Patient and Public Involvement and Engagement (PPIE) in AI Healthcare Evaluations: A Patient-Centric Framework for Healthcare AI

Soumya Banerjee [1]
1 University of Cambridge, Cambridge, United Kingdom
sb2333@cam.ac.uk

Abstract

Artificial intelligence (AI) is transforming healthcare, yet its evaluation has remained narrowly focused on quantitative performance metrics such as accuracy and sensitivity. This technical emphasis overlooks the social dimensions of AI deployment: particularly the perspectives, values, and trust of patients and the public. In this paper, I argue that Patient and Public Involvement and Engagement (PPIE) must be systematically integrated into the evaluation pipeline for healthcare AI. I present case studies that demonstrate meaningful PPIE in practice, including patient-guided model design, public deliberation, and co-produced communication tools. I propose a framework for embedding PPIE at three critical stages: pre-deployment (design and development), deployment (real-time trust and feedback), and post-deployment (long-term impact on outcomes and equity). I also examine methodological challenges, ethical imperatives, and policy shifts supporting this integration. I suggest that incorporating patient and public perspectives is not only ethically necessary but also instrumental in building trustworthy, inclusive, and effective AI systems in healthcare. I conclude with concrete recommendations for researchers, developers, and evaluators seeking to centre human experience in the lifecycle of medical AI.

Keywords

Patient and public involvement and engagement, AI evaluations, AI in healthcare

Introduction

Advances inAI are transforming healthcare, from imaging and predictive modelling to clinical decision support. Patient and Public Involvement and Engagement (PPIE) is increasingly seen as essential to ensure these technologies are safe, equitable, and trustworthy. For example, involving patients with lived experience and co-designing models can build trust and set realistic expectations [1,2]. Similarly, Lammons report from a UK focus group that early PPIE is “crucial not only to safeguard patients but also to increase the chances of acceptance of AI” in clinical care [3]. A systematic review in mental health by Zidaru also highlights multiple ethical issues and calls for public engagement at every stage of AI from development to implementation stage [4]. Together, these sources point to a growing consensus that PPIE is not an afterthought, but a fundamental part of AI evaluation in healthcare.

Background: AI in Healthcare and the Need for PPIE

AI technologies in healthcare promise to improve diagnosis, treatment planning, and resource allocation. However, there is concern that data-driven AI can exacerbate biases or alienate

patients if developed in isolation [4]. As Zidaru observe, digital health datafication raises questions about whether AI truly empowers patients and whether patients have a meaningful say in designing these systems [4]. In this context, PPIE brings patient perspectives into the design and evaluation process.

PPIE is the meaningful, sustained partnership of patients, carers and members of the public with researchers, clinicians and decision-makers across the full lifecycle of research and technology development. It includes co-design, priority-setting, governance, dissemination and evaluation activities in which lay contributors actively shape questions, methods, outcomes and implementation, rather than serving only as research participants.

It has long been recognized that involving patients and the public improves research relevance and outcomes [5]. Adus cite broad evidence that patient engagement "can improve patient outcomes, quality of life, and safety, as well as decrease hospital admissions and health care costs" [5]. Hence, embedding PPIE in AI evaluations can help to ensure the technology aligns with patient values, addresses patient and publics' concerns (e.g. privacy, fairness), and gains social license for use.

In the remainder of this paper, I outline case studies and examples of PPIE in AI projects, and frameworks for PPIE in AI evaluations. I then present some methodological challenges in applying PPIE, ethical implications, and outline governance and policy considerations. Following on, I make recommendations on assessing the impact of PPIE and then suggest how to incorporate PPIE into AI evaluations. Finally, I make practical recommendations for how patients can contribute to and be involved in healthcare AI projects.

Case Studies and Examples of PPIE in AI Projects

Real-world projects illustrate how PPIE can be incorporated into AI research and evaluation. In one UK mental health analytics project, I described two participatory studies on secondary care mental health data [1]. In the first, a patient advisory group proposed a study on lithium's effect on kidney function; patients co-designed the linear model approach and were informed about uncertainties and trade-offs [1]. In another study, patients motivated the development of an explainable AI model for predicting mortality in schizophrenia, leading researchers to choose transparent, explainable AI models [1]. These cases show how patients can guide hypotheses and demand model explainability.

Lammons (England) convened public collaborators to discuss AI in healthcare [3]. Participants identified AI advantages such as reducing errors and improving access, but also raised challenges like data security concerns, algorithmic bias, and loss of human touch [2]. They stressed the importance of including diverse communities in development

to mitigate bias. The study itself exemplifies PPIE in practice: public collaborators co-analyzed focus group transcripts and used the findings to shape AI project plans [3].

