AI in Nursing a Year in Review

May 17, 2024

Andrew Phillips PhD, RN, FAMIA

Conflict of Interest

Have no real or apparent conflicts of interest to report.

Learning Objectives

- Evaluate themes that impact nursing informatics coming from the AI nursing literature.
- Identify gaps in the literature.
- Generate logical next steps in advancing research.

Methods – Scoping Study

Arksey and O'Malley[1]

- Step 1 – Identify the Research Question
- Step 2 – Identify Relevant Studies
- Step 3 – Study Selection (Iterative process which can change over time)
- Step 4 – Charting the Data
- Step 5 – Collating, summarizing, and reporting the results
- Step 6 – Consultation – This is you guys

[1]Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International journal of social research methodology*, *8*(1), 19-32.

Step 1: Research Question

- What trends and themes emerge from a survey of the published literature in the area of AI in nursing informatics during the past year

Step 2: Identify Relevant Studies

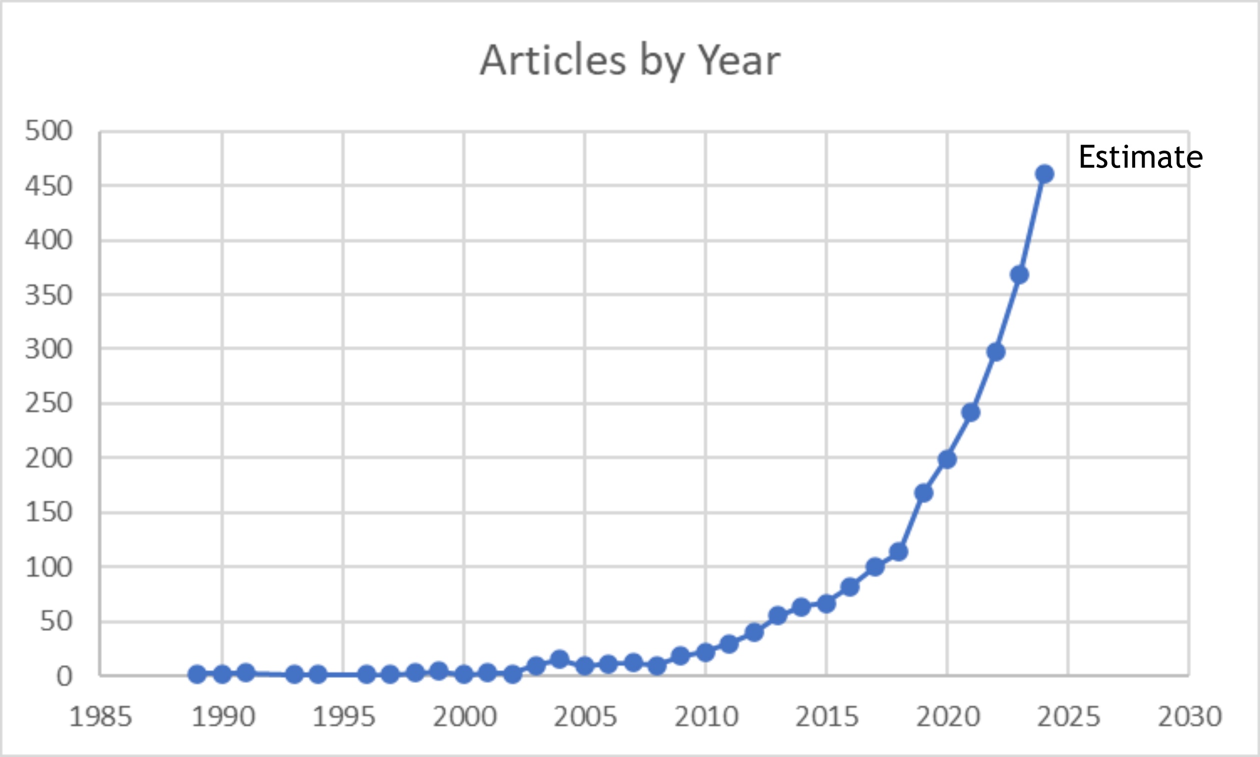
- Search Strategy
  - Databases: PubMed and CINAHL
  - Search terms
- (“artificial intelligence” OR “AI” OR “large language models”) AND healthcare AND nurs*
- Publication Dates 1/1/2023 – 3/31/2024

