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

Key Questions When Reading a Paper

1. Motivation
    1. What is the problem being solved?
    2. Why is it important?
2. Approach
    1. What methods were used and why?
    2. What datasets were used and why?
3. Results
    1. How well did the approach solve the problem with simulated and/or real data?
    2. How did the approach compare to other solutions?
    3. What conclusions can be drawn?
4. Contributions
    1. How does this work compare to previous work?
    2. What makes the paper “new” or “novel"?
5. Limitations
    1. What might the issues be in applying the approach to another dataset or problem?
    2. What results are missing from the paper?
    3. Are the authors' conclusions well-informed?

Schedule for Our Reading Group

- June 16: ?
- July 7: ?
- July 29 (Thursday): ?
- August 18: ?
- Sign-up here

Papers: EHR Research

- EHR Phenotyping + Risk Prediction
  - Goldstein, Benjamin A., et al. "Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review." *Journal of the American Medical Informatics Association* 24.1 (2017): 198-208.
  - Banda, Juan M., et al. "Advances in electronic phenotyping: from rule-based definitions to machine learning models." *Annual review of biomedical data science* 1 (2018): 53-68.
  - Zhang, Yichi, et al. "High-throughput phenotyping with electronic medical record data using a common semi-supervised approach (PheCAP)." *Nature protocols* 14.12 (2019): 3426-3444.
  - Weng, Chunhua, Nigam H. Shah, and George Hripcsak. "Deep phenotyping: embracing complexity and temporality—towards scalability, portability, and interoperability." *Journal of biomedical informatics* 105 (2020): 103433.
  - Tong, Jiayi, et al. "An efficient distributed algorithm with application to COVID-19 data from heterogeneous clinical sites." *medRxiv* (2020).
- EHR + Genetics
  - Kohane, Isaac S. "Using electronic health records to drive discovery in disease genomics." *Nature Reviews Genetics* 12.6 (2011): 417-428.
  - Li, Ruowang, et al. "Electronic health records and polygenic risk scores for predicting disease risk." *Nature Reviews Genetics* 21.8 (2020): 493-502.
  - Beesley, Lauren J., et al. "The emerging landscape of health research based on biobanks linked to electronic health records: Existing resources, statistical challenges, and potential opportunities." *Statistics in medicine* 39.6 (2020): 773-800.
- General
  - Kohane, Isaac S., et al. "What every reader should know about studies using electronic health record data but may be afraid to ask." *Journal of medical Internet research* 23.3 (2021).

Papers: ML in Healthcare

- Panch, Trishan, Heather Mattie, and Leo Anthony Celi. "The "inconvenient truth" about AI in healthcare." *NPJ digital medicine* 2.1 (2019): 1-3.
- Ghassemi, Marzyeh, et al. "Practical guidance on artificial intelligence for health-care data." *The Lancet Digital Health* 1.4 (2019): e157-e159.
- Beam, Andrew L., Arjun K. Manrai, and Marzyeh Ghassemi. "Challenges to the reproducibility of machine learning models in health care." *Jama* 323.4 (2020): 305-306.
- Chen, Irene Y., et al. "Ethical Machine Learning in Healthcare." *Annual Review of Biomedical Data Science* 4 (2020).

Papers: Clinical Informatics + ML

- Huang, Jing, et al. "PIE: A prior knowledge guided integrated likelihood estimation method for bias reduction in association studies using electronic health records data." *Journal of the American Medical Informatics Association* 25.3 (2018): 345-352.
- Duan, Rui, et al. "ODAL: A one-shot distributed algorithm to perform logistic regressions on electronic health records data from multiple clinical sites." *BIOCOMPUTING 2019: Proceedings of the Pacific Symposium*. 2018.
- Rajkomar, Alvin, et al. "Scalable and accurate deep learning with electronic health records." *NPJ Digital Medicine* 1.1 (2018): 1-10. (A lot written about this one...)
- Li, Ruowang, Yong Chen, and Jason H. Moore. "Integration of genetic and clinical information to improve imputation of data missing from electronic health records." *Journal of the American Medical Informatics Association* 26.10 (2019): 1056-1063.
- Ning, Wenxin, et al. "Feature extraction for phenotyping from semantic and knowledge resources." *Journal of biomedical informatics* 91 (2019): 103122.
- Zhang, Lingjiao, et al. "A maximum likelihood approach to electronic health record phenotyping using positive and unlabeled patients." *Journal of the American Medical Informatics Association* 27.1 (2020): 119-126.