In Japan, Katirai held workshops with a Japanese PPI panel to elicit expectations and concerns about AI in healthcare [7]. Participants envisioned improved care, efficiency, and reduced disparities, but also feared loss of autonomy, accountability gaps, and new inequities [7]. This balanced view reinforces the need for public deliberation.

On a broader scale, the EU's Mobilise-D consortium (digital mobility measurement, EU Horizon 2020) established structured PPIE across multiple countries and patient groups [4]. Patients were involved in topics like data sharing, algorithm usability, and visualization of complex data. By formalizing advisory boards and providing contributor training, the consortium ensured that lay perspectives shaped outcome measures for real-world health monitoring (EU Horizon 2020).

NHS England's AI in Health and Care Award embedded PPIE into its evaluation program [5]. An external advisory group of PPIE experts vetted study designs [9]. NHS evaluation domains explicitly asked how each AI technology fits clinical and patient needs and recommended analyzing impacts on health inequalities [9]. These initiatives demonstrate system-level integration of PPIE in the AI evaluation pipeline.

Each example underscores different facets of involvement. Collectively, they suggest that PPIE can and should be woven into AI evaluation at both project and policy levels.

Each project highlights a distinct role that patients and the public can play: from shaping research questions and co-designing analytic methods, to co-analysing data, demanding explainability, and advising on governance and equity. They therefore illustrate involvement at different stages (design, implementation, evaluation) and at different intensities (consultation, partnership, formal advisory roles). Together these examples map the varied functions, mechanisms and impacts of PPIE, showing it is multi-dimensional rather than a single, uniform activity.

Frameworks for PPIE in AI Evaluations

Based on case studies and the literature, a general framework for PPIE in AI can be outlined. First, *early involvement in concept and design* is critical: patients should be involved from problem definition through model selection. Focus groups and advisory panels can shape study hypotheses [5]. Second, *inclusive recruitment and representation* is essential: diverse demographics and underrepresented groups must be sought (e.g. through community outreach and social media) to ensure that all relevant perspectives inform the AI design [5].

Third, *education and empowerment* of contributors is needed: effective PPIE requires that participants understand AI basics. Training modules or visual tools can build contributors' confidence [1]. Fourth, *co-design and feedback loops* should be established: forming Research Advisory Groups (RAGs) or similar bodies allows patients to review data, models, and interim results iteratively [1]. I suggest that these groups enable patient insight to be applied throughout model development, not just at the end [1].

Fifth, *tools for communication and explainability* are required: use lay-friendly materials (infographics, heatmaps) to explain AI progress to participants, and solicit their input on interpretations [1]. For example, class-contrastive heatmaps were used to explain a mortality model to mental health patient advisors [1]. Sixth, *addressing diversity and inclusion* means designing AI evaluation to account for diverse patient needs and prevent inequities. PPIE frameworks should explicitly include tasks such as sensitivity analysis by subgroup and monitoring for bias.

Finally, *integration with governance and ethics review* should ensure alignment with regulatory oversight. McKay recommends integrating traditional PPI groups with institutional data committees and citizens' juries to create a unified governance model for medical AI [10]. In practice, this could mean involving patient representatives in data access decisions and on ethics boards for AI research [10]. Additionally, the framework calls for defining how PPIE itself will need to be evaluated (e.g. tracking implementation changes driven by input, or measuring stakeholder satisfaction) [5]. Future research must assess how PPIE influences AI project outcomes [5].

Embedding PPIE requires structured frameworks and best practices. Recent work proposes explicit co-design processes, such as forming Research Advisory Groups (RAGs) with patients and clinicians to guide data science projects [13,17]. For instance, teams set up RAG meetings every few months to **jointly define problems and build models**, walking patients through simple models and iterating based on their feedback [13,16]. Larger consortia (such as Mobilise-D) emphasise a consortium-wide PPIE strategy: research teams should **embed PPIE from the design phase**, educate all members on its importance, appoint a dedicated PPIE coordinator, and routinely refine the engagement approach as projects evolve [17]. Handbooks or plain-language overviews for patient contributors are recommended so that lay members can meaningfully participate [17]. To measure impact, projects are advised to document all patient recommendations and resulting changes (for example, tracking how patient input altered data collection or analysis) [17].

