# Goals of Our Reading Group

- Get to know each other (the most important aspect of grad school)
- Stay up to date on EHR research - it keeps growing!
- Practice reading and discussing papers
- Practice giving presentations
- Eventually you'll present your research

PubMed_Timeline_Results_by_Year

| Search query: electronic health records | |
|---|---|
| **Year** | Count |
| **2020** | 6847 |
| **2019** | 6543 |
| **2018** | 5898 |
| **2017** | 5412 |
| **2016** | 5205 |
| **2015** | 4958 |
| **2014** | 4497 |
| **2013** | 3550 |
| **2012** | 2468 |
| **2011** | 2148 |
| **2010** | 1826 |

# Structure of Our Reading Group

- Meet every three weeks (Wednesdays at 9:30)
- Everyone reads a paper(s)
- One person presents and leads the discussion
- Upload papers to a shared drive
- We'll update the schedule for the Fall semester

# DO NO HARM: A ROADMAP FOR RESPONSIBLE MACHINE LEARNING FOR HEALTH CARE

WIENS ET AL

# *Do no harm: a roadmap for responsible machine learning for health care*

- **By Jenna Wiens, Suchi Saria[2], Mark Sendak, Marzyeh Ghassemi, Vincent X. Liu, Finale Doshi-Velez, Kenneth Jung, Katherine Heller, David Kale, Mohammed Saeed, Pilar N. Ossorio, Sonoo Thadaney-Israni and Anna Goldenberg**
- ***Nature medicine*. 2019 Sep 25(9): 1337-40.**

# Focus of the Paper

- Interest in machine learning (ML) in medicine is growing, but there are few examples of deployment in patient care
- Put into context the key issues that interdisciplinary teams need to consider in translating ML-based interventions to health care

# Interest in ML in Medicine Is Growing

WILL KNIGHT BUSINESS 04.13.2021 08:00 AM

**Microsoft Makes a $16 Billion Entry Into Health Care AI**

The company plans to buy Nuance, a speech-recognition firm that grasps the specialized language of medicine—tech that won't be easy for others to replicate.

The New York Times

*Google to Store and Analyze Millions of Health Records*

CNBC MARKETS BUSINESS INVESTING TECH POLITICS CNBC TV

TECH

**Amazon confirms it's working on a project to mine patient records and more accurately diagnose diseases**

# Interest in ML in Medicine Is Growing


- Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H. and Wang, Y., 2017. Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology*, *2*(4).
- Yu, K.H., Beam, A.L. and Kohane, I.S., 2018. Artificial intelligence in healthcare. *Nature biomedical engineering*, *2*(10), pp.719-731
- Panch, T., Szolovits, P. and Atun, R., 2018. Artificial intelligence, machine learning and health systems. *Journal of global health*, *8*(2).
- Char, D.S., Shah, N.H. and Magnus, D., 2018. Implementing machine learning in health care—addressing ethical challenges. *The New England journal of medicine*, *378*(11), p.981.
- Beam, A.L. and Kohane, I.S., 2018. Big data and machine learning in health care. *Jama*, *319*(13), pp.1317-1318.
- Rajkomar, A., Dean, J. and Kohane, I., 2019. Machine learning in medicine. *New England Journal of Medicine*, *380*(14), pp.1347-1358.
- Davenport, T. and Kalakota, R., 2019. The potential for artificial intelligence in healthcare. *Future healthcare journal*, *6*(2), p.94.
- Kelly, C.J., Karthikesalingam, A., Suleyman, M., Corrado, G. and King, D., 2019. Key challenges for delivering clinical impact with artificial intelligence. *BMC medicine*, *17*(1), pp.1-9.
- Nagendran, M., Chen, Y., Lovejoy, C.A., Gordon, A.C., Komorowski, M., Harvey, H., Topol, E.J., Ioannidis, J.P., Collins, G.S. and Maruthappu, M., 2020. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *bmj*, *368*.
- Matheny, M.E., Whicher, D. and Israni, S.T., 2020. Artificial intelligence in health care: a report from the National Academy of Medicine. *Jama*, *323*(6), pp.509-510.
- Sendak, M.P., D'Arcy, J., Kashyap, S., Gao, M., Nichols, M., Corey, K., Ratliff, W. and Balu, S., 2020. A path for translation of machine learning products into healthcare delivery. *EMJ Innov*, *10*, pp.19-00172.
- Ghassemi, M., Naumann, T., Schulam, P., Beam, A.L., Chen, I.Y. and Ranganath, R., 2020. A review of challenges and opportunities in machine learning for health. *AMIA Summits on Translational Science Proceedings*, *2020*, p.191.

