# Using a Multilingual AI Care Agent to Reduce Disparities in Colorectal Cancer Screening: Higher FIT Test Adoption Among Spanish-Speaking Patients

Meenesh Bhimani, MD, MHA[1*]; R. Hal Baker, MD[2]; Markel Sanz Ausin, PhD[1]; Gerald Meixiong, MS[1]; Rae Lasko, BS[1]; Mariska Raglow-Defranco, BA[1]; Alex Miller, BS[1]; Subhabrata Mukherjee, PhD[1]; Saad Godil, MEng[1]; Anderson Cook, MFA[1]; Jonathan D. Agnew, PhD, MBA[3]; Ashish Atreja, MD, MPH[4]

**Affiliations**
1. Hippocratic AI, Palo Alto, California, USA
2. WellSpan Health, York, Pennsylvania, USA
3. School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada
4. UC Davis Health, Davis, California, USA

***Corresponding Author**
Meenesh Bhimani, MD, MHA
Co-founder & Chief Medical Officer
Hippocratic AI
167 Hamilton Ave 3rd Floor
Palo Alto, CA 94301
USA
Email: research@hippocraticai.com

**Abstract**

**Background**: Colorectal cancer (CRC) screening rates remain disproportionately low among Hispanic and Latino populations. While artificial intelligence (AI) has shown promise in healthcare delivery, its impact on health equity remains unclear.

**Objective**: To evaluate the effectiveness of a bilingual generative AI voice agent outreach program in engaging Spanish-speaking patients for CRC screening compared to English-speaking patients.

**Methods**: We conducted a retrospective analysis of AI-powered outreach calls for CRC screening at a large integrated health system serving central Pennsylvania and northern Maryland in September 2024. The study included 1,878 patients (517 Spanish-speaking, 1,361 English-speaking) eligible for colorectal cancer screening. The AI care agent conducted personalized phone calls in the patient's preferred language to discuss screening and facilitate fecal immunochemical test (FIT) kit requests. The primary outcome was FIT test opt-in rate. Secondary outcomes included call connect rates and duration.

**Results**: Spanish-speaking patients demonstrated significantly higher engagement across all measures compared to English-speaking patients: FIT test opt-in rates (18.2% vs. 7.1%, $p<0.001$), connect rates (88.8% vs. 53.3%, $p<0.001$), and call duration (6.05 vs. 4.03 minutes, $p<0.001$). In multivariate analysis, Spanish language preference remained an independent predictor of FIT test opt-in (adjusted OR 2.012, 95% CI 1.340-3.019, $p<0.001$) after controlling for demographic and other factors (gender, age, state of residence, and call duration).

**Conclusions**: Contrary to concerns about technology exacerbating disparities, AI-powered outreach achieved significantly higher engagement among Spanish-speaking patients. These findings suggest that

language-concordant AI interactions may help address healthcare disparities and improve preventive care engagement in traditionally underserved populations.

**Keywords**
Colorectal cancer screening, artificial intelligence, healthcare disparities, Hispanic Americans, preventive health services

**Declaration of Conflicts of Interest**
MB, MSA, GM, RL, MRD, AM, SM, SG and AC are employees of Hippocratic AI, which provided funding for this study. RHB is an employee of WellSpan, which provided data for this study. JDA is an Adjunct Professor at the University of British Columbia and received compensation for work performed on this project. AA is an employee of UC Davis Health and received compensation for work performed on this project. All authors have reviewed and approved the manuscript and materials included in this submission.

**Funding Statement**
This research was supported by Hippocratic AI, Inc.

**Data Availability Statement**
Data supporting the results can be accessed by contacting the corresponding author.

**Ethics Statement**
This study analyzed anonymized administrative data that were exempt from ethics review. No individual patient consent was required as all data were de-identified prior to analysis. The study was conducted in accordance with relevant privacy and data protection regulations.

**Declaration of AI Tool Usage**
AI tools were used in preparation of this manuscript for initial draft assistance with manuscript language and structure, grammar and style refinement, and reference formatting. All AI-generated content was thoroughly reviewed, verified, and edited by the authors. All statistical analyses, interpretations, and conclusions were conducted and validated by the human authors. The authors take full responsibility for the final content of this manuscript.