Step 3: Study Selection

Inclusion and Exclusion Criteria

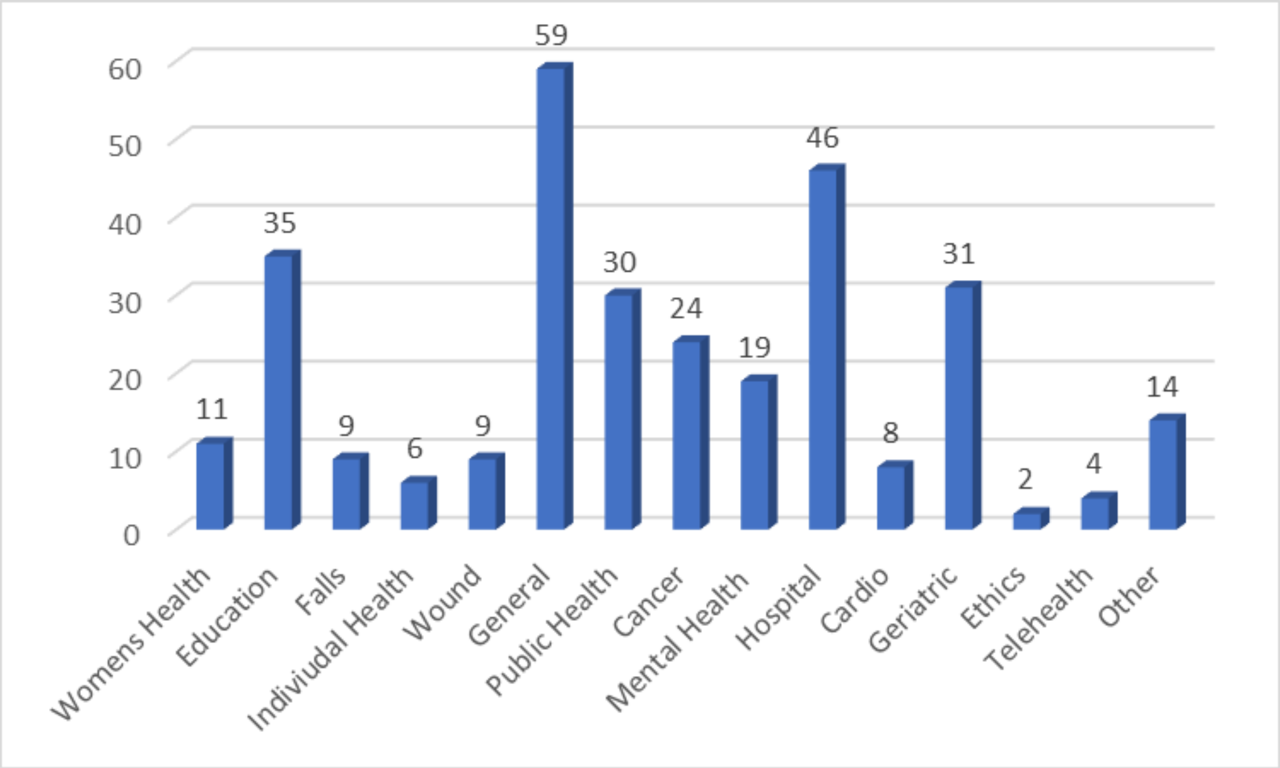
- Inclusion criteria: healthcare delivery, relevant to nursing
- Exclusions: Articles without an AI informatics focus, i.e. genomics, medical treatments

Search Results

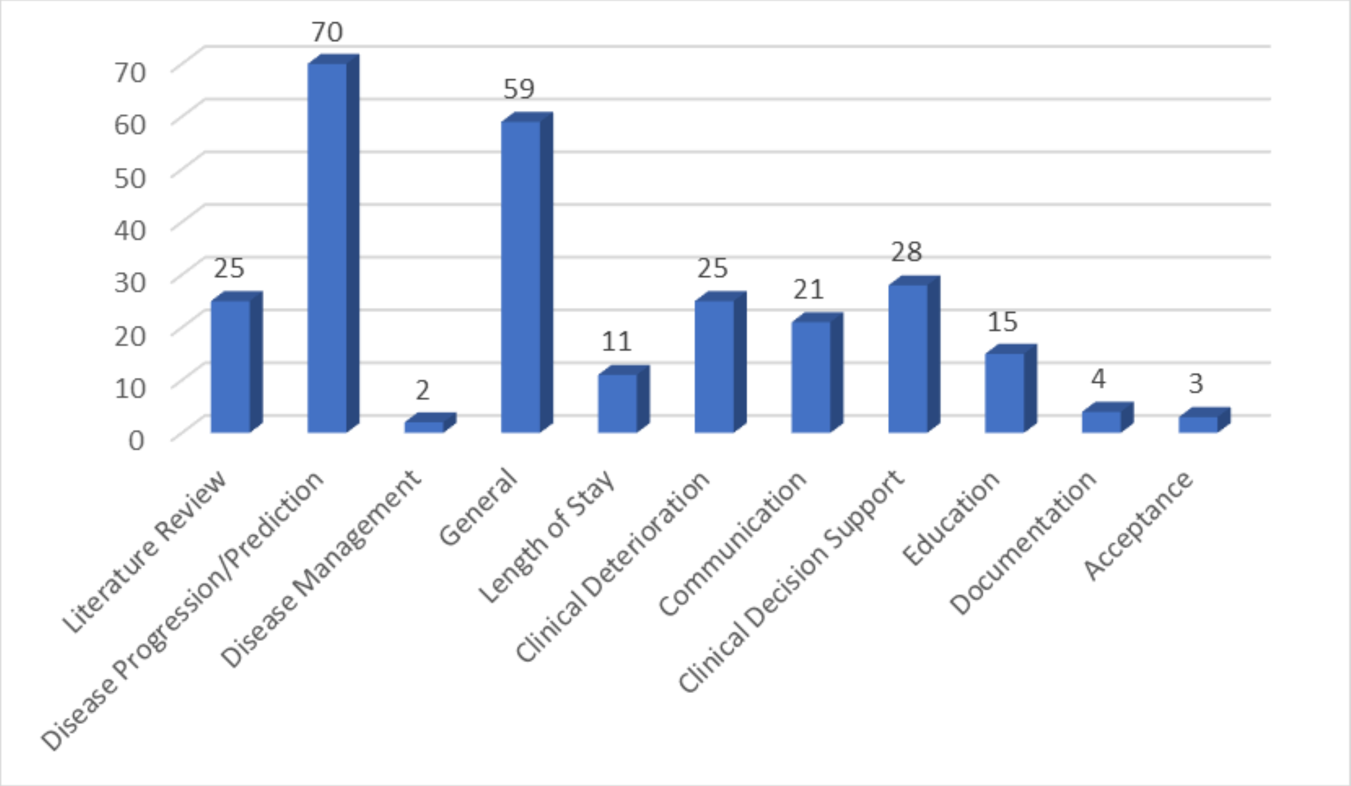


Step 4: Charting the Data

Articles by Topic/Setting



Articles by Clinical Goal



Step 5 – Collating, summarizing, and reporting the results

Themes Identified

1. Acceptance/Attitudes/Intention to Use
   - Clinicians
   - Patients
2. Education
   - Training of novice nurses
   - Preparing RNs for future of healthcare
   - Patient education
3. Ethics
   - Replace humans
   - Bias
   - Error
   - Training
4. Robots
   - Distraction
   - Emotional support
5. New sources of data to “feed” AI tools
   - Wearables
   - PROMs
   - Clinical notes
   - Patient-clinician communication
6. Comparisons to “correct” answer
   1. Trying to “perfect” the training data
7. Communication
   1. Mental health
8. ED/ICU Settings overrepresented
   1. Triage
9. How far do/can we take it as far as trust and autonomy
10. Visualization of AI results
    1. See/understand how recommendation made
11. Not as much radiology as I would have thought – nursing focus?
12. Reduce burden/fatigue
13. Cautious optimism
14. Patient privacy
15. AI generally over cautious
16. “getting your head around” the topic of AI – many lit reviews

Representative Citations

Bienefeld, N., Kolbe, M., Camen, G., Huser, D., & Buehler, P. K. (2023). Human-AI teaming: leveraging transactive memory and speaking up for enhanced team effectiveness. Frontiers in Psychology, 14, 1208019.