# Few Examples of Deployment in Clinical Care

editorial

## All eyes are on AI

Artificial intelligence may eventually help diagnose eye conditions and the risk of cardiovascular disease, solely from retinal images.

- ***"The only machine-learning application currently approved by the United States Food and Drug Administration (FDA), Arterys Cardio DL, segments MRI images of the heart, and an algorithm for the detection of diabetic retinopathy (the IDx-DR system) is under expedited FDA review."***

**All eyes are on AI. *Nat Biomed Eng* 2, 139 (2018).**

# Example: Diagnostic Imaging

ARTIFICIAL INTELLIGENCE

## Google and Verily reveal algorithm for diabetic eye disease screening

The organizations have announced that the first real-world clinical use of the algorithm is happening at the Aravind Eye Hospital in Madurai, India.

By ERIN DIETSCHE

CHI 2020 Paper | CHI 2020, April 25–30, 2020, Honolulu, HI, USA

## A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy

# Example: Diagnostic Imaging

ARTIFICIAL INTELLIGENCE

## Google and Verily reveal algorithm for diabetic eye disease screening

The organizations have announced that the first real-world clinical use of the algorithm is happening at the Aravind Eye Hospital in Madurai, India.

By ERIN DIETSCHE

CHI 2020 Paper — CHI 2020, April 25–30, 2020, Honolulu, HI, USA

## A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy

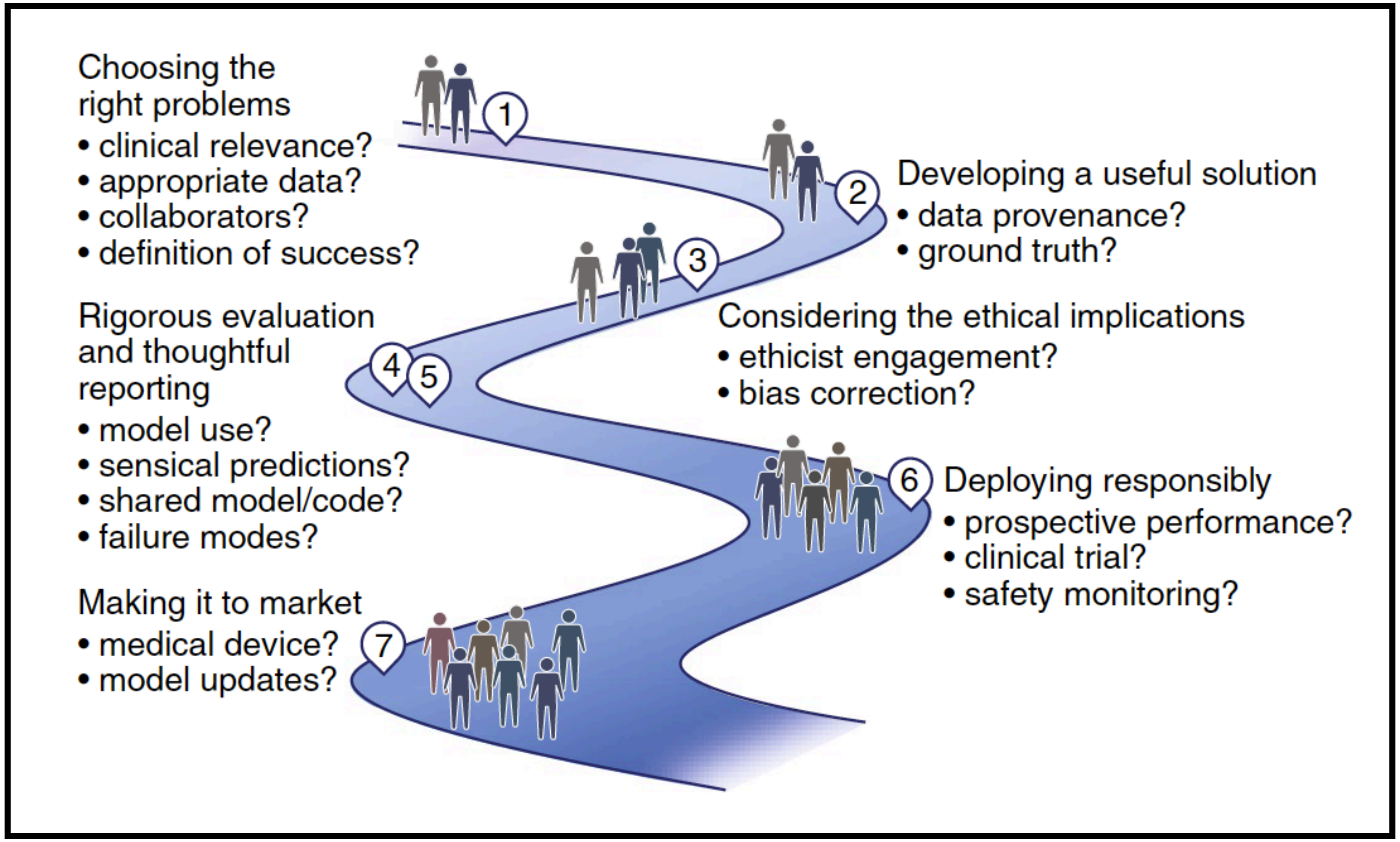
Artificial intelligence / Machine learning