Papers: Statistical Methodology

- Model Evaluation
  - Le, Wang, et al. "Evaluating risk-prediction models using data from electronic health records." *The annals of applied statistics* 10.1 (2016): 286.
  - Zhang, Lingjiao, et al. "Testing calibration of phenotyping models using positive-only electronic health record data." *Biostatistics* (2021).
- Measurement Error and Misclassification
  - Hubbard, Rebecca A., et al. "Accounting for differential error in time-to-event analyses using imperfect electronic health record-derived endpoints." *New Advances in Statistics and Data Science*. Springer, Cham, 2017. 239-255.
  - Beesley, Lauren J., and Bhramar Mukherjee. "Statistical inference for association studies using electronic health records: handling both selection bias and outcome misclassification." *Biometrics* (2020).
- Missing Data
  - Thaweethai, Tanayott, et al. "Robust inference when combining inverse-probability weighting and multiple imputation to address missing data with application to an electronic health records-based study of bariatric surgery." *The Annals of Applied Statistics* 15.1 (2021): 126-147.
- Risk Prediction
  - Hubbard, Rebecca A., et al. "A Bayesian latent class approach for EHR-based phenotyping." *Statistics in medicine* 38.1 (2019): 74-87.
  - Tan, W. Katherine, and Patrick J. Heagerty. "Surrogate-guided sampling designs for classification of rare outcomes from electronic medical records data." *arXiv preprint arXiv:1904.00412* (2019).
- Sampling Bias
  - Beesley, Lauren J., Lars G. Fritsche, and Bhramar Mukherjee. "An analytic framework for exploring sampling and observation process biases in genome and phenome-wide association studies using electronic health records." *Statistics in medicine* 39.14 (2020): 1965-1979.
- Estimating Treatment Effects
  - Wu, Peng, Donglin Zeng, and Yuanjia Wang. "Matched learning for optimizing individualized treatment strategies using electronic health records." *Journal of the American Statistical Association* 115.529 (2020): 380-392.

Papers: Statistical Theory

- **Weakly supervised learning, Measurement Error + Misclassification**
  - **Chakrabortty, Abhishek, et al. "Surrogate Aided Unsupervised Recovery of Sparse Signals in Single Index Models for Binary Outcomes." *arXiv preprint arXiv:1701.05230* (2017).**
  - **Kallus, Nathan, and Xiaojie Mao. "On the role of surrogates in the efficient estimation of treatment effects with limited outcome data." *arXiv preprint arXiv:2003.12408* (2020).**
  - **Duan, Rui, et al. "On the global identifiability of logistic regression models with misclassified outcomes." *arXiv preprint arXiv:2103.12846* (2021).**
- **Semi-supervised learning**
  - **Chakrabortty, Abhishek, and Tianxi Cai. "Efficient and adaptive linear regression in semi-supervised settings." *Annals of Statistics* 46.4 (2018): 1541-1572.**
  - **Zhang, Yuqian, Abhishek Chakrabortty, and Jelena Bradic. "Double Robust Semi-Supervised Inference for the Mean: Selection Bias under MAR Labeling with Decaying Overlap." *arXiv preprint arXiv:2104.06667* (2021).**
- **Estimation of Treatment Effects**
  - **Cai, Tianxi, Tony Cai, and Zijian Guo. "Optimal statistical inference for individualized treatment effects in high-dimensional models." *arXiv preprint arXiv:1904.12891* (2019).**

Papers: Non-EHR Stats Papers

- Li, Ker-Chau, and Naihua Duan. "Regression analysis under link violation." *The Annals of Statistics* (1989): 1009-1052.
- Tibshirani, Robert. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1 (1996): 267-288.
- Zou, Hui. "The adaptive lasso and its oracle properties." *Journal of the American statistical association* 101.476 (2006): 1418-1429.
- Varin, Cristiano, Nancy Reid, and David Firth. "An overview of composite likelihood methods." *Statistica Sinica* (2011): 5-42.

- More to come....