Dlugatch, R., Georgieva, A., & Kerasidou, A. (2024). AI-driven decision support systems and epistemic reliance: a qualitative study on obstetricians’ and midwives’ perspectives on integrating AI-driven CTG into clinical decision making. BMC Medical Ethics, 25(1), 6.

Economou-Zavlanos, N. J., Bessias, S., Cary Jr, M. P., Bedoya, A. D., Goldstein, B. A., Jelovsek, J. E., ... & Poon, E. G. (2024). Translating ethical and quality principles for the effective, safe and fair development, deployment and use of artificial intelligence technologies in healthcare. Journal of the American Medical Informatics Association, 31(3), 705-713.

Garcia, P., Ma, S. P., Shah, S., Smith, M., Jeong, Y., Devon-Sand, A., ... & Sharp, C. (2024). Artificial Intelligence–Generated Draft Replies to Patient Inbox Messages. JAMA Network Open, 7(3), e243201-e243201.

Jabbour, S., Fouhey, D., Shepard, S., Valley, T. S., Kazerooni, E. A., Banovic, N., ... & Sjoding, M. W. (2023). Measuring the impact of AI in the diagnosis of hospitalized patients: a randomized clinical vignette survey study. JAMA, 330(23), 2275-2284

Staes, C. J., Beck, A. C., Chalkidis, G., Scheese, C. H., Taft, T., Guo, J. W., ... & McPherson, J. P. (2024). Design of an interface to communicate artificial intelligence-based prognosis for patients with advanced solid tumors: a user-centered approach. Journal of the American Medical Informatics Association, 31(1), 174-187.

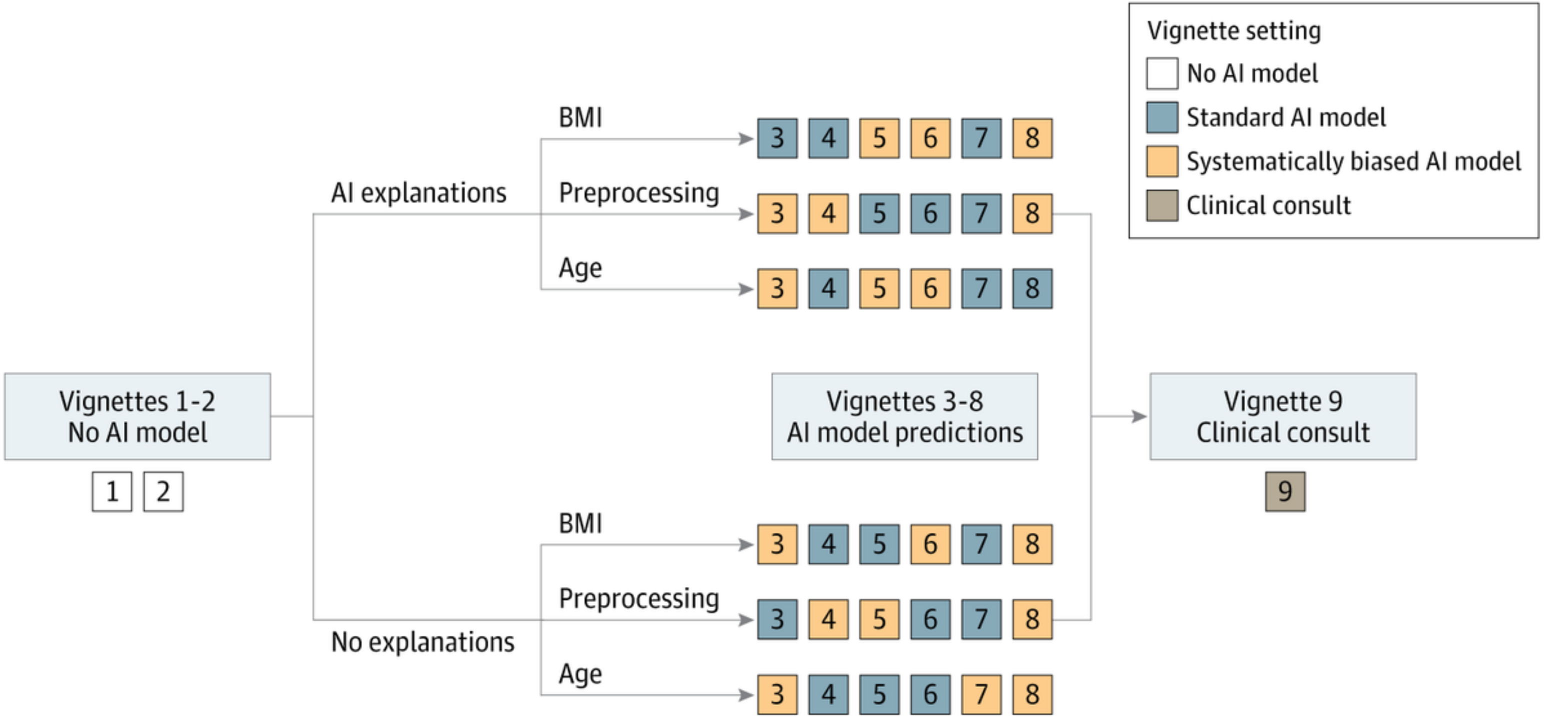
Zolnoori, M., Sridharan, S., Zolnour, A., Vergez, S., McDonald, M. V., Kostic, Z., ... & Topaz, M. (2024). Utilizing patient-nurse verbal communication in building risk identification models: the missing critical data stream in home healthcare. Journal of the American Medical Informatics Association, 31(2), 435-444.