## Google's medical AI was super accurate in a lab. Real life was a different story.

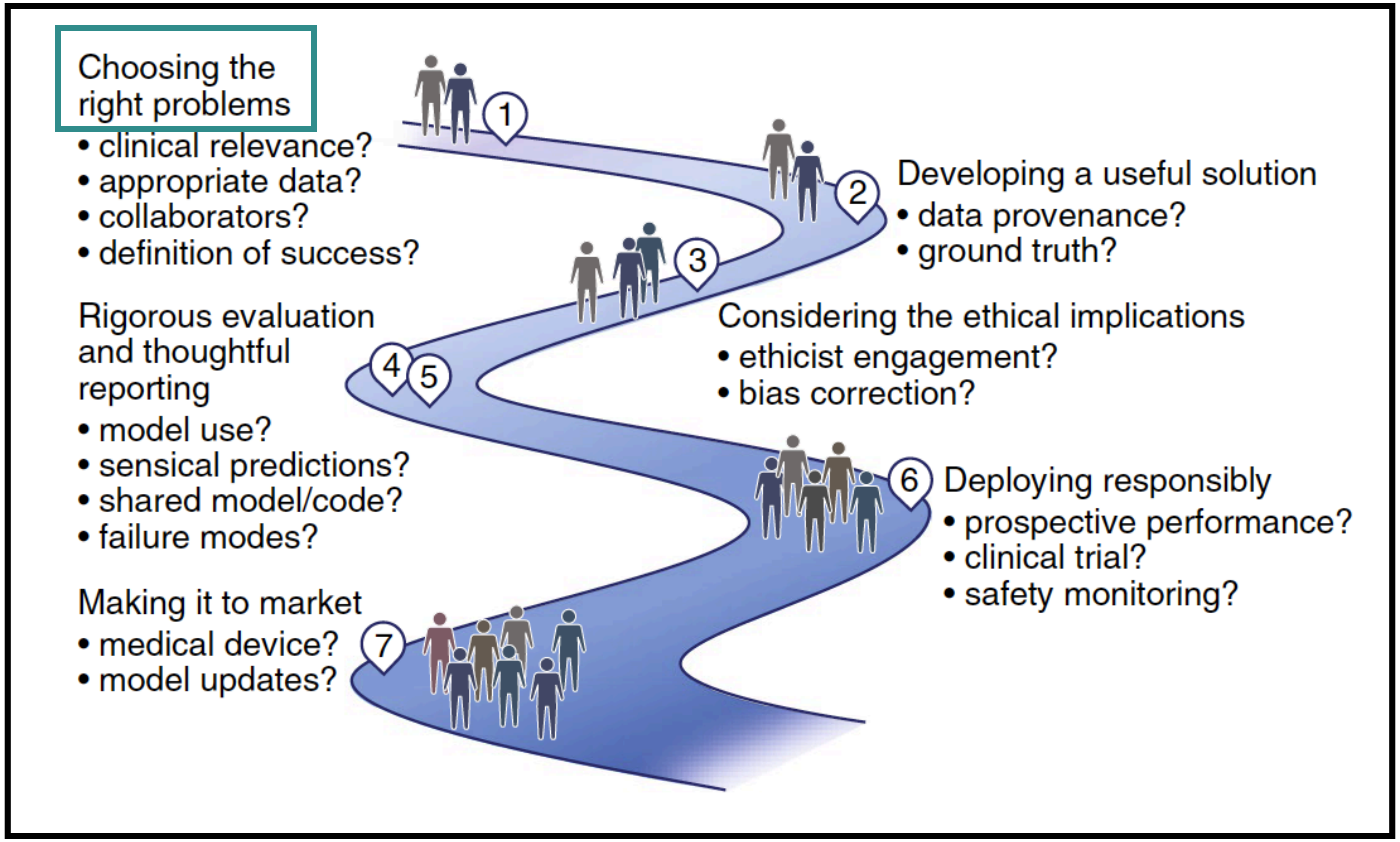
If AI is really going to make a difference to patients we need to know how it works when real humans get their hands on it, in real situations.