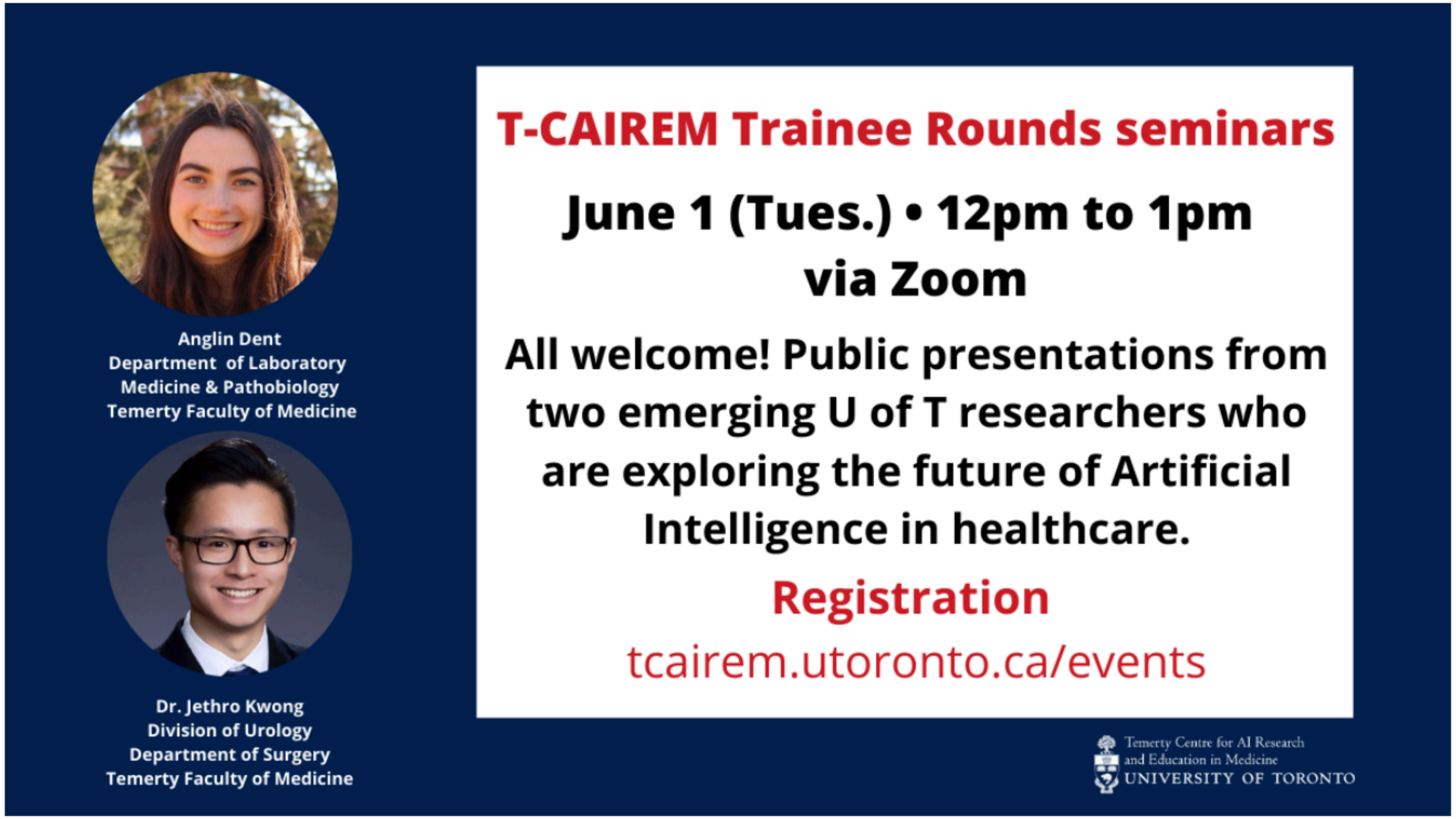
Papers: Clinical NLP

- Alsentzer, Emily, et al. "Publicly available clinical BERT embeddings." *arXiv preprint arXiv:1904.03323* (2019).
- Beam, Andrew L., et al. "Clinical concept embeddings learned from massive sources of multimodal medical data." *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2020*. 2019. Agrawal, Monica, et al.
- "Robust Benchmarking for Machine Learning of Clinical Entity Extraction." *Machine Learning for Healthcare Conference*. PMLR, 2020.

Papers: Other

- Kass, Robert E., et al. "Ten simple rules for effective statistical practice." (2016): e1004961.
- Meng, Xiao-Li. "Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election." *Annals of Applied Statistics* 12.2 (2018): 685-726.
- Gebru, Timnit, et al. "Datasheets for datasets." *arXiv preprint arXiv:1803.09010* (2018).
- Hutchinson, Ben, and Margaret Mitchell. "50 years of test (un) fairness: Lessons for machine learning." *Proceedings of the Conference on Fairness, Accountability, and Transparency*. 2019.
- Mitchell, Margaret, et al. "Model cards for model reporting." Proceedings of the conference on fairness, accountability, and transparency. 2019.
- Gelman, Andrew, and Aki Vehtari. "What are the most important statistical ideas of the past 50 years?." *arXiv preprint arXiv:2012.00174* (2020).
- Apley, Daniel W., and Jingyu Zhu. "Visualizing the effects of predictor variables in black box supervised learning models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82.4 (2020): 1059-1086.