Bienefeld, N., Kolbe, M., Camen, G., Huser, D., & Buehler, P. K. (2023). Human-AI teaming: leveraging transactive memory and speaking up for enhanced team effectiveness. Frontiers in Psychology, 14, 1208019.

- Objective: to investigate the role of transactive memory and speaking up in human-AI teams comprising 180 intensive care (ICU) physicians and nurses working with AI in a simulated clinical environment
- Methods: Resident and attending physicians and nurses from the Institute of Intensive Care Medicine at a large teaching hospital in Switzerland. In this prospective observational study, 180 ICU physicians and nurses comprising 45 teams collaborated with an AI agent to diagnose and provide medical treatment to a simulated patient suffering from a life-threatening condition. Documented and coded team interactions.
- Results: Out of the 45 teams, 22 teams (48.89%) were above the median (i.e., higher-performing), and 23 teams (51.11%) were below the median (lower-performing).
- Conclusions: **The results demonstrate that in higher-performing teams accessing knowledge from an AI agent is positively associated with a team's ability to develop new hypotheses and speaking up with doubts or concerns. In contrast, accessing knowledge from a human team member appeared to be negatively associated with hypothesis-building and speaking up, regardless of team performance.**

Jabbour, S., Fouhey, D., Shepard, S., Valley, T. S., Kazerooni, E. A., Banovic, N., ... & Sjoding, M. W. (2023). Measuring the impact of AI in the diagnosis of hospitalized patients: a randomized clinical vignette survey study. JAMA, 330(23), 2275-2284.

- Objective: To evaluate the impact of systematically biased AI on clinician diagnostic accuracy and to determine if image-based AI model explanations can mitigate model errors.
- Methods: Randomized clinical vignette survey study administered between April 2022 and January 2023 across 13 US states involving hospitalist physicians, nurse practitioners, and physician assistants. Clinicians were shown 9 clinical vignettes of patients hospitalized with acute respiratory failure, including their presenting symptoms, physical examination, laboratory results, and chest radiographs. Clinicians were then asked to determine the likelihood of pneumonia, heart failure, or chronic obstructive pulmonary disease as the underlying cause(s) of each patient's acute respiratory failure. To establish baseline diagnostic accuracy, clinicians were shown 2 vignettes without AI model input. Clinicians were then randomized to see 6 vignettes with AI model input with or without AI model explanations. Among these 6 vignettes, 3 vignettes included standard-model predictions, and 3 vignettes included systematically biased model predictions.
- Results: Clinicians' baseline diagnostic accuracy was 73.0% (95% CI, 68.3% to 77.8%) for the 3 diagnoses. When shown a standard AI model without explanations, clinician accuracy increased over baseline by 2.9 percentage points (95% CI, 0.5 to 5.2) and by 4.4 percentage points (95% CI, 2.0 to 6.9) when clinicians were also shown AI model explanations. Systematically biased AI model predictions decreased clinician accuracy by 11.3 percentage points (95% CI, 7.2 to 15.5) compared with baseline and providing biased AI model predictions with explanations decreased clinician accuracy by 9.1 percentage points (95% CI, 4.9 to 13.2) compared with baseline, representing a nonsignificant improvement of 2.3 percentage points (95% CI, −2.7 to 7.2) compared with the systematically biased AI model.
- Conclusions: **Although standard AI models improve diagnostic accuracy, systematically biased AI models reduced diagnostic accuracy, and commonly used image-based AI model explanations did not mitigate this harmful effect.**