by **Will Douglas Heaven** — April 27, 2020

# Proposed Framework



# Proposed Framework



# Choosing the Right Problems

- Conflict between clinically relevant questions and available data
- Ex. Predicting in-hospitality mortality within 48 hrs of admission to ICU
    - Mortality is annotated so there is sufficient training data to build a model with a high AUC
    - Accuracy can be due to learning patterns of end of life care, which may not be explicitly charted
    - This oversight results in a model that doesn’t convey anything new to clinicians

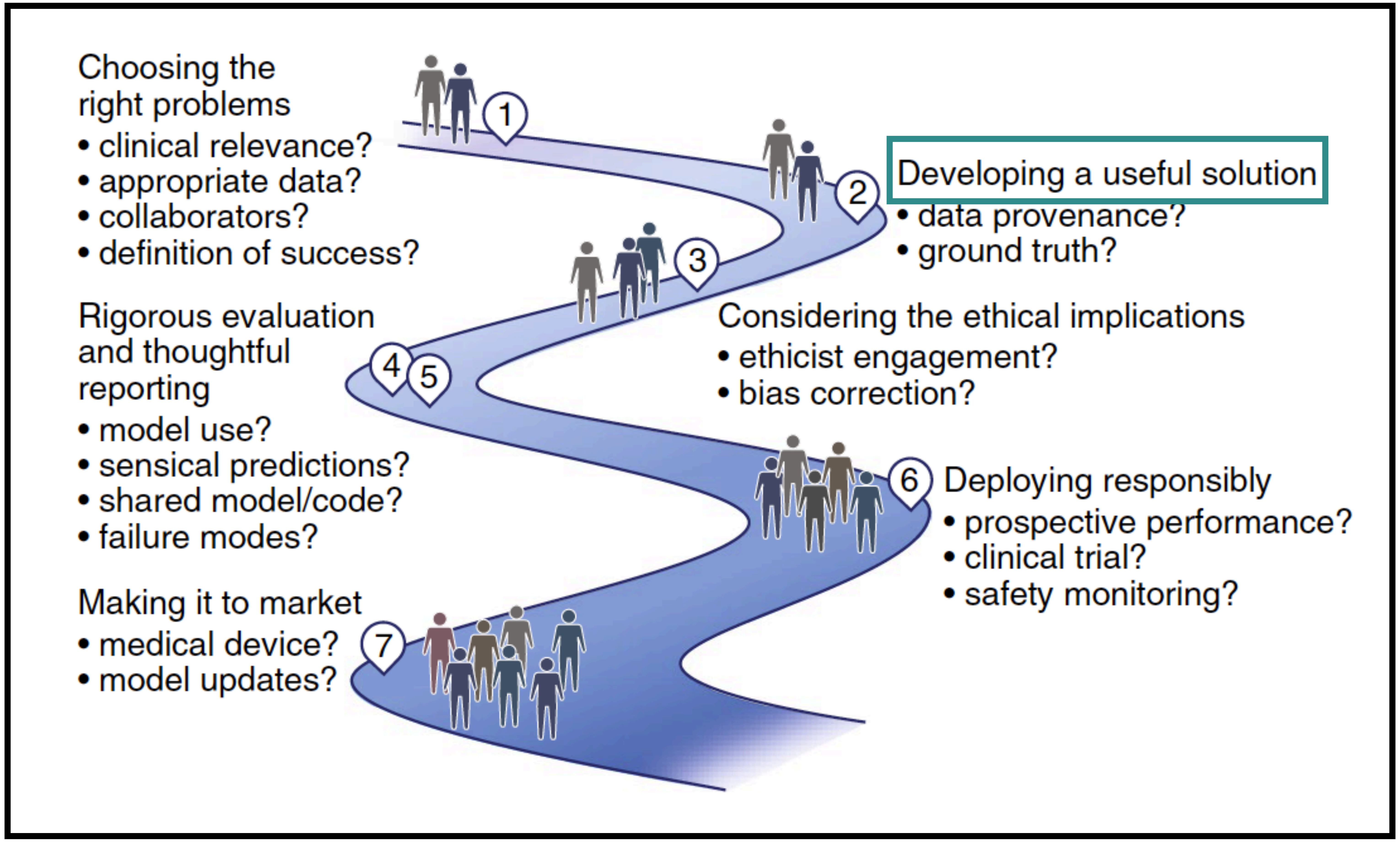
# Importance of an Interdisciplinary Team

- **Engage stakeholders who know the data, the problem, and the use case early on**