Conceptual frameworks from fields like design justice highlight that PPIE must actively promote inclusion and equity [18]. In practice, this means recruiting diverse patient advisors, explicitly including perspectives of historically marginalized groups, and building social accountability into evaluation [18]. A systematic review of PPI in AI-assisted mental health care similarly calls for "new methods of PPI at every stage, from concept design to the final review of technology," guided by design justice [19].

Beyond research studies, patient advocacy groups have issued recommendations. The European Patients' Forum (EPF) report emphasizes early, representative involvement [20]. EPF also calls for building *AI literacy* among patients and providers and for supporting developers to use unbiased data that respect patients' rights [20].

Governance and Guidance: The NHS AI Lab and NIHR are developing toolkits for involving patients in digital health projects [14]. The "AI in Health and Care Award" program mandates PPIE experts on advisory boards and co-produced evaluation plans [22]. International frameworks like the WHO's mERA checklist emphasize social impact evaluation [21].

Assessing PPIE Impact: Projects should record patient insights, resulting changes, and effects on outcomes [17]. Satisfaction and challenges should be documented using tools like the GRIPP2 checklist [23].

Methodological Challenges

Implementing PPIE in AI research presents several challenges. One challenge is *recruitment and diversity*: engaging a sufficiently representative sample of patients can be difficult. PPIE studies must reach beyond the willing volunteers (often the same few individuals) by using varied recruitment channels (clinics, community groups, online) and minimizing barriers (e.g. providing devices or transportation). *Technical complexity* is another challenge: AI concepts (algorithms, data privacy) are often opaque. Researchers must translate complex material into lay terms. This requires time and creative methods and may strain project resources.

A third challenge is *aligning expectations*: patients may have different priorities or understanding of AI capabilities. Facilitators must manage expectations and explain limitations. For instance, I had to explain to patients that lithium is a critical treatment despite potential kidney concerns [1]. Fourth, *measuring the impact of involvement* is difficult: quantifying how PPIE influences outcomes is not straightforward. Adus et al. note that dedicated research is needed to evaluate PPIE effectiveness [5]. Traditional clinical trial metrics do not capture these effects, so mixed qualitative–quantitative approaches may be required. Fifth, *time and resource constraints* pose a barrier: true co-production can be time-consuming. Allocating funds for PPIE (training, meetings, compensation) and integrating these tasks into project timelines must be planned. NHS AI evaluation reports suggest several years may be needed for multi-site studies [9].

Ethical Implications

PPIE intersects with several ethical issues in AI. One issue is *privacy and data use*: patients are concerned about who accesses their data. Borondy Kitts observes that individuals generally support sharing de-identified data for the public good, but worry about privacy breaches and data misuse [11]. PPIE can guide the creation of transparent data-sharing agreements and consent models [11].

Another issue is *bias and fairness*: AI systems trained on biased data can perpetuate discrimination. Underrepresented groups must be included in datasets and development to avoid unfair outcomes. PPIE can help identify potential bias early and advocate for corrective measures [11]. *Trust and transparency* are also ethical concerns: patients often distrust black-box algorithms. Building transparency through explainable models (e.g. class-contrastive heatmaps [1]) and patient education can mitigate this distrust. Borondy Kitts recommends that radiologists openly disclose AI use and explain its role to patients to maintain trust [11].

Autonomy and the human touch is another consideration: patients may worry that AI may depersonalize care. Borondy Kitts notes concern about losing human empathy [11]. Involving patients in AI design can ensure these technologies are positioned as support tools, not replacements, preserving the patient–provider relationship [11]. *Inclusion and justice* is yet another ethical dimension: Zidaru frames PPIE as an issue of design justice, emphasizing equality and diversity in digital health [4]. PPIE itself must be inclusive, or else AI may only serve privileged groups. For example, Katirai's panel warned that AI might introduce new disparities [7]. Governance should therefore require that PPIE reaches marginalized communities.

Overall, ethical AI development in healthcare demands involving those affected by the technology. PPIE serves as an ethical safeguard by giving patients a voice in risk–benefit tradeoffs and by aligning AI use with societal values [5,11].

Governance and Policy Considerations

National and institutional policies increasingly recognize the need for public involvement in AI governance. For instance, McKay et al. propose a multi-scale model in the UK where lay members of data-access committees, PPIE groups, and citizens' juries coordinate to oversee medical AI research [10]. This model leverages the complementary strengths of each approach to improve transparency and trust.

In practice, some programs have begun integrating PPIE into regulatory processes. The NHS AI Award mandated PPIE input at the evaluation design stage and explicitly recommended studying AI's impact on health inequalities [9]. Regulators like NICE and the MHRA (UK) now encourage patient representation in AI-related guideline development. On a broader scale, professional societies and patient advocacy groups are developing standards for AI. While the EU's AI Act and WHO's guidance do not yet mandate PPIE, they emphasize human oversight and risk management, which opens a role for patient perspectives in implementation.