Upcoming Talk



Another Journal Club

- **EXITE AI Journal Club led by Dr. Andrew Pinto**
- **Once per month**
- **Let me know if you’d like to join**
- **June 2: Siyue on EHR phenotyping!**

EXITE-AI Journal Club

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

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

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

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

TECH • VERILY

Trump hyped Verily's coronavirus testing tool. It led to less than 1% of all tests in 2020

BY DANIELLE ABRIL

December 29, 2020 4:05 PM EST

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

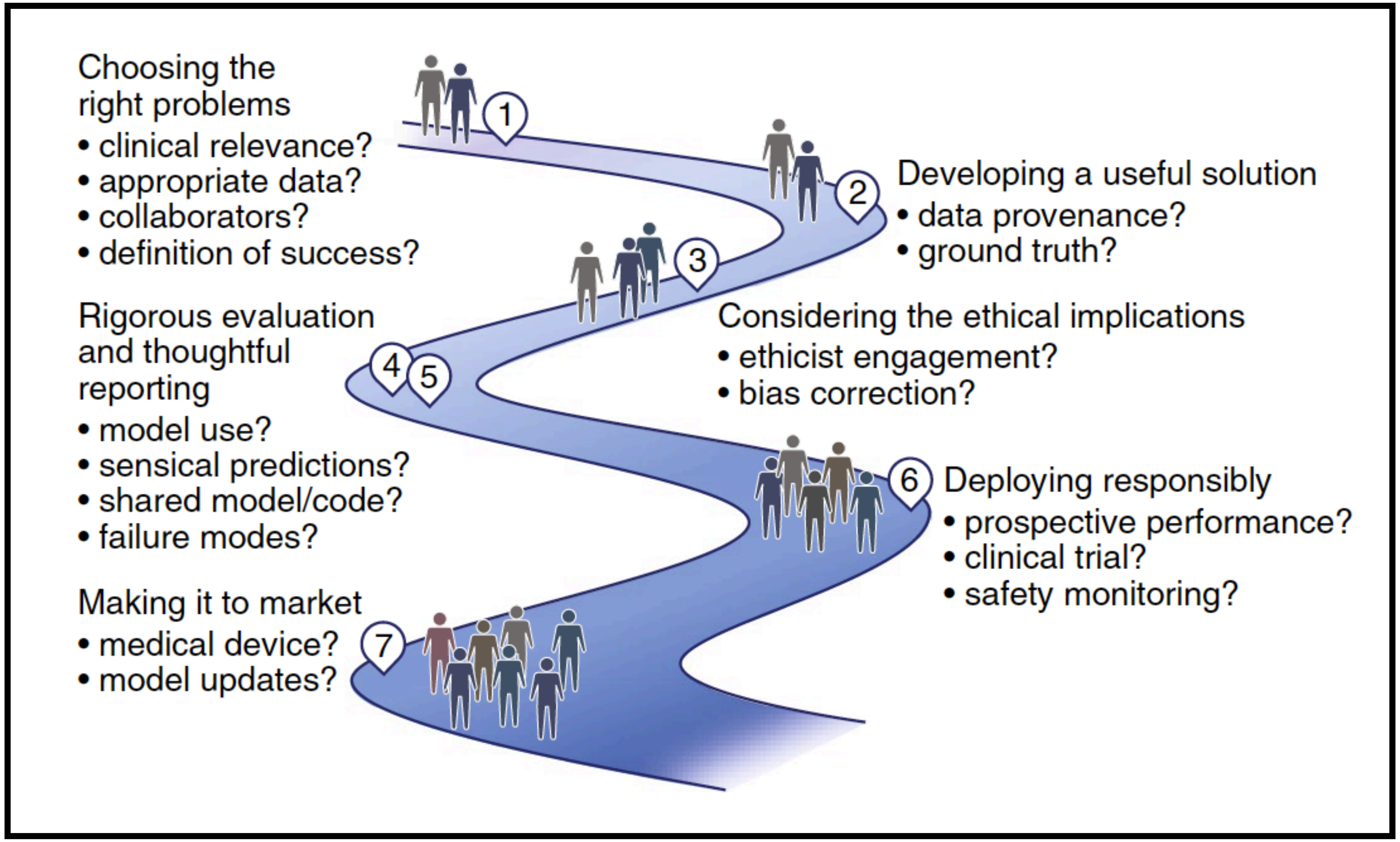
Artificial intelligence / Machine learning