**Table 1 | Interdisciplinary teams may consist of stakeholders from different categories**

| Stakeholder categories | Examples |
|---|---|
| Knowledge experts | • Clinical experts<br>• ML researchers<br>• Health information and technology experts<br>• Implementation experts |
| Decision-makers | • Hospital administrators<br>• Institutional leadership<br>• Regulatory agencies<br>• State and federal government |
| Users | • Nurses<br>• Physicians<br>• Laboratory technicians<br>• Patients<br>• Friends and family (framily) |

# Proposed Framework



# Developing a Useful Solution

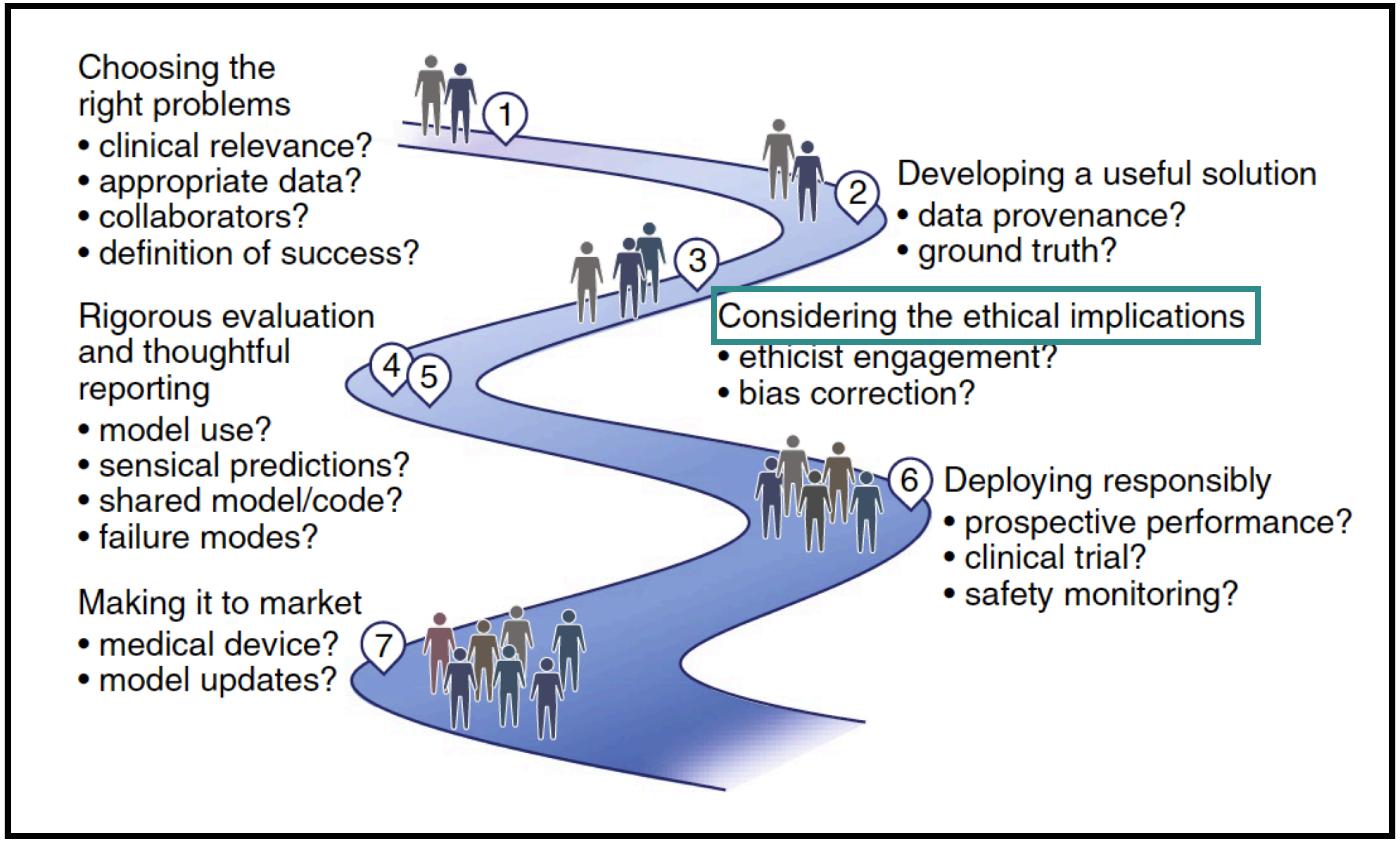
- Understand the data to determine if they are appropriate for the question

# Developing a Useful Solution

- **Understand the data to determine if they are appropriate for the question**

1. **How, when, why the inputs and outputs were collected**
    - **Ex. Cost vs. health state, ICD codes vs. diseases diagnoses**
2. **Differences in collection (e.g. across departments or health systems)**
    - **Ex. Domain specific codes such as Fyler codes**
3. **Who is represented**
    - **Ex. Patients who have access to care and visit the same institution**
4. **Subtle biases**
    - **Ex. Provider specific patterns, time-dependent patterns**