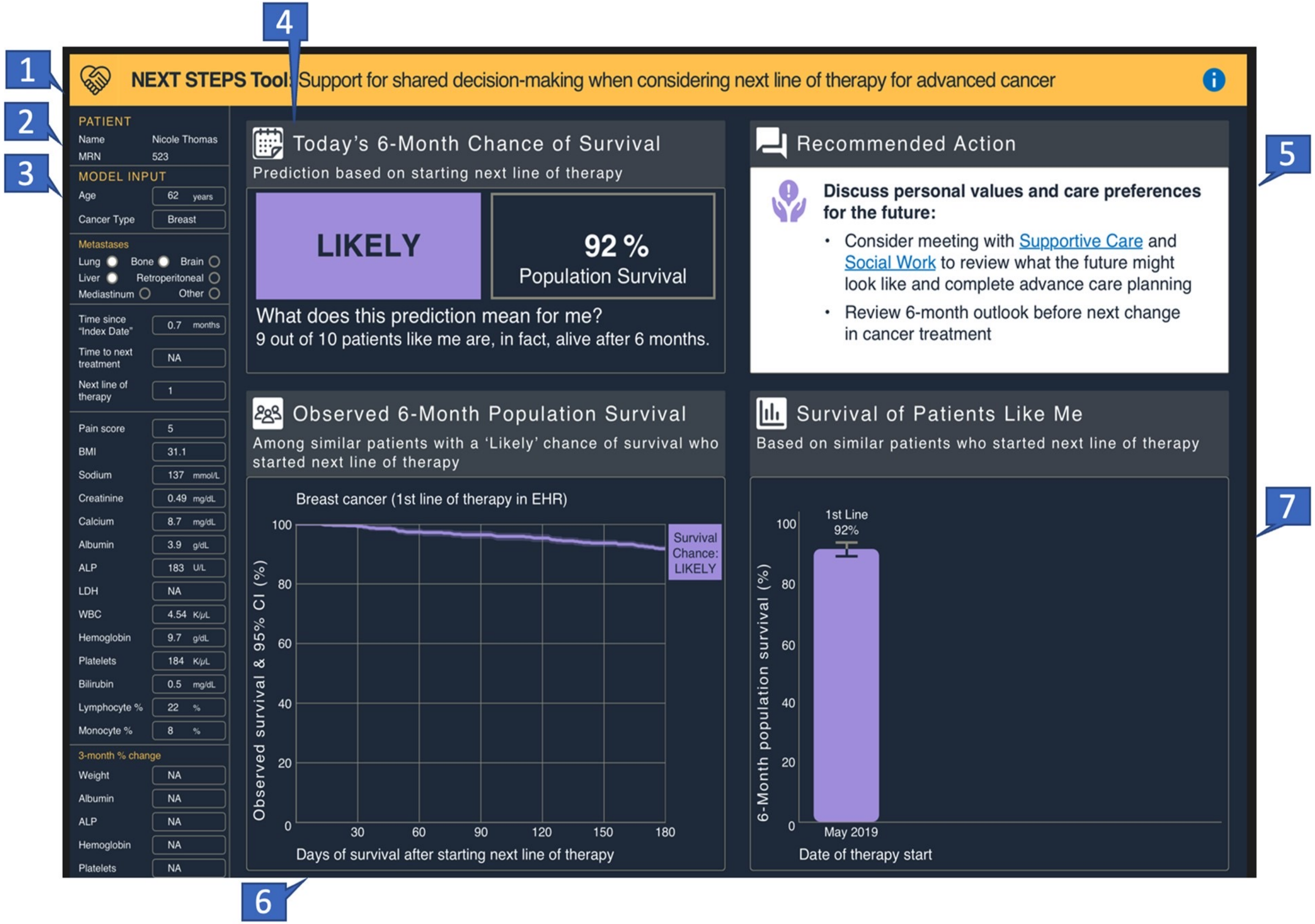

Vignette setting
No AI model
Standard AI model
Systematically biased AI model
Clinical consult
BMI
AI explanations
Preprocessing
Age
3 4 5 6 7 8
3 4 5 6 7 8
3 4 5 6 7 8
Vignettes 1-2
No AI model
1 2
Vignettes 3-8
AI model predictions
Vignette 9
Clinical consult
9
BMI
Preprocessing
No explanations
Age
3 4 5 6 7 8
3 4 5 6 7 8
3 4 5 6 7 8

Garcia, P., Ma, S. P., Shah, S., Smith, M., Jeong, Y., Devon-Sand, A., ... & Sharp, C. (2024). Artificial Intelligence–Generated Draft Replies to Patient Inbox Messages. JAMA Network Open, 7(3), e243201-e243201.

- Objective: To evaluate the implementation of a large language model used to draft responses to patient messages in the electronic inbox.
- Methods: A 5-week, prospective, single-group quality improvement study was conducted from July 10 through August 13, 2023, at a single academic medical center(Stanford Health Care). Draft replies to patient portal messages generated by a HIPAA–compliant electronic health record–integrated large language model The primary outcome was AI-generated draft reply utilizationas a percentage of total patient message replies. Secondary outcomes included changes in time measures and clinician experience as assessed by survey.
- Results: The mean AI-generated draft response utilization rate across clinicians was 20%. There was no change in reply action time, write time, or read time between the prepilot and pilot periods. There were statistically significant reductions in the 4-item physician task load score derivative (mean [SD], 61.31 [17.23] presurvey vs 47.26 [17.11] postsurvey; paired difference,−13.87; 95% CI, −17.38 to −9.50; $P < .001$) and work exhaustion scores (mean [SD], 1.95 [0.79]presurvey vs 1.62 [0.68] postsurvey; paired difference, −0.33; 95% CI, −0.50 to −0.17; $P < .001$).
- Conclusions: There was notable adoption, usability, and improvement in assessments of burden and burnout. There was no improvement in time. Further code-to-bedside testing is needed to guide future development and organizational strategy.

Staes, C. J., Beck, A. C., Chalkidis, G., Scheese, C. H., Taft, T., Guo, J. W., ... & McPherson, J. P. (2024). Design of an interface to communicate artificial intelligence-based prognosis for patients with advanced solid tumors: a user-centered approach. Journal of the American Medical Informatics Association, 31(1), 174-187.

- Objectives: To design an interface to support communication of machine learning (ML)-based prognosis for patients with advanced solid tumors, incorporating oncologists' needs and feedback throughout design.
- Methods: Using an interdisciplinary user-centered design approach, we performed 5 rounds of iterative design to refine an interface, involving expert review based on usability heuristics, input from a color-blind adult, and 13 individual semi-structured interviews with oncologists. Individual interviews included patient vignettes and a series of interfaces populated with representative patient data and predicted survival for each treatment decision point when a new line of therapy (LoT) was being considered. Ongoing feedback informed design decisions and directed qualitative content analysis of interview transcripts was used to evaluate usability and identify enhancement requirements.
- Results: Design processes resulted in an interface with 7 sections, each addressing user-focused questions, supporting oncologists to "tell a story" as they discuss prognosis during a clinical encounter. The iteratively enhanced interface both triggered and reflected design decisions relevant when attempting to communicate ML-based prognosis, and exposed misassumptions. Clinicians requested enhancements that emphasized interpretability over explainability. Qualitative findings confirmed that previously identified issues were resolved and clarified necessary enhancements (eg, use months not days) and concerns about usability and trust (eg, address LoT received elsewhere). Appropriate use should be in the context of a conversation with an oncologist.
- Conclusions: User-centered design, ongoing clinical input, and a visualization to communicate ML-related outcomes are important elements for designing any decision support tool enabled by artificial intelligence, particularly when communicating prognosis risk.