National and institutional policies increasingly recognize the need for public involvement in AI governance. For instance, McKay propose a multi-scale model in the UK where lay members of data-access committees, PPIE groups, and citizens' juries coordinate to oversee medical AI research [10]. This model leverages complementary strengths of each approach to improve transparency and trust.

In practice, some programs have begun integrating PPIE in regulatory processes. The NHS AI Award mandated PPIE input at the evaluation design stage, and also recommended explicitly studying AI's impact on health inequalities [9]. Regulators like NICE and MHRA (in UK) now encourage patient representation in AI-related guideline development.

On a broader scale, professional societies and patient advocacy groups are developing standards for AI. While the EU's AI Act and WHO's guidance do not yet mandate PPIE, they emphasize human oversight and risk management [28] , opening a role for patient perspectives in implementation.

Assessing the Impact of PPIE

The literature underscores the expectation that PPIE improves outcomes, but evidence in the AI context is still emerging. General reviews have found that PPIE makes research more patient-centered and can improve health outcomes [5]. In AI specifically, one might measure impacts such as increased model acceptability, reduced errors from patient-informed criteria, or even better clinical outcomes due to AI tools that reflect patient needs.

To assess these, studies should build in PPIE evaluation metrics. This could include pre- and post-intervention surveys of patient trust, or analyses of how PPI input changed study design (e.g. adding bias checks). Future work must rigorously evaluate PPIE effectiveness in AI development. Developing validated scales or mixed-methods frameworks will be important, as they have been in other areas of health research.

Reframing AI Evaluation: Integrating Patient and Public Perspectives

Beyond Technical Metrics

AI systems in healthcare are predominantly evaluated using quantitative metrics such as accuracy, sensitivity, specificity, and area under the curve (AUC) [24]. While these metrics are essential for assessing performance, they often overlook the perspectives of patients and the public. Trust, acceptance, and perceived value of AI tools by end-users are critical factors influencing the successful implementation and adoption of these technologies [25,26].

The Role of Trust in AI Adoption

Studies have highlighted that patients' trust in AI varies depending on the application. Concerns about the lack of human touch, empathy, and understanding in AI-driven healthcare solutions have been consistently reported [27]. These concerns underscore the need to incorporate patient feedback into the evaluation process.

Incorporating PPIE into AI Evaluations

To address these gaps, I propose an **expanded AI evaluation framework** that integrates PPIE at multiple stages:

- *Pre-deployment*: Engaging patients in the design and development phases to ensure that AI tools align with patient needs and values.
- *Deployment*: Including patient feedback mechanisms to monitor real-time experiences and trust levels.
- *Post-deployment*: Assessing the long-term impact of AI tools on patient satisfaction, trust, and health outcomes.

This framework (shown in Figure 1) acknowledges that *technical performance alone does not guarantee AI success in healthcare settings*. By actively involving patients and the public, we can ensure that AI tools are not only effective, but also accepted and trusted by those they are designed to serve.

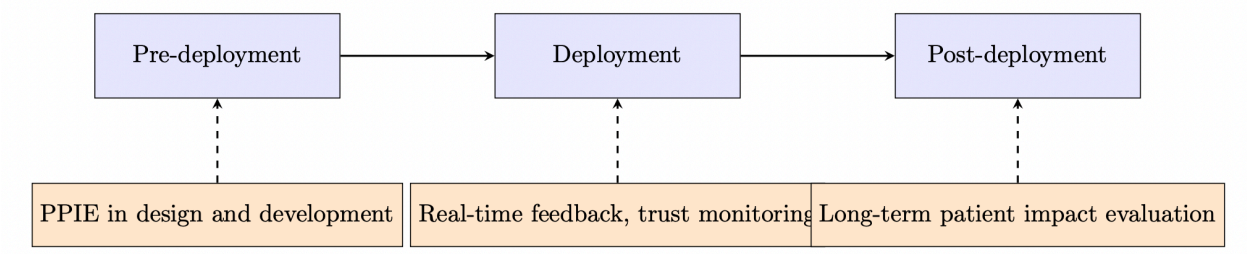

Figure 1. *An expanded AI evaluation framework integrating Patient and Public Involvement and Engagement (PPIE) across three key stages: pre-deployment, deployment, and post-deployment. Each stage includes distinct opportunities for involving patients and the public to ensure AI systems are not only technically effective but also aligned with user values, and beneficial in the long term.*