**Word Count**
Abstract: 267 words
Main Text (including headers and tables): 2233 words

**References**: 24

**Date of submission**: December 16, 2024

# Background

Colorectal cancer (CRC) screening is crucial for reducing CRC-related mortality by detecting and removing precancerous adenomas or identifying cancer at earlier stages.[1,2] Various screening methods are available, including colonoscopy, fecal immunochemical tests (FIT), stool DNA tests and CT colonography, each with different levels of effectiveness and potential harms.[3] Recent advances in understanding CRC pathogenesis and improvements in screening technologies have enhanced early detection capabilities.[4] Although some concerns remain regarding the appropriate balance of benefits, risks, and costs, CRC screening is recommended in major guidelines and viewed as a critical public health intervention that can significantly reduce the burden of this prevalent cancer.[2,5–7]

Despite these benefits, data consistently show lower CRC screening rates among the Hispanic population in the United States compared to non-Hispanic whites, with studies reporting disparities ranging from 13.5% to 17%.[8–10] Common barriers to CRC screening among Hispanics include fear, cost, lack of awareness, low literacy/education levels, and limited English proficiency.[11] Effective interventions to increase screening rates include culturally tailored patient education, navigation services, and provider training.[12–14] Nonetheless, disparities persist, with significant regional variations across the United States.[10]

Previous research has demonstrated that technological innovations can improve preventive health screening rates among Hispanic populations, suggesting promising applications for artificial intelligence in this domain. Multiple digital approaches have already shown success: mHealth interventions featuring educational videos and interactive multimedia have increased cancer screening participation, while a combination of bilingual patient navigators, secure SMS messaging, and at-home testing achieved an increase in CRC screening rates.[15,16]

Earlier successes with phone-based interventions in Hispanic and psychiatric populations, along with traditional ‘promotora’ programs that improved preventive exam compliance by 35%, suggest that AI could effectively augment these established approaches.[17–19] Given that tailored patient-centered technologies have consistently improved health outcomes and reduced disparities, AI represents a potential next step in this technological evolution.[20] Nonetheless, its impact in healthcare remains unclear, especially with respect to AI care agents. Indeed, some have suggested that AI-based interventions in healthcare may disadvantage non-English-speaking and minority populations due to biases in development and deployment.[21,22] Implementation science is considered a key factor in determining the optimal outcome of AI-based initaitives.[23]

# Methods

We hypothesized that AI-powered outreach using a bilingual AI care agent would demonstrate comparable or higher engagement and FIT test opt-in rates among Spanish-speaking patients compared to English-speaking patients, potentially helping to reduce existing screening disparities between these populations. Our primary aim was to evaluate the effectiveness of a multilingual AI-powered outreach system in improving CRC screening rates among Spanish-speaking patients compared to English-speaking patients.

## Study Design

This retrospective observational study analyzed data from an AI-powered outreach initiative conducted during the week of September 16-24, 2024, at WellSpan Health, an integrated health system serving central Pennsylvania and northern Maryland. The study focused on comparing the effectiveness of multilingual AI-powered outreach for CRC screening between Spanish- and English-speaking patients. As the study analyzed anonymized administrative data, it was exempt from ethics review, and no individual patient consent was required.

## Patient Population

The study population comprised 1,878 patients eligible for CRC screening who did not have active online health system profiles. Of these, 517 (28%) were Spanish-speaking, and 1,361 (72%) were English-speaking patients. Patients were included if they were due for CRC screening and had a verified date of birth. The study specifically targeted patients who were not actively engaged with the healthcare system and were unlikely to be reached through traditional outreach methods such as email.

## Intervention

The intervention used an AI-powered conversational agent named Ana, designed to conduct personalized outreach calls for CRC screening. The platform was developed to engage patients in natural, empathetic telephone conversations while providing education about CRC screening and facilitating FIT test kit requests. The AI care agent was programmed with full bilingual capabilities in English and Spanish, allowing for natural conversation in the patient's preferred language. The system was designed to engage in culturally appropriate dialogue, explain the importance of CRC screening, address patient questions and concerns, and facilitate the scheduling of FIT test kit deliveries. Calls were initiated to patients' registered phone numbers, with language preference determined at the start of each interaction by the AI care agent.[24] We used process models to map out the processes required to integrate AI into clinical workflow; identify potential issues before they become barriers in the patient outreach; and adjust strategies for better outcomes.[23]

## Outcome Measures

Data were collected automatically through the AI platform's integrated tracking system. The primary outcome measure was the FIT test opt-in rate, defined as the percentage of patients who agreed to receive a FIT test kit during the AI-powered conversation. Secondary outcomes included connect rate (percentage of successful connections with patients) and call duration.