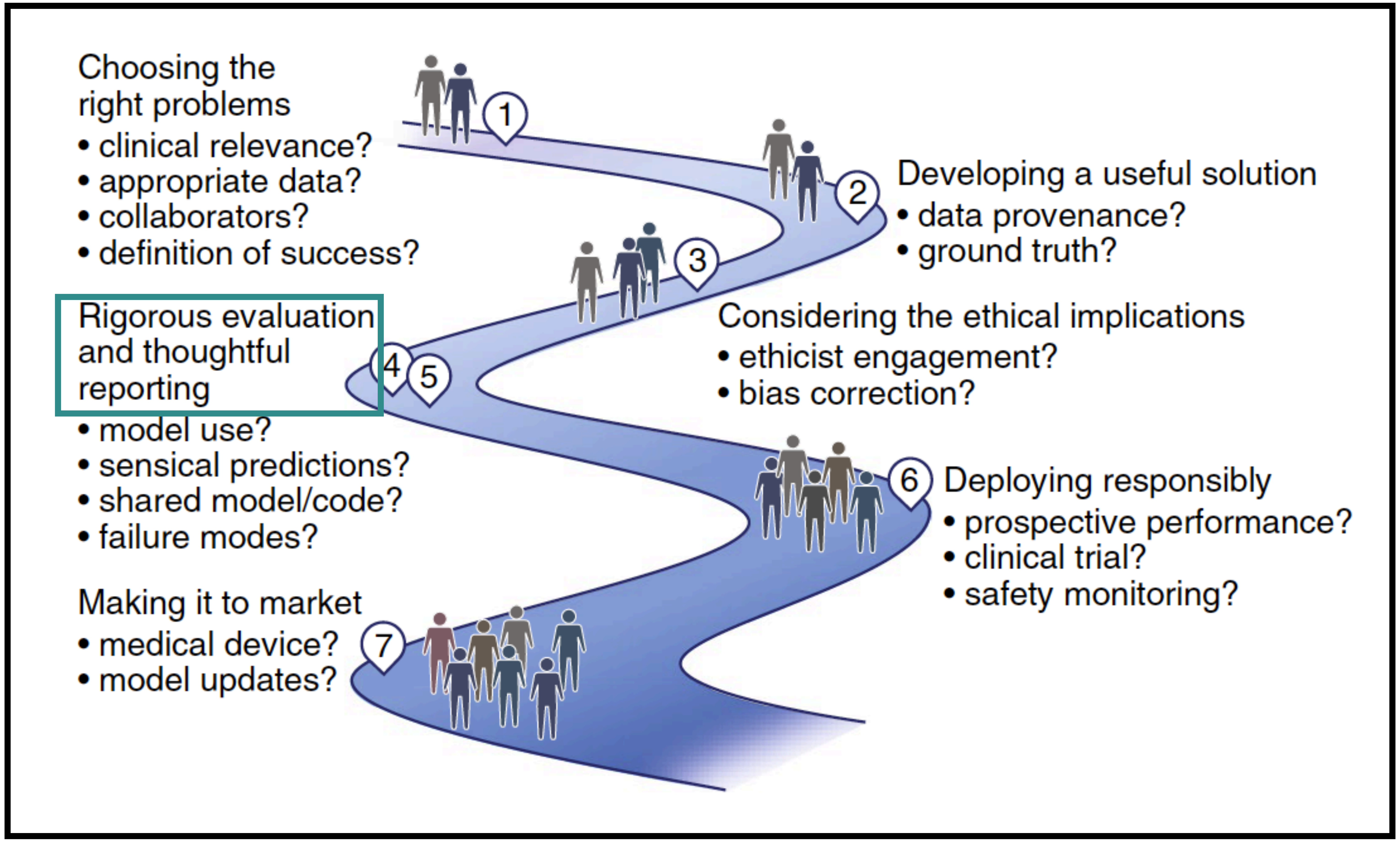
# Proposed Framework



# Considering Ethical Implications

- Health data is prone to race, sex, and other biases
  - <u>Ex</u>. A model predicting who should have surgery could favour those with access to care, ability to take time off work, etc.
- The question of interest may pose ethical concerns
  - <u>Ex</u>. Imputation of missing data such as HIV and smoking status
- Requires that ethicists, ML experts, and other stakeholders work together