1
2
3
4
5
6
7
NEXT STEPS Tool: Support for shared decision-making when considering next line of therapy for advanced cancer
PATIENT
Name Nicole Thomas
MRN 523
MODEL INPUT
Age 62 years
Cancer Type Breast
Metastases
Lung Bone Brain
Liver Retroperitoneal
Mediastinum Other
Time since "Index Date" 0.7 months
Time to next treatment NA
Next line of therapy 1
Pain score 5
BMI 31.1
Sodium 137 mmol/L
Creatinine 0.49 mg/dL
Calcium 8.7 mg/dL
Albumin 3.9 g/dL
ALP 183 U/L
LDH NA
WBC 4.54 K/μL
Hemoglobin 9.7 g/dL
Platelets 184 K/μL
Bilirubin 0.5 mg/dL
Lymphocyte % 22 %
Monocyte % 8 %
3-month % change
Weight NA
Albumin NA
ALP NA
Hemoglobin NA
Platelets NA
Today's 6-Month Chance of Survival
Prediction based on starting next line of therapy
LIKELY
92 %
Population Survival
What does this prediction mean for me?
9 out of 10 patients like me are, in fact, alive after 6 months.
Recommended Action
Discuss personal values and care preferences for the future:
Consider meeting with Supportive Care and Social Work to review what the future might look like and complete advance care planning
Review 6-month outlook before next change in cancer treatment
Observed 6-Month Population Survival
Among similar patients with a 'Likely' chance of survival who started next line of therapy
Breast cancer (1st line of therapy in EHR)
Survival Chance: LIKELY
Observed survival & 95% CI (%)
Days of survival after starting next line of therapy
Survival of Patients Like Me
Based on similar patients who started next line of therapy
1st Line
92%
6-Month population survival (%)
May 2019
Date of therapy start

Zolnoori, M., Sridharan, S., Zolnour, A., Vergez, S., McDonald, M. V., Kostic, Z., ... & Topaz, M. (2024). Utilizing patient-nurse verbal communication in building risk identification models: the missing critical data stream in home healthcare. Journal of the American Medical Informatics Association, 31(2), 435-444.

- Objective: To measure the added value of integrating audio-recorded home healthcare patient-nurse verbal communication into a risk identification model built on home healthcare EHR data and clinical notes.
- Methods: This pilot study was conducted at one of the largest not-for-profit home healthcare agencies in the United States. We audio-recorded 126 patient-nurse encounters for 47 patients, out of which 8 patients experienced ED visits and hospitalization. The risk model was developed and tested iteratively using: (1) structured data from the Outcome and Assessment Information Set, (2) clinical notes, and (3) verbal communication features. We used various natural language processing methods to model the communication between patients and nurses.
- Results: Using a Support Vector Machine classifier, trained on the most informative features from OASIS, clinical notes, and verbal communication, we achieved an AUC-ROC = 99.68 and an F1-score = 94.12. By integrating verbal communication into the risk models, the F-1 score improved by 26%. The analysis revealed patients at high risk tended to interact more with risk-associated cues, exhibit more “sadness” and “anxiety,” and have extended periods of silence during conversation.
- Conclusions: This innovative study underscores the immense value of incorporating patient-nurse verbal communication in enhancing risk prediction models for hospitalizations and ED visits, suggesting the need for an evolved clinical workflow that integrates routine patient-nurse verbal communication recording into the medical record.

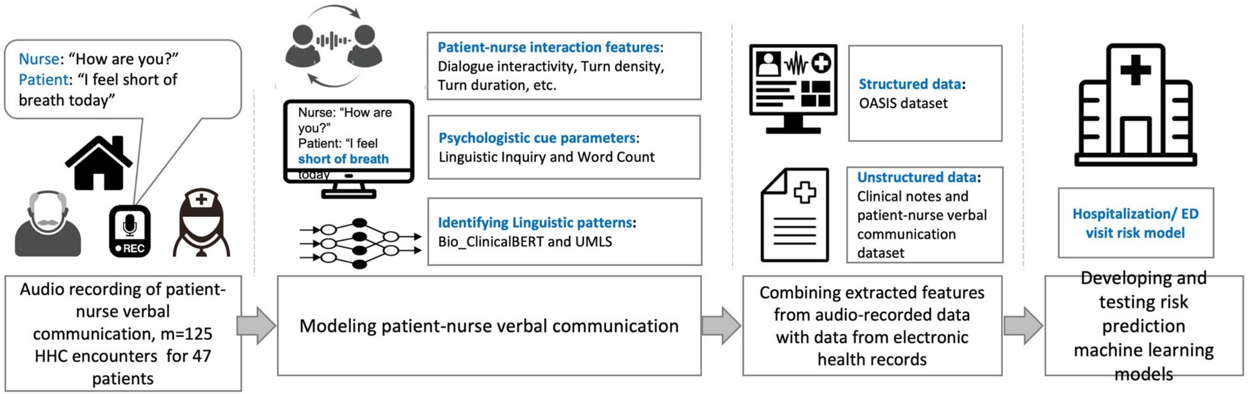
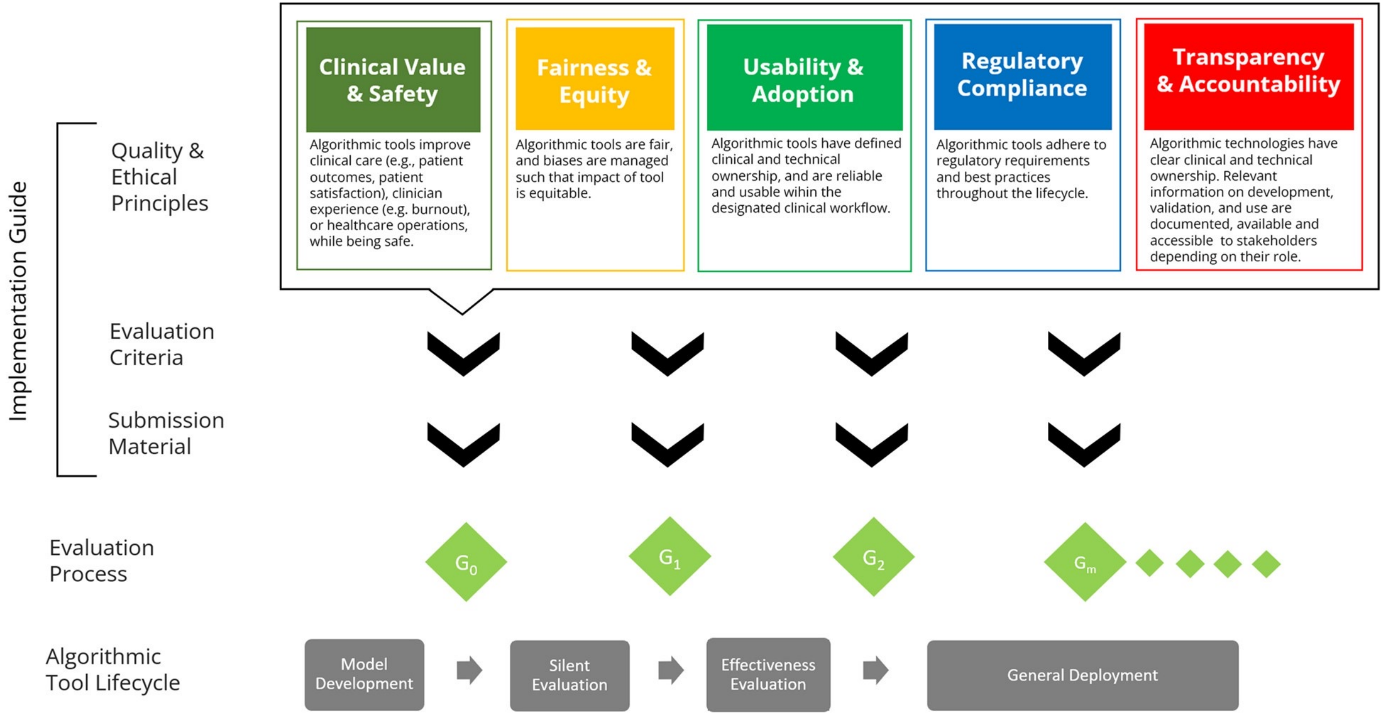