## Statistical Analysis

Analyses were performed using SPSS version 28.0.1.1(15). Descriptive statistics were calculated for demographic characteristics and outcome measures. Differences between Spanish- and English-speaking groups were assessed using independent samples t-tests for continuous variables and chi-square tests for categorical variables. A logistic regression analysis was conducted to evaluate factors associated with FIT test opt-in rates, controlling for age, gender, geographic location, and call duration. The model's goodness of fit was assessed using the Hosmer-Lemeshow test, and model performance was evaluated using Nagelkerke $R^2$. Statistical significance was set at $p < 0.001$. No missing data were present in the analyzed dataset.

# Results

## Baseline Characteristics

A total of 1,878 patients were included in the analysis, comprising 517 Spanish-speaking (27.5%) and 1,361 English-speaking (72.5%) individuals (Table 1). Significant demographic differences were observed between the language groups. Spanish-speaking patients were younger (mean age 57 ± 8.5 years vs. 61 ± 8.7 years, p<0.001) and more likely to be female (49.1% vs. 38.4%, p<0.001) compared to English-speaking patients. The majority of patients in both groups resided in Pennsylvania, though this proportion was higher among Spanish-speaking patients (99.3% vs. 95.6%, p<0.001).

### Table 1. Baseline Characteristics by Language Group

| Characteristic | Spanish-Speaking (n=517) | English-Speaking (n=1,361) | P-value |
|---|---|---|---|
| **Age, years (SD)** | 57 (8.5) | 61 (8.7) | <.001* |
| **Gender, % (n)** | | | |
| - Female | 49.1% (n=254) | 38.4% (n=523) | <.001** |
| - Male | 50.9% (n=263) | 61.6% (n=838) | |
| **Geographic Location, % (n)** | | | |
| - Pennsylvania | 99.3% (n=513) | 95.6% (1301) | <.001** |
| - Maryland | 0.6% (n=3) | 3.5% (n=48) | |
| - Other | 0.2% (n=1) | 0.9% (n=12) | |

*Independent samples t-test. **Chi-square test. SD = standard deviation

## Primary and Secondary Outcomes

Spanish-speaking patients demonstrated significantly higher engagement with the AI-powered outreach compared to English-speaking patients across all measured outcomes (Tables 2a, b, and c). The FIT test opt-in rate was more than twice as high among Spanish-speaking patients (18.2% vs. 7.1%, p<0.001). Similarly, connect rates were substantially higher in the Spanish-speaking group (88.8% vs. 53.3%, p<0.001). Additionally, Spanish-speaking patients engaged in longer conversations with the AI care agent (6.05 minutes vs. 4.03 minutes, p<0.001).

### Table 2a: Outcomes by Language Group

| Outcome | Spanish-Speaking (n=517) | English-Speaking (n=1,361) | P-value |
|---|---|---|---|
| **FIT Test Opt-in, % (n)** | 18.2% (94) | 7.1% (97) | <.001** |
| **Connect Rate, % (n)** | 88.8% (459) | 53.3% (725) | <.001** |
| **Call Duration, minutes (SD)*** ** | 6.05 (3.20) | 4.03 (3.44) | <.001* |

*Independent samples t-test. **Chi-square test. SD = standard deviation. ***Excludes observations where the call connect was not verified.