# Proposed Framework



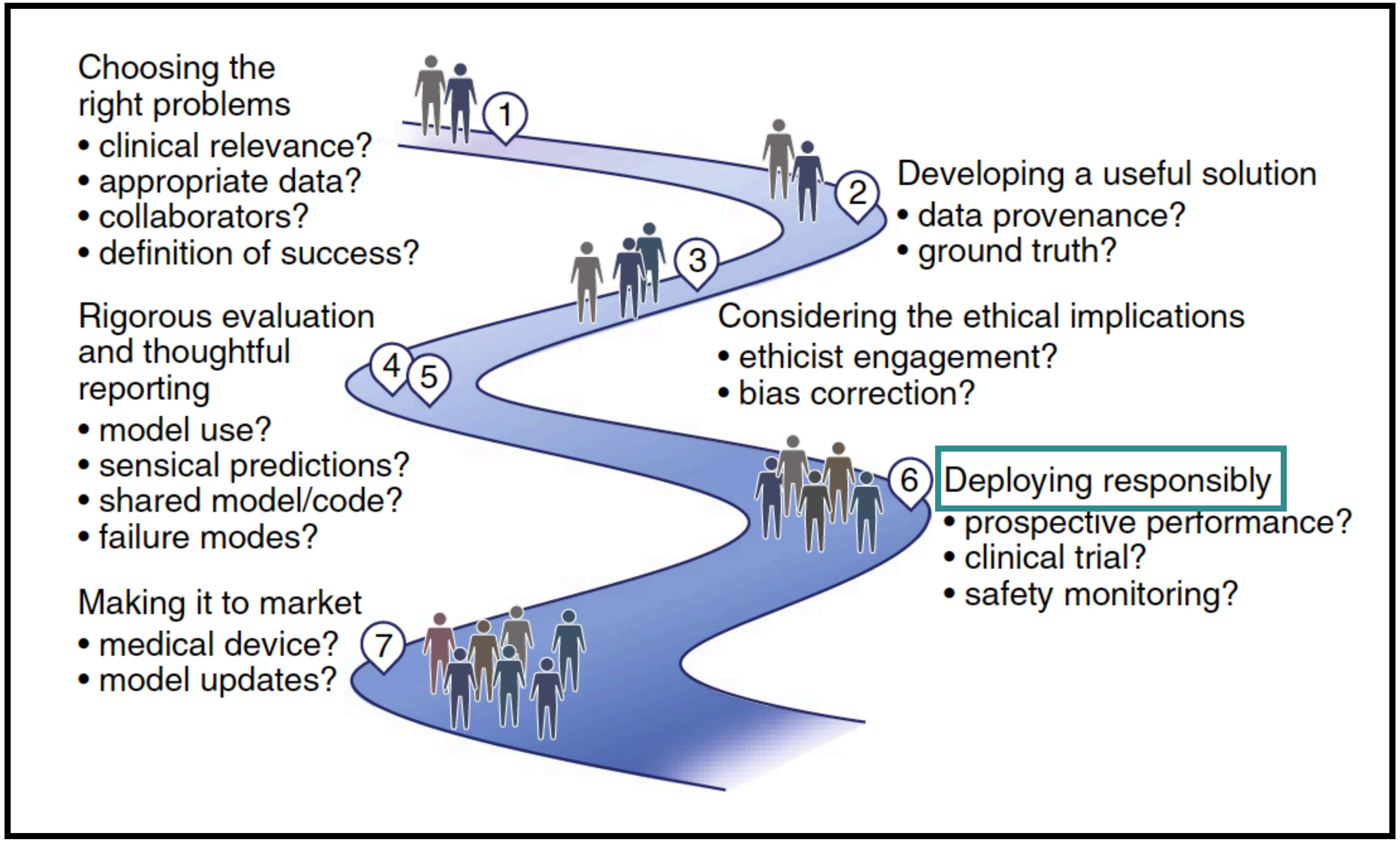
# Rigorously Evaluating the Model

1. Avoiding label leakage
   - Ex. Appropriate data splitting
2. Evaluating and reporting the scope in which the model is likely to succeed
   - Ex. Sepsis-prediction models in adults vs. children
3. Reporting clinically relevant evaluation metrics
   - Ex. AUC, number needed to benefit
4. Assessing qualitative measures
   - Ex. Model predicting an outcome depends on its treatment

# Thoughtfully Reporting Results

- Clear descriptions of the source of the data, participants, outcomes and predictors
- Model description (but, be careful)
  - Ex. Variables associated with increased risk of healthcare-associated infection may be protective in another hospital
- Share code, packages and inputs used to generate the reported results
  - Ex. Complex models can be sensitive to the random seed
- Go beyond predictive performance when comparing models
  - Ex. Trade-off between accuracy, complexity, computational burden

