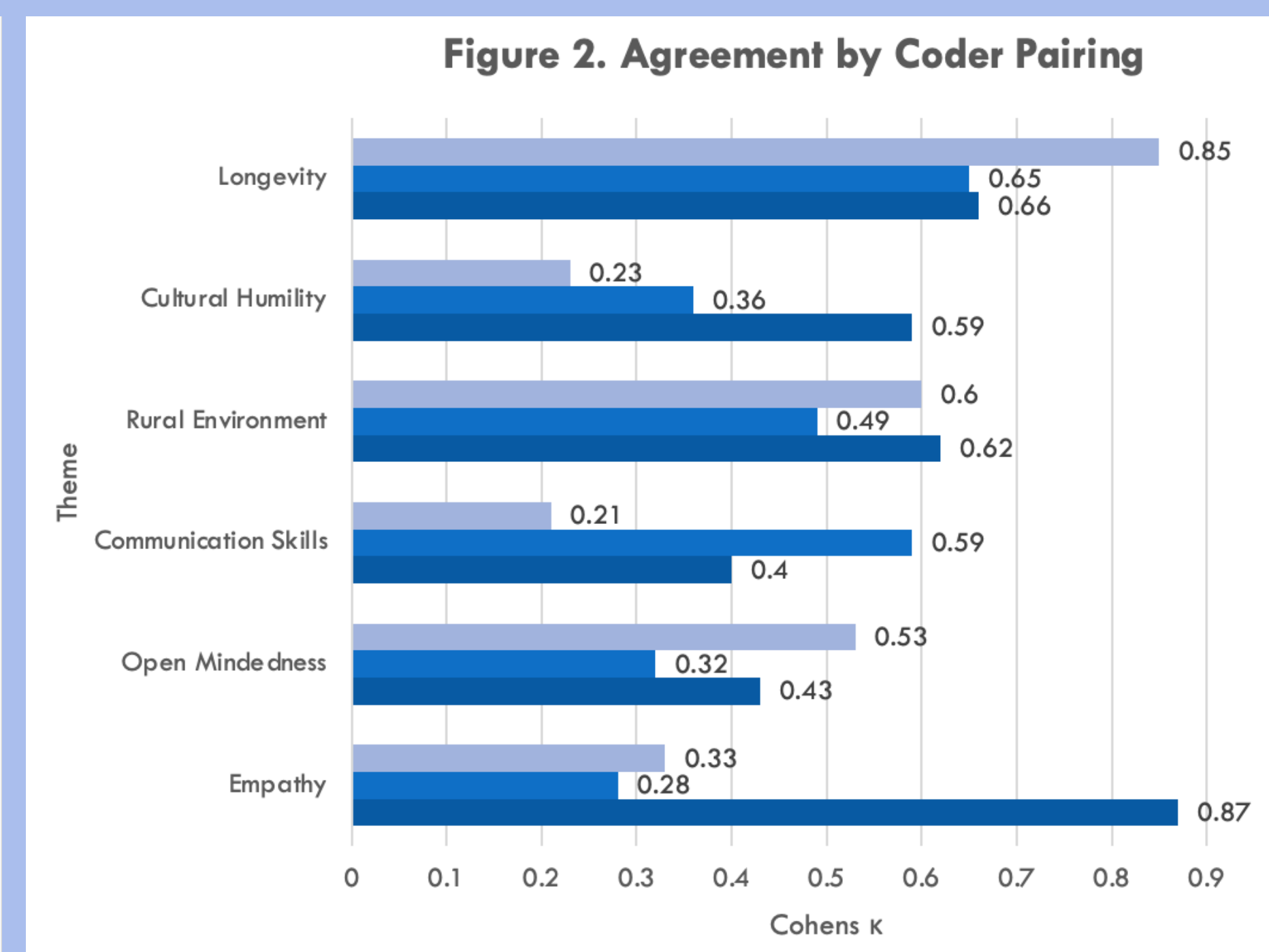


Figure 2. Agreement by Coder Pairing. Human-human agreement was generally stronger than human-AI agreement across most themes, with the largest differences seen for empathy and cultural humility.

Category	Prevalence	Exact Agreement	Fleiss' κ	HC1-HC2 κ	HC2-Gemini κ	HC1-Gemini κ
Empathy	0.21	0.66	0.56	0.87	0.33	0.28
Open Mindedness	0.15	0.71	0.59	0.43	0.53	0.32
Communication Skills	0.11	0.78	0.61	0.4	0.21	0.59
Rural Environment	0.21	0.76	0.45	0.62	0.6	0.49
Cultural Humility	0.27	0.63	0.53	0.59	0.23	0.36
Longevity	0.04	0.95	0.65	0.66	0.85	0.65

Table 1. Theme prevalence and inter-rater agreement across human and AI coders. Prevalence represents the proportion of coded excerpts assigned to each theme. Exact agreement is the proportion of coding decisions shared across coders. Fleiss' κ summarizes overall agreement among the three coders, while pairwise Cohen's κ values compare agreement between the two human coders and between each human coder and Gemini.

Discussion

Key Patterns in Inter-Rater Reliability

- Substantial Human Baseline**
 - Human coders (HC1-HC2) achieved strong baseline consensus (Pooled κ = 0.64).
- Moderate Overall Reliability**
 - The 3 Raters reached moderate agreement (Pooled Fleiss κ = 0.50), with overall reliability dampened by AI divergence from human raters.
- Coding Definition Clarity**
 - Lower agreement between human coders on Communication Skills (HC1-HC2 κ = 0.40 vs. HC1/GM κ = 0.59) highlighted an ambiguous definition requiring clearer boundaries between general interaction and specific clinical dialogue.
- AI tends to overcode or misinterpret implicit emotional context and nuance, but excels at identifying factual and explicit categories.

Conclusions

Cultural Humility, Open Mindedness, & Empathy

- All themes are in moderate agreement or more, but these are the most prevalent.

Human coders are essential, while LLMs are Task-Dependent AI Utility

- AI may assist with structured or concrete coding tasks, but culturally grounded qualitative interpretation requires human oversight.

Future Steps

- Detailed Prompt Formation (zero-shot prompting to few-shot prompt)
- Recruitment of more Human Coders.
- Inform the development of culturally responsive augmented intelligence in medical education

References

- Withy K. Annual Report on Findings from the Hawai'i Physician Workforce Assessment Project. 2025. Accessed August 2, 2026. https://www.hawaii.edu/govrel/docs/reports/2026/act18-sslh2009_2026_physician-workforce_annual-report_508.pdf
- Anderson CM, McAllaster SA, Yoneoka GM, et al. Hawai'i Rural Health Program: Shaping the Next Generation of Rural Doctors. Hawaii J Health Soc Welf. 2022;81(4 Suppl 2):20-23.
- HANA: Health and Nurturing through Artificial Intelligence | John A. Burns School of Medicine. Hawaii.edu2026. Accessed August 2, 2026. <https://nhcoe.jabsom.hawaii.edu/project-h%C4%81nai.html>

Acknowledgments

Thank you to Dr. Hosoda and Dr. Favreau for generously sharing their time and expertise throughout this project. Their mentorship fostered my understanding and skills. Additionally, thank you to Lori Tsue and the QMC Office of Research and Development for supporting the 2026 Queen's Summer Research Internship. This project was supported in part by the AMA ChangeMedEd® Initiative at the University of Hawai'i John A. Burns School of Medicine.

Scan for more info about Project HANA