**Table 2b: FIT Test Opt-in and Connect Rates by Language Group**

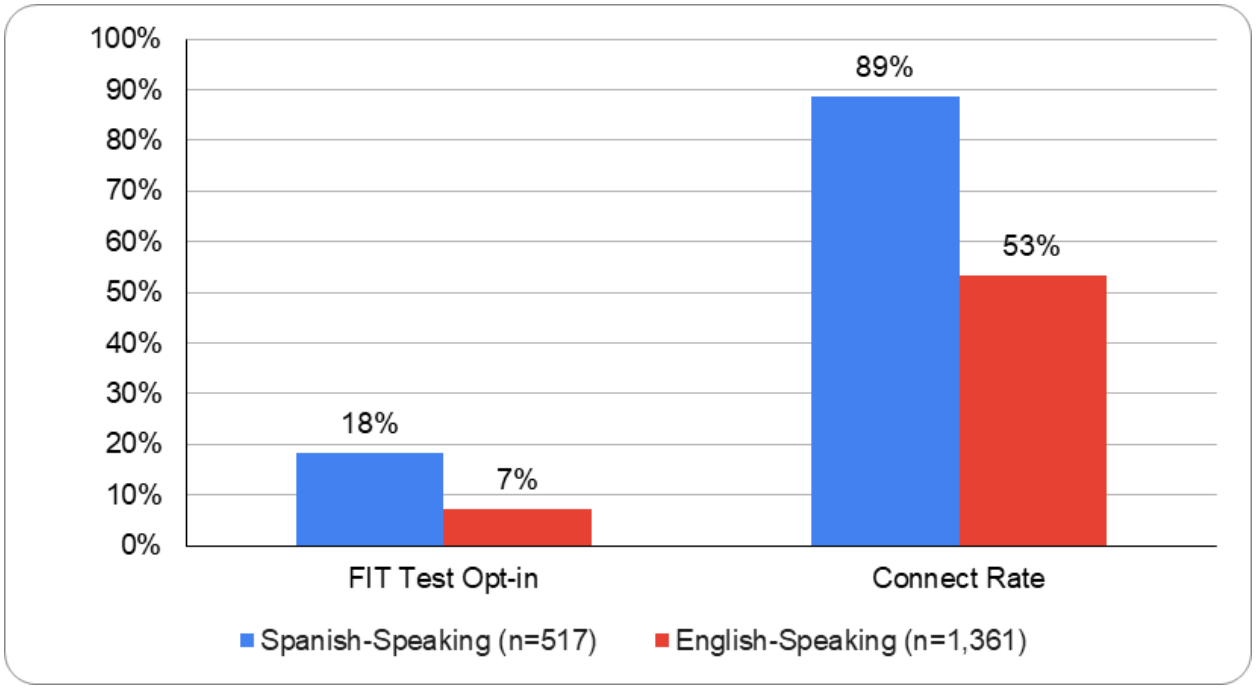
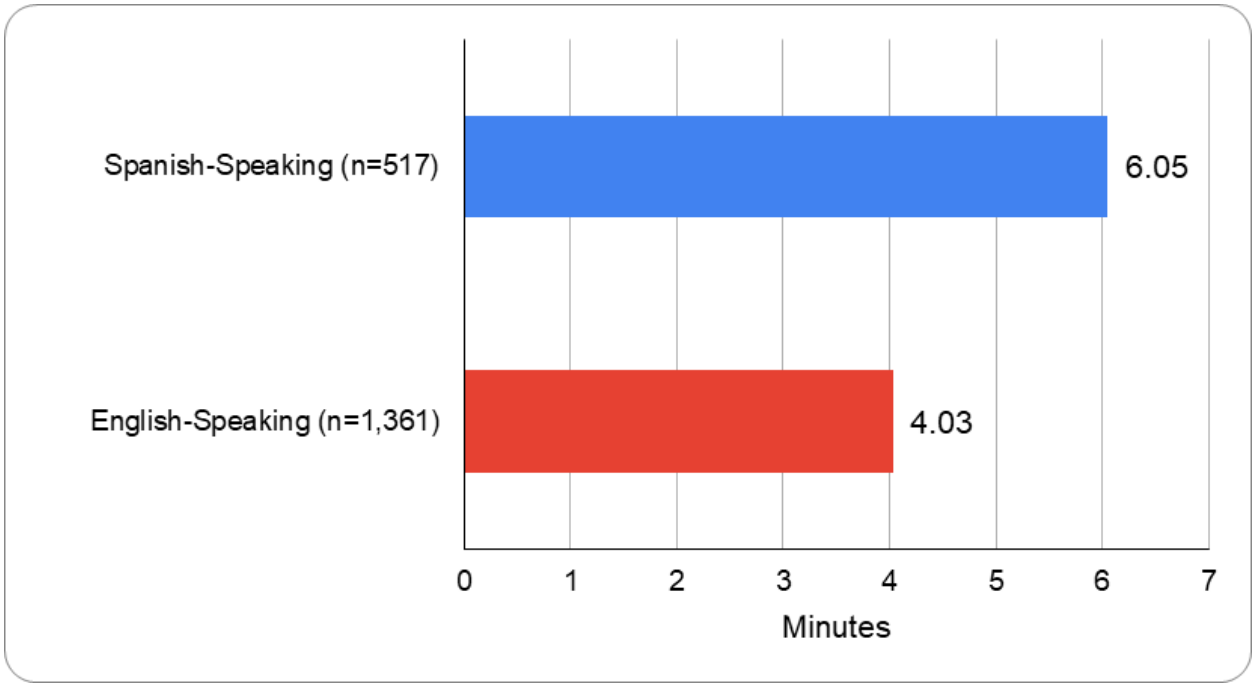