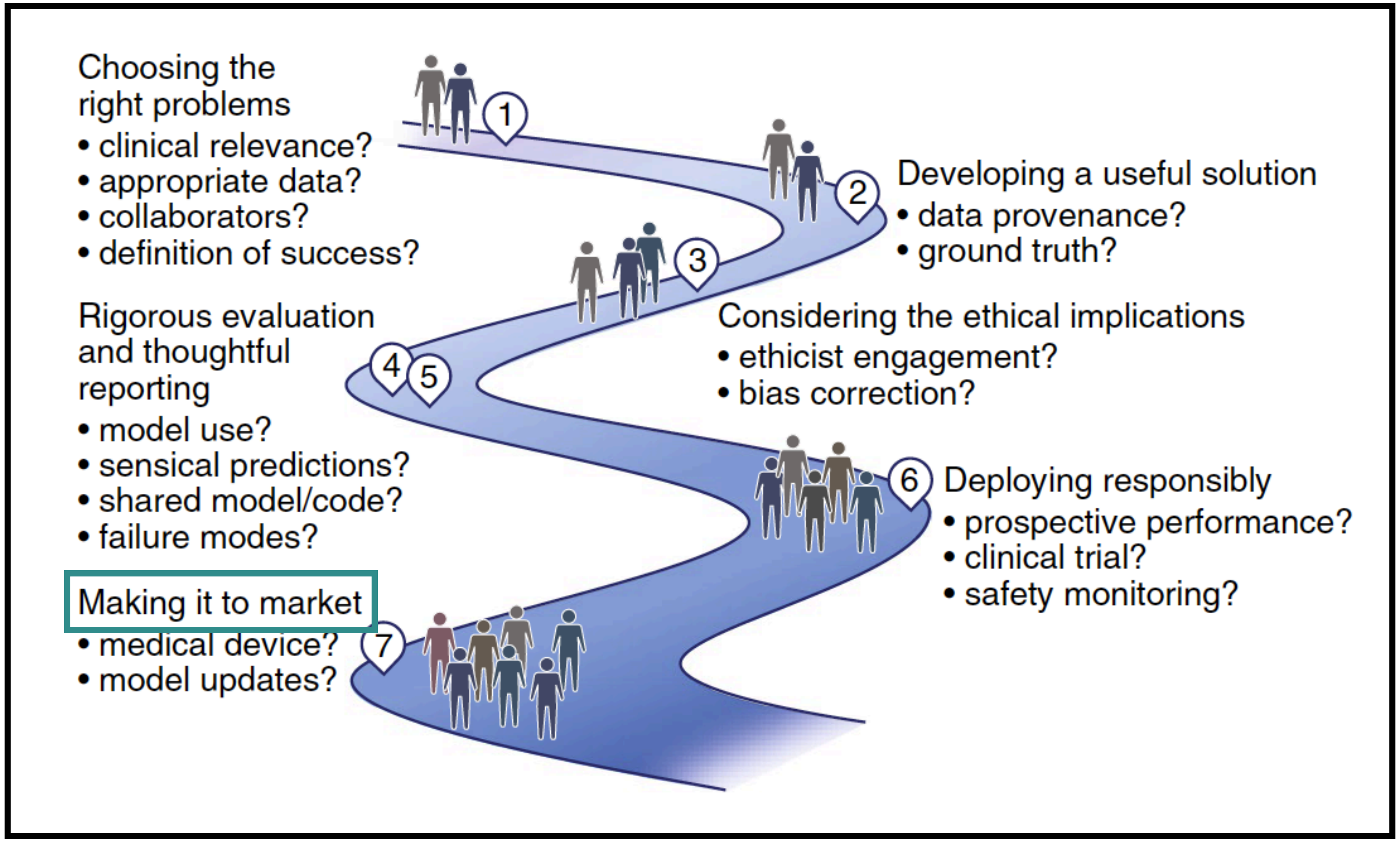
# Proposed Framework



# Deploying Responsibly

- Prospective validation
  - Testing in *silent mode*: Review predictions without acting upon them
- Assessment of efficacy in clinical studies
  - Randomization is hard in real-world settings
  - Introducing the model over time may serve as an alternative
- Integration into clinical workflow
- Frequent monitoring

# Proposed Framework



# Making It to Market

- ML tools must be validated with the government-required regulatory steps in mind
  - Ex. In the US, some types of medical software or clinical decision support systems are considered and regulated as medical devices
- Ability to interrogate predictions
  -

# Conclusions

- There is still a long way to go
- Outstanding questions
  - How much accuracy is sufficient for deployment?
  - What level of model transparency is required?
  - Do we understand when the model outputs are likely to be unreliable and therefore should not be trusted?