Table 3. Top-performing ML model on the OASIS dataset, as the baseline, along with the combination of clinical notes and audio-recorded patient-nurse verbal communication.

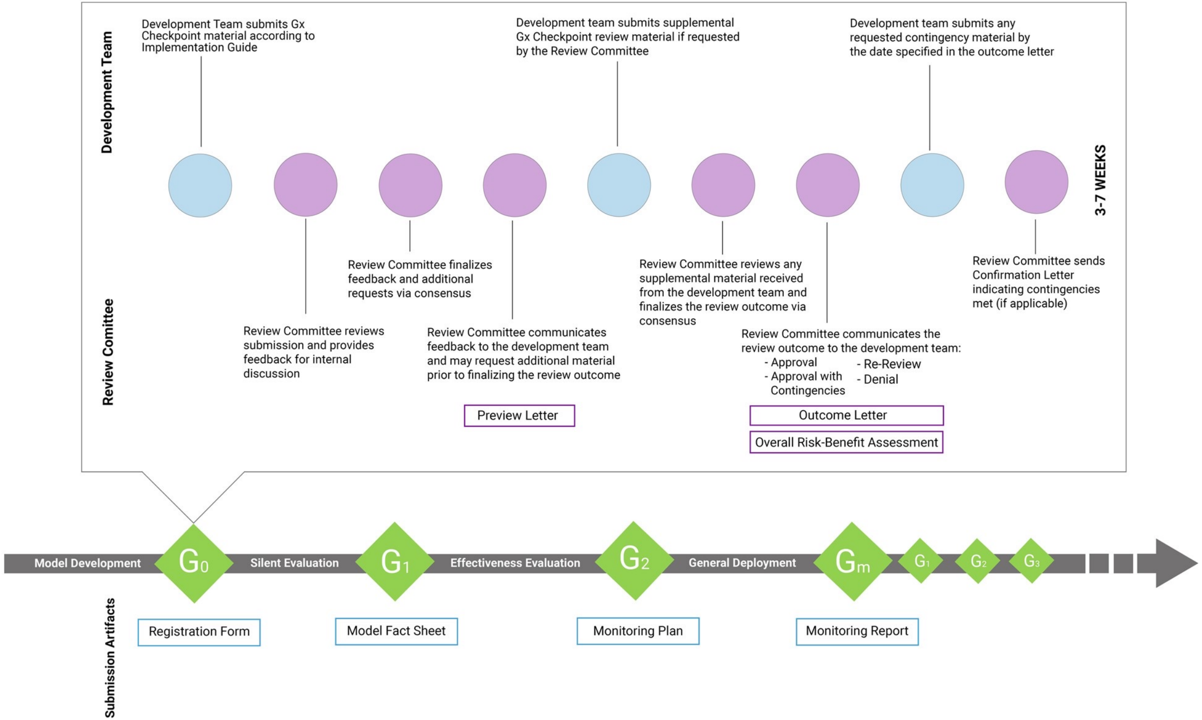
Feature generation methods	Best performing algorithm	AUC-ROC	F1-score
Sample: $N = 47$			
Baseline dataset			
OASIS dataset	XG-boost	67.63	48.01
Combination of OASIS and clinical notes and audio-recorded encounter for the most recent encounter			
OASIS+features extracted from clinical notes	SVM-RBF	79.17	73.68
OASIS+features extracted from clinical notes+features extracted from the patient's speech during an encounter	SVM-RBF	94.55	85.72
OASIS+features extracted from clinical notes+features extracted from the patient's speech during an encounter+the nurse's speech during an encounter	**SVM-RBF**	**96.15**	**87.5**
Combination of OASIS and clinical notes and audio-recorded encounters for all available encounters			
OASIS+features extracted from clinical notes	SVM-RBF	86.54	75.01
OASIS+features extracted from clinical notes + features extracted from the patient's speech during an encounter	XGB	96.79	87.5
OASIS+features extracted from clinical notes+features extracted from the patient's speech during an encounter+the nurse's speech during an encounter	**SVM-RBF**	**99.68**	**94.12**

Economou-Zavlanos, N. J., Bessias, S., Cary Jr, M. P., Bedoya, A. D., Goldstein, B. A., Jelovsek, J. E., ... & Poon, E. G. (2024). Translating ethical and quality principles for the effective, safe and fair development, deployment and use of artificial intelligence technologies in healthcare. Journal of the American Medical Informatics Association, 31(3), 705-713.