**Table 2c: Average Call Duration, in Minutes, by Language Group**



## Multivariate Analysis

A logistic regression analysis revealed that Spanish language preference remained an independent predictor of FIT test opt-in after adjusting for age, gender, geographic location, and call duration (Table 3). Spanish-speaking patients were more than twice as likely to opt in for FIT testing compared to English-speaking patients (adjusted OR 2.012, 95% CI 1.340-3.019, p<0.001). Call duration was also significantly associated with opt-in likelihood (adjusted OR 1.682 per minute, 95% CI 1.583-1.787, p<0.001). Age, gender, and geographic location were not significantly associated with FIT test opt-in rates.

### Table 3. Logistic Regression Analysis of Factors Associated with FIT Test Opt-in

| Variable | B Coefficient (SE) | Adjusted OR (95% CI) | P-value |
|---|---|---|---|
| **Age (per year)** | -0.010 (0.11) | 0.990 (0.968 – 1.1012) | N/S |
| **State** | | | |
| **- Pennsylvania** | Reference | 1.00 | -- |
| **- Maryland or Other** | -1.527 (0.908) | 0.217 (0.037 – 1.288) | N/S |
| **Gender** | | | |
| **- Male** | Reference | 1.00 | -- |
| **- Female** | -0.296 (0.200) | 0.744 (0.503 – 1.101) | N/S |
| **Call Duration (per minute)** | 0.520 (0.031) | 1.682 (1.583 – 1.787) | <0.001 |
| **Language** | | | |
| **- English** | Reference | 1.00 | -- |
| **- Spanish** | 0.699 (0.207) | 2.012 (1.340 – 3.019) | <0.001 |
| **Constant** | -3.195 (0.715) | -- | <0.001 |

OR = Odds Ratio; CI = Confidence Interval; SE = Standard Error. N/S = not significant. The dependent variable was FIT test opt-in (yes/no). Model fit statistics: Hosmer-Lemeshow $\chi^2(8) = 30.486$, $p = 0.001$; Nagelkerke $R^2 = 0.445$; Model $\chi^2(5) = 452.681$, $p = <0.001$.

The regression model demonstrated good explanatory power (Nagelkerke $R^2 = 0.445$) and overall significance (Model $\chi^2(5) = 452.681$, $p<0.001$), though the Hosmer-Lemeshow test suggested potential issues with model fit ($\chi^2(8) = 30.486$, $p = 0.001$). These findings indicate that language preference was a robust predictor of engagement with CRC screening outreach, independent of other demographic factors.

# Discussion

Our study suggests that AI-powered outreach can effectively engage Spanish-speaking populations in CRC, with outcomes exceeding those observed in English-speaking patients. The 2.6-fold higher FIT test opt-in rate among Spanish-speaking patients (18.2% vs. 7.1%) represents a departure from historical patterns of healthcare disparities – a noteworthy finding given that previous studies have documented consistently lower CRC screening rates among Hispanic and Latino populations (53.4%) compared to non-Hispanic whites (70.4%).[10] Moreover, higher engagement levels observed among Spanish-speaking patients, evidenced by superior connect rates (88.8% vs. 53.3%) and longer call durations (6.05 vs. 4.03 minutes), suggest that AI outreach may help overcome traditional barriers to healthcare access and engagement.