Recommendations and Best Practices

Key recommendations for patient and public involvement and engagement (PPIE) in AI evaluation emphasise partnership, capacity-building, inclusivity, transparency, and resourcing. AI projects should be co-designed with patients, embedding them within research teams and advisory panels from the earliest stages to ensure that lived experience shapes priorities and design decisions [1,8]. Contributors must be educated and empowered to participate meaningfully: providing accessible training on AI fundamentals and on the specific healthcare problem enables patients to engage with technical material and contribute substantive critique and insight [5,12]. Active steps to ensure diversity (recruiting across age, ethnicity, socio-economic status, and levels of digital literacy) are essential to capture a range of perspectives and avoid reinforcing existing inequities [3,5]. Transparency is also critical; using explainable AI methods, openly communicating results and limitations, and clearly describing data use foster trust and support informed scrutiny [1,11]. Evaluation frameworks should explicitly incorporate PPIE into their criteria by assessing patient acceptability, user experience, and equity alongside technical performance [9]. Finally, meaningful engagement requires planning and funding: grant proposals and evaluation protocols must allocate specific time and budget for PPIE activities and for assessing their impact.

These practices align with ethical imperatives and evidence-based governance; when followed, they can make healthcare AI evaluations more patient-centred and increase the likelihood that resulting tools are effective, fair, and accepted by those who will use them.

Practical Recommendations for PPIE in Healthcare AI

Here I outline practical recommendations for how patients and the public can contribute to healthcare AI projects. For example, *initial scoping workshops* can be convened where patients, family members, and caregivers (especially those with lived experience of a condition) join data scientists and clinicians to map out real-world care pathways. Activities in such workshops might include *walkthroughs of current workflows*, where patients describe step by step what happens when they present with symptoms. For example, a diabetic patient might explain how they self-monitor, how alerts are sent to clinicians, and where miscommunications have occurred in the past. Another activity could be *"pain point" brainstorming*, in which participants list scenarios where technology

or human error could cause harm. For instance, a patient might recall when a lab result was misreported, leading to delayed intervention (highlighting where an AI tool could either compound mistakes or potentially catch them earlier). The outcome of these workshops would be a prioritized list of potential safety hazards categorized by likelihood and severity.

Ongoing co-development meetings can then be established. For example, *regular check-ins* might involve monthly "AI Safety Rounds" in which developers demo a prototype of the model's decision logic or user interface (UI) and patients provide feedback. An example activity could present an early UI mockup of a radiology report annotated by AI and ask patients to interpret what each color or icon means. If multiple patients misunderstand a red/yellow/green urgency scheme, developers can adjust labels or add explanatory tooltips before any code is finalized. These check-ins catch misinterpretations early and surface hidden biases: for instance, patients might say, "I am not sure why the algorithm flags my age 75 as 'high risk' for pneumonia when I'm physically active and have not smoked in decades" prompting developers to reconsider how they weigh age versus comorbidities.

Real-world risk mapping can be led by patients. In *"risk detective" exercises*, small teams of diverse patients can be asked to imagine worst-case scenarios. For example, they might ask: "What happens if the AI fails to recognize your congenital heart condition because it's rare?" They can then trace downstream effects (delayed diagnosis, hospitalization, malpractice concerns). The deliverable would be a patient-authored risk register labeling each potential AI safety threat with three attributes. *Likelihood* asks how often patients believe this could happen. *Impact* asks how severe it would be personally. *Detectability* asks whether they would notice the failure (e.g. odd user-interface behaviour).

Ethical Governance and Oversight. Patient representatives should be meaningfully included in institutional AI governance, not merely as token seats. Clear role definitions and empowerment are important: for example, patient members could have voting rights on all agenda items, so their views carry weight, and rotating the "patient chair" role among patient members ensures their perspective drives meeting agendas. Providing dedicated onboarding and early access to meeting materials (white papers, performance metrics, incident logs) allows patient representatives to prepare questions rather than react passively during discussions.

Ensuring diversity of lived experience on governance bodies is also vital. Multiple patient seats can be designated: for example, one seat for a chronic disease patient (someone navigating multiple specialist appointments), one for a caregiver representative (family member supporting someone with dementia or advanced cancer), and one for a community advocate (representing under-represented groups like non-English speakers or

low-income communities). Term limits and staggered rotation (e.g. 2-year terms) ensure new patient voices cycle in periodically, preventing stagnation or burnout.