- Objective: The complexity and rapid pace of development of algorithmic technologies pose challenges for their regulation and oversight in healthcare settings. We sought to improve our institution's approach to evaluation and governance of algorithmic technologies used in clinical care and operations by creating an Implementation Guide that standardizes evaluation criteria so that local oversight is performed in an objective fashion.
- Methods: Building on a framework that applies key ethical and quality principles (clinical value and safety, fairness and equity, usability and adoption, transparency and accountability, and regulatory compliance), we created concrete guidelines for evaluating algorithmic technologies at our institution. Committee members analyzed notes from each breakout group to develop a list of action items.
- Results: An Implementation Guide articulates evaluation criteria used during review of algorithmic technologies and details what evidence supports the implementation of ethical and quality principles for trustworthy health AI. Application of the processes described in the Implementation Guide can lead to algorithms that are safer as well as more effective, fair, and equitable upon implementation, as illustrated through 4 examples of technologies at different phases of the algorithmic lifecycle that underwent evaluation at our academic medical center.
- Conclusions: We developed a scalable, adaptable framework for translating principles into evaluation criteria and specific requirements that support trustworthy implementation of algorithmic technologies in patient care and healthcare operations.



Clinical Value & Safety
Algorithmic tools improve clinical care (e.g., patient outcomes, patient satisfaction), clinician experience (e.g. burnout), or healthcare operations, while being safe.
Fairness & Equity
Algorithmic tools are fair, and biases are managed such that impact of tool is equitable.
Usability & Adoption
Algorithmic tools have defined clinical and technical ownership, and are reliable and usable wihin the designated clinical workflow.
Regulatory Compliance
Algorithmic tools adhere to regulatory requirements and best practices throughout the lifecycle.
Transparency & Accountability
Algorithmic technologies have clear clinical and technical ownership. Relevant information on development, validation, and use are documented, available and accessible to stakeholders depending on their role.
Implementation Guide
Quality & Ethical Principles
Evaluation Criteria
Submission Material
Evaluation Process
G_0
G_1
G_2
G_m
Algorithmic Tool Lifecycle
Model Development
Silent Evaluation
Effectiveness Evaluation
General Deployment



Development Team
Review Comittee
3-7 WEEKS
Development Team submits Gx Checkpoint material according to Implementation Guide
Development team submits supplemental Gx Checkpoint review material if requested by the Review Committee
Development team submits any requested contingency material by the date specified in the outcome letter
Review Committee reviews submission and provides feedback for internal discussion
Review Committee finalizes feedback and additional requests via consensus
Review Committee communicates feedback to the development team and may request additional material prior to finalizing the review outcome
Preview Letter
Review Committee reviews any supplemental material received from the development team and finalizes the review outcome via consensus
Review Committee communicates the review outcome to the development team:
- Approval
- Approval with Contingencies
- Re-Review
- Denial
Outcome Letter
Overall Risk-Benefit Assessment
Review Committee sends Confirmation Letter indicating contingencies met (if applicable)
Model Development
G_0
Silent Evaluation
G_1
Effectiveness Evaluation
G_2
General Deployment
G_m
G_1
G_2
G_3
Submission Artifacts
Registration Form
Model Fact Sheet
Monitoring Plan
Monitoring Report

Dlugatch, R., Georgieva, A., & Kerasidou, A. (2024). AI-driven decision support systems and epistemic reliance: a qualitative study on obstetricians' and midwives' perspectives on integrating AI-driven CTG into clinical decision making. BMC Medical Ethics, 25(1), 6.

- Objective: Given that AI-driven decision support systems (AI-DSS) are intended to assist in medical decision making, it is essential that clinicians are willing to incorporate AI-DSS into their practice. This study takes as a case study the use of AI-driven cardiotography (CTG), a type of AI-DSS, in the context of intrapartum care. Focusing on the perspectives of obstetricians and midwives regarding the ethical and trust-related issues of incorporating AI-driven tools in their practice, this paper explores the conditions that AI-driven CTG must fulfill for clinicians to feel justified in incorporating this assistive technology into their decision-making processes regarding interventions in labor.
- Methods: This study is based on semi-structured interviews conducted online with eight obstetricians and five midwives based in England. Participants were asked about their current decision-making processes about when to intervene in labor, how AI-driven CTG might enhance or disrupt this process, and what it would take for them to trust this kind of technology. Interviews were transcribed verbatim and analyzed with thematic analysis. NVivo software was used to organize thematic codes that recurred in interviews to identify the issues that mattered most to participants. Topics and themes that were repeated across interviews were identified to form the basis of the analysis and conclusions of this paper.
- Results: There were four major themes that emerged from our interviews with obstetricians and midwives regarding the conditions that AI-driven CTG must fulfill: (1) the importance of accurate and efficient risk assessments; (2) the capacity for personalization and individualized medicine; (3) the lack of significance regarding the type of institution that develops technology; and (4) the need for transparency in the development process.
- Conclusions: Accuracy, efficiency, personalization abilities, transparency, and clear evidence that it can improve outcomes are conditions that clinicians deem necessary for AI-DSS to meet in order to be considered reliable and therefore worthy of being incorporated into the decision-making process. Importantly, healthcare professionals considered themselves as the epistemic authorities in the clinical context and the bearers of responsibility for delivering appropriate care. Therefore, what mattered to them was being able to evaluate the reliability of AI-DSS on their own terms, and have confidence in implementing them in their practice.

Step 6 – Consultation

Feedback and Professional Input…