Our multivariate analysis revealed that Spanish language preference remained a significant and independent predictor of FIT test opt-in even after controlling for demographic factors and call duration, with Spanish-speaking patients twice as likely to opt-in to screening compared to their English-speaking

counterparts. This finding challenges the narrative that AI-based interventions inherently disadvantage non-English-speaking and minority populations.

The study benefits from several strengths, including a large sample size with substantial representation of Spanish-speaking patients and complete data capture through automated systems. The real-world implementation in a diverse healthcare setting, coupled with statistical analysis controlling for potential confounders, provides evidence for the effectiveness of the intervention. The direct comparison of language groups within the same intervention period further strengthens our findings.

However, important limitations must be considered. The single healthcare system setting may limit generalizability to other contexts, and the short study duration may not capture longer-term engagement patterns. The absence of follow-up data on FIT test completion rates prevents assessment of the intervention's impact on actual screening behaviors. We were unable to assess socioeconomic, medical, and other factors that might influence engagement, and potential selection bias in the initial patient identification process cannot be ruled out.

Our findings suggest that AI-powered outreach can effectively complement existing care delivery systems, particularly for traditionally underserved populations. The higher engagement rates among Spanish-speaking patients indicate that language-concordant AI interactions may help address longstanding disparities in preventive care access and utilization. Healthcare systems seeking to improve screening rates among diverse populations should consider implementing multilingual AI outreach as part of their comprehensive screening strategy.

The success of this intervention challenges the assumption that technological solutions inherently exacerbate healthcare disparities. Rather, our results suggest that thoughtfully designed AI systems can help bridge care gaps and promote health equity. This has important implications for healthcare policy, particularly regarding the allocation of resources for technological innovations in healthcare delivery.

The markedly higher engagement among Spanish-speaking patients suggests that AI-powered outreach may be particularly effective in reaching traditionally underserved populations. This finding is especially relevant given the historical challenges in engaging Hispanic and Latino communities in preventive care programs. The success of this intervention demonstrates that AI, when properly implemented, can serve as a tool for promoting health equity rather than perpetuating disparities.

Future research should evaluate long-term engagement patterns and screening completion rates, while also assessing the impact on clinical outcomes and cancer detection rates. Investigation of AI outreach effectiveness in other languages and cultural contexts would provide valuable insights for expansion. Studies examining cost-effectiveness, resource utilization, and patient experience with AI-powered interactions would further inform implementation strategies.

The successful implementation of this program in a large healthcare system suggests potential for broader scalability. Key considerations for expansion include infrastructure requirements for implementing multilingual AI systems, integration with existing healthcare workflows and electronic health records, training requirements for healthcare staff, and cost implications of system-wide implementation.

# Conclusion

This study provides compelling evidence that carefully designed AI-powered outreach through AI care agents can effectively engage diverse populations in preventive care, particularly benefiting traditionally underserved communities. As healthcare systems increasingly adopt technological solutions, careful attention to language access and cultural competency will be crucial in ensuring these innovations promote rather than hinder health equity.

# References


1. Nicholson FB, Barro JL, Atkin W, et al. Review article: Population Screening for Colorectal Cancer. *Aliment Pharmacol Ther*. 2005;22(11-12):1069-1077.

2. Levin B, Lieberman DA, McFarland B, et al. Screening and Surveillance for the Early Detection of Colorectal Cancer and Adenomatous Polyps, 2008: A Joint Guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA Cancer J Clin*. 2008;58(3):130-160.

3. Lin JS, Perdue LA, Henrikson NB, Bean SI, Blasi PR. Screening for Colorectal Cancer. *JAMA*. 2021;325(19):1978.

4. Li D. Recent advances in colorectal cancer screening. *Chronic Dis Transl Med*. 2018/9;4(3):139-147.

5. Qaseem A. Screening for Colorectal Cancer: A Guidance Statement From the American College of Physicians. *Ann Intern Med*. 2012;156(5):378.

6. Shaukat A, Levin TR. Current and future colorectal cancer screening strategies. *Nat Rev Gastroenterol Hepatol*. 2022;19(8):521-531.