Training and Support for Patient Members. To understand technical aspects well enough to ask probing questions, patient members may need dedicated training. Foundational workshops on AI concepts could include an "AI 101 for Patients" curriculum covering basics (e.g. machine learning vs rule-based systems), performance metrics (accuracy, sensitivity, specificity, AUC), and common failure modes (overfitting, data drift, adversarial inputs). Each module should include a live Q&A with data scientists where patient members can ask clarifying questions (e.g. "What does data drift look like if the regional patient population changes?"). Concise one-page glossaries or cheat sheets explaining terms like "confusion matrix," "training vs validation sets," and "explainability vs interpretability" can be provided; patient members can refer to these when technical terms arise in meetings. Mentorship pairing is another strategy: pair each patient representative with a data scientist mentor who can answer follow-up questions outside formal meetings.

Inclusive AI Literacy Programs. Inclusive AI literacy begins by recognising that basic digital skills are a safety issue: when communities lack digital literacy, misinterpreting an AI output (such as misunderstanding a risk score) can cause real harm. To prevent that, patients and local advocates should co-design outreach and training that meets people where they are. One practical approach is to run short, community-based "Digital Health Days" in familiar places such as libraries, faith centres, or senior centres; these two-hour workshops teach concrete skills (how to read AI notifications, who to contact if something is unclear, and common pitfalls like assuming a low risk score means you are healthy right now) and can point learners to curated resources (for example, materials available at https://github.com/neelsoumya/ai_outreach)

Managing patient expectations of AI. Managing expectations about AI in healthcare begins with recognising how overblown headlines. For example, "AI will replace radiologists!" can create unrealistic hopes or unnecessary fears.

A practical first step is identifying and cataloguing those misconceptions through "myth-busting" workshops. I suggest inviting a diverse group of patients to work in pairs and list headline statements they have seen or heard (such as "AI reads my scans in five seconds," "AI never makes mistakes," or "AI can prescribe medication on its own"). For each claim, ask participants to explain why it might be misleading: for example, suggestions that imply "no human oversight" or that "all hospitals have it, so I do not need to double-check." The expected outcome is a prioritised list of five to ten myths ordered by how frequently patients encounter them in social media, TV news, or word-of-mouth.

The next stage is co-creating “reality check” messaging with patients so communications strike the right balance. Patients can help craft framing statements that emphasise both the value and the limits of AI: for instance, “AI aids, it does not replace: AI helps your healthcare team by giving them extra insights, but your doctor still makes the final decision,” and “AI has strengths and limits: AI may spot patterns humans miss, but it can also be wrong if it has never seen a rare condition.” Finally, I suggest testing the tone of these messages in small patient groups, comparing versions that stress “AI is a powerful helper” with versions that highlight “AI sometimes makes mistakes,” to see which comes across as balanced (rather than either overly reassuring or unduly discouraging).

Global and cultural contexts. A complementary strategy is a train-the-trainer programme: recruit tech-comfortable volunteers from the community (high-school students, retirees, or other local champions) then give them deeper, modular explanations of AI so they can serve as Digital Health Ambassadors. These ambassadors run smaller, neighbourhood sessions that are sensitive to local culture and language, and that reinforce learning through trusted, repeated contact.

Materials must also be multilingual and culturally tailored rather than verbatim translations. Translating infographics and videos into a community’s primary language (Somali, for instance) is necessary but not sufficient; examples and metaphors should be adapted so they make sense locally (e.g., swapping a “baseball pitch” analogy for “kicking a football” where that resonates more). That kind of contextualisation improves comprehension and trust.

When we shift focus to global and cultural contexts, especially low- and middle-income countries (LMICs), different safety priorities emerge because of infrastructure limitations and cultural attitudes toward automation. Engaging local patient communities uncovers regional risks that one-size-fits-all regulations can miss. For example, intermittent electricity in some rural districts makes cloud-dependent medical devices unacceptable unless they include offline fail-safe modes or manual fallback controls; patients and frontline health workers often highlight exactly these operational vulnerabilities. Cultural perspectives matter too: in places where faith healers are central to care, people may fear that digital tools will supplant traditional counselling or spiritual support, so co-designers should frame AI recommendations as tools that support clinicians and community practices rather than replace them.

Participatory research in LMIC settings is crucial for surfacing these realities. For example, spending several weeks in a rural clinic to observe patient interactions, conduct semi-structured interviews, and document communication patterns may reveal preferences (such as a strong inclination for verbal explanations from trusted elders rather than written or app-based alerts) that should shape both how tools communicate and how training is delivered. Likewise, adapting AI models trained in high-income settings is essential: an

HIV-risk prediction algorithm developed for an urban UK population may misclassify patients in sub-Saharan Africa if local cofactors (for example, malaria coinfections) are not incorporated; community health workers and patients are the best sources for identifying those cofactors.

Policy and regulatory variability across countries adds another layer of concern. Many LMICs currently lack AI-specific medical regulations, so patient groups must often advocate for baseline safeguards: mandatory adverse event reporting to national databases, prohibition of unvalidated AI tools in critical care, and simple operational requirements such as paper-based backup protocols and power-outage plans. Cross-border collaborations can accelerate capacity building: patient advocacy groups in the UK and Europe can share templates for informed consent and study governance while LMIC partners contribute low-resource considerations that make those templates workable in practice.

Finally, equity in AI deployment must be actively protected to prevent "AI colonialism" [31]. Patients and local advocates should be empowered to challenge deployments of externally developed tools that lack local validation (such as skin-lesion detectors trained predominantly on lighter skin tones that perform poorly on darker skin) and to demand evidence of representative performance. Where appropriate, communities can insist on collective data governance models, for example community cooperatives or tribal oversight, to ensure that data collected for AI are not commodified without consent and that benefits accrue locally. Together, these approaches of co-designed literacy, culturally adapted materials, participatory research, adaptive policy, and community stewardship of data may help create safer, fairer AI systems that respect both local realities and global obligations.

Conclusion

Patient and public involvement is a key pillar for successful AI in healthcare. It addresses ethical concerns, enhances trust, and improves the relevance of technical evaluations. Recent studies across multiple countries consistently advocate embedding PPIE from concept to dissemination. I have outlined case studies, challenges, and an AI evaluation framework based on this emergent literature. Going forward, it will be important to empirically study PPIE models and to develop formal guidelines tailored to AI evaluations. By integrating PPIE into the lifecycle of AI research, we can help ensure these powerful technologies truly serve patients and communities.

Integrating PPIE into AI evaluations provides a more holistic understanding of an AI tool's effectiveness. It ensures that evaluations capture not just technical performance but also the social and ethical dimensions critical to patient care. Future research should focus on developing

standardized methods for incorporating patient and public feedback into AI evaluation frameworks.

Contemporary discussions about AI often swing between two extremes: boundless techno-optimism that promises easy fixes, and dystopian fear that imagines catastrophe. Involving patients and the public in building, testing and learning about AI helps us find a middle way: techno-realism [29]. When people who will actually use or be affected by systems help shape them, we get clearer ideas of what AI can and cannot do, what trade-offs are acceptable, and which harms need guarding against. That grounded perspective leads to better choices about design, priorities, and rules, and it builds trust. Ultimately, PPIE may help us build AI that is truly trustworthy.

In summary, I advocate for an evaluation framework that transcends technical benchmarks to prioritize the patient experience. By integrating PPIE across the *pre-deployment*, *deployment*, and *post-deployment* stages, we can ensure that AI development is continuously informed by patient needs and values. This human-centric approach is critical for fostering the trust required to translate AI innovation into meaningful health outcomes.

Declarations

Acknowledgements

We acknowledge the help and support of the Accelerate technical team.

Funding statement

This work was funded by an Accelerate Programme for Scientific Discovery Fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The views expressed are those of the authors and not necessarily those of the funders.

Conflicts of interests

All authors declare they have no conflicts of interest to disclose.

Ethics

No ethics approval was necessary.

Data accessibility

This study does not generate any clinical data.

References

1. Banerjee S, Alsop P, Jones L, Cardinal RN (2022) Patient and public involvement to build trust in artificial intelligence: A framework, tools, and case studies. Patterns 3: 100506.

2. Banerjee S, Griffiths S (2023) Involving patients in artificial intelligence research to build trustworthy systems. AI & Society 1: 1-3

3. Lammons W, Silkens M, Hunter J, Shah S, Stavropoulou C (2023) Centering public perceptions on translating ai into clinical practice: Patient and public involvement and engagement consultation focus group study. J Med Internet Res 25: e49303.

4. Zidaru T, Morrow EM, Stockley R (2021) Ensuring patient and public involvement in the transition to ai-assisted mental health care: A systematic scoping review and agenda for design justice. Health Expectations 24: 1072–1124.

5. Adus S, Macklin J, Pinto A (2023) Exploring patient perspectives on how they can and should be engaged in the development of artificial intelligence (ai) applications in health care. BMC Health Services Research 23.