7. Ness RM, Llor X, Abbass MA, et al. NCCN guidelines® insights: Colorectal Cancer screening, version 1.2024: Featured updates to the NCCN Guidelines. *J Natl Compr Canc Netw*. 2024;22(7):438-446.

8. Pollack LA, Blackman DK, Wilson KM, Seeff LC, Nadel MR. Colorectal cancer test use among Hispanic and non-Hispanic U.S. populations. *Prev Chronic Dis*. 2006;3(2):A50.

9. May FP, Yang L, Corona E, Glenn BA, Bastani R. Disparities in Colorectal Cancer Screening in the United States Before and After Implementation of the Affordable Care Act. *Clin Gastroenterol Hepatol*. 2020;18(8):1796-1804.e2.

10. Viramontes O, Bastani R, Yang L, Glenn BA, Herrmann AK, May FP. Colorectal cancer screening among Hispanics in the United States: Disparities, modalities, predictors, and regional variation. *Prev Med*. 2020;138(106146):106146.

11. Wang J, Moehring J, Stuhr S, Krug M. Barriers to colorectal cancer screening in Hispanics in the United States: An integrative review. *Appl Nurs Res*. 2013;26(4):218-224.

12. Gonzalez SA, Ziebarth TH, Wang J, Noor AB, Springer DL. Interventions Promoting Colorectal Cancer Screening in the Hispanic Population: A Review of the Literature. *J Nurs Scholarsh*. 2012;44(4):332-340.

13. Naylor K, Ward J, Polite BN. Interventions to Improve Care Related to Colorectal Cancer Among Racial and Ethnic Minorities: A Systematic Review. *J Gen Intern Med*. 2012;27(8):1033-1046.

14. Mojica CM, Parra-Medina D, Vernon S. Interventions Promoting Colorectal Cancer Screening Among Latino Men: A Systematic Review. *Prev Chronic Dis*. 2018;15. doi:10.5888/pcd15.170218

15. Watanabe-Galloway S, Ratnapradipa K, Subramanian R, et al. Mobile Health (mHealth) Interventions to Increase Cancer Screening Rates in Hispanic/Latinx Populations: A Scoping Review. *Health Promot Pract*. 2022;24(6):1215-1229.

16. Rozario MA, Walton A, Kang M, Padilla BI. Colorectal Cancer Screening: A Quality Improvement Initiative Using a Bilingual Patient Navigator, Mobile Technology, and Fecal Immunochemical Testing to Engage Hispanic Adults. *Clin J Oncol Nurs*. 2021;25(4):423-429.

17. Villegas N, Cianelli R, de Tantillo L, et al. Assessment of Technology Use and Technology Preferences for HIV Prevention Among Hispanic Women. *Hisp Health Care Int*. 2018;16(4):197-203.

18. Duarte AC, Thomas SA. The Use of Phone Technology in Outpatient Populations: A Systematic Review. *Open Nurs J*. 2016;10(1):45-58.

19. Hunter JB, de Zapien JG, Papenfuss M, Fernandez ML, Meister J, Giuliano AR. The Impact of a Promotora on Increasing Routine Chronic Disease Prevention among Women Aged 40 and Older at the U.S.-Mexico Border. *Health Educ Behav*. 2004/8;31(4_suppl):18S - 28S.

20. Tarver WL, Haggstrom DA. The Use of Cancer-Specific Patient-Centered Technologies Among Underserved Populations in the United States: Systematic Review. *J Med Internet Res*. 2019;21(4):e10256.

21. Alford J, Rathod N. AI could worsen health inequities for UK's minority ethnic groups - new report. Imperial News. February 24, 2022. Accessed November 19, 2024. https://www.imperial.ac.uk/news/230413/ai-could-worsen-health-inequities-uks/

22. Nazer LH, Zatarah R, Waldrip S, et al. Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digit Health*. 2023;2(6):e0000278.

23. Longhurst CA, Singh K, Chopra A, Atreja A, Brownstein JS. A call for artificial intelligence implementation science centers to evaluate clinical effectiveness. *NEJM AI*. 2024;1(8). doi:10.1056/aip2400223

24. Hear Our GenAI Agent in Action. Hippocratic AI. Accessed December 4, 2024. https://www.hippocraticai.com/video