6. Banerjee S, Lio P, Jones PB, Cardinal RN (2021) A class-contrastive human-interpretable machine learning approach to predict mortality in severe mental illness. npj Schizophrenia 7: 1-13.

7. Katirai A, Yamamoto BA, Kogetsu A, Kato K (2023) Perspectives on artificial intelligence in healthcare from a patient and public involvement panel in Japan: An exploratory study. Frontiers in Digital Health 5: 1229308.

8. Keogh A, Mc Ardle R, Diaconu MG, Ammour N, Arnera V, et al. (2023) Mobilizing patient and public involvement in the development of real-world digital technology solutions: Tutorial. J Med Internet Res 25: e44206.

9. NHS England (2024). Planning and implementing real-world artificial intelligence (AI) evaluations: lessons from the AI in health and care award. URL https://www.england.nhs.uk/long-read/ planning-and-implementing-real-world-ai-evaluations-lessons-from-the-ai-in-health-and-care-awa

10. McKay F, Williams BJ, Prestwich G, Treanor D, Hallowell N (2022) Public governance of medical artificial intelligence research in the uk: An integrated multi-scale model. Research Involvement and Engagement 8.

11. Borondy Kitts A (2023) Patient perspectives on artificial intelligence in radiology. J Am Coll Radiol 20: 863–867.

12. Teodorowski P, Gleason K, Gregory JJ, Martin M, Punjabi R, et al. (2023) Participatory evaluation of the process of co-producing resources for the public on data science and artificial intelligence. Research Involvement and Engagement 9.

13. Banerjee S (2021) Emergent rules of computation in the universe lead to life and consciousness: a computational framework for consciousness. Interdisciplinary Description of Complex Systems 19: 31-41.

14. Aguirre A, et al (2022) Public involvement in health AI: current state and future directions. Ada Lovelace Institute Working Paper.

15. Sato Y, et al (2022) Patient perspectives on AI in healthcare: A Japanese workshop study. BMJ Health & Care Informatics.

16. Banerjee S, Ghose J, Banerjee T, Banerjee K (2023) Beauty of life in dynamical systems: Philosophical musings and resources for students. Journal of Humanistic Mathematics 13.

17. Mobilise-D Consortium MD (2022) Mobilise-D PPIE strategy guide. Project Internal Report. https://mobilise-d.eu/patient-and-public-involvement-and-engagement-structures/

18. Izquierdo L, et al (2021) Design justice and artificial intelligence: A framework for inclusion. AI & Society . 19. Stark A, et al (2023) Participatory methods in mental health AI: a systematic review. Digital Mental Health Review.

19. Stark A, et al (2023) Participatory methods in mental health ai: a systematic review. Digital Mental Health Review.

20. European Patient Forum (2022) Empowering patient voices in digital health innovation. Policy Report.

21. Ada Lovelace Institute (2022) Algorithmic impact assessment in healthcare: lessons and frameworks. https://www.adalovelaceinstitute.org/project/algorithmic-impact-assessment-healthcare/

22. (2022) Evaluation of bias and inequality in NHS AI projects. https://www.england.nhs.uk/long-read/planning-and-implementing-real-world-ai-evaluations-lessons-from-the-ai-in-health-and-care-award/

23. Staniszewska S, et al (2017). Gripp2 reporting checklist: Guidance for reporting involvement of patients and the public. Available at: https://www.equator-network.org/reportingguidelines/gripp2/.

24. Gareth J, Daniela W, Trevor H, Robert T (2017) Introduction to Statistical Learning with Applications in R. Springer. URL http://www-bcf.usc.edu/~gareth/ISL/.

25. Richardson JP, Curtis S, Smith C, Pacyna J, Zhu X, et al. (2022) A framework for examining patient attitudes regarding applications of artificial intelligence in healthcare. Digital Health 8: 20552076221089084.

26. Woodcock C, Mittelstadt B, Busbridge D, Blank G (2022) The impact of explanations on layperson trust in artificial intelligence-driven symptom checker apps: Experimental study. arXiv preprint arXiv:220213444 .

27. Witkowski K, Dougherty R, Neely S (2023) Public perceptions of artificial intelligence in healthcare: ethical concerns and opportunities for patient-centered care. BMC Medical Ethics.

28. World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: WHO

29. Vardi, M. Y. (2026). Techno-Optimism, Techno-Pessimism, and Techno-Realism. Communications of the ACM, 69(1), 5

30. https://github.com/neelsoumya/ai_outreach, URL accessed December 2025

31. Couldry, N., & Mejias, U. A. (2019). The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism. Stanford University Press