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

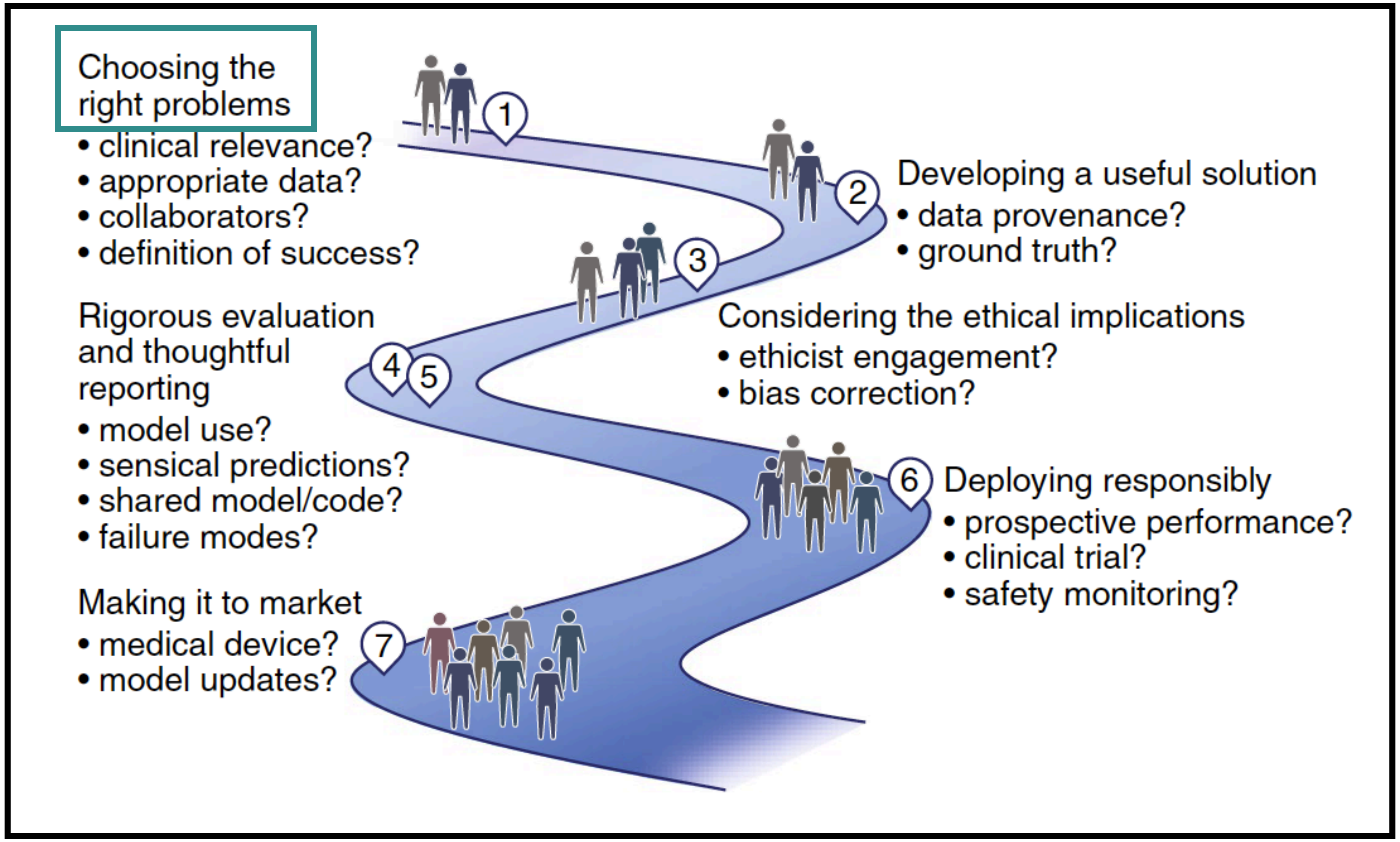
If AI is really going to make a difference to patients we need to know how it works when real humans get their hands on it, in real situations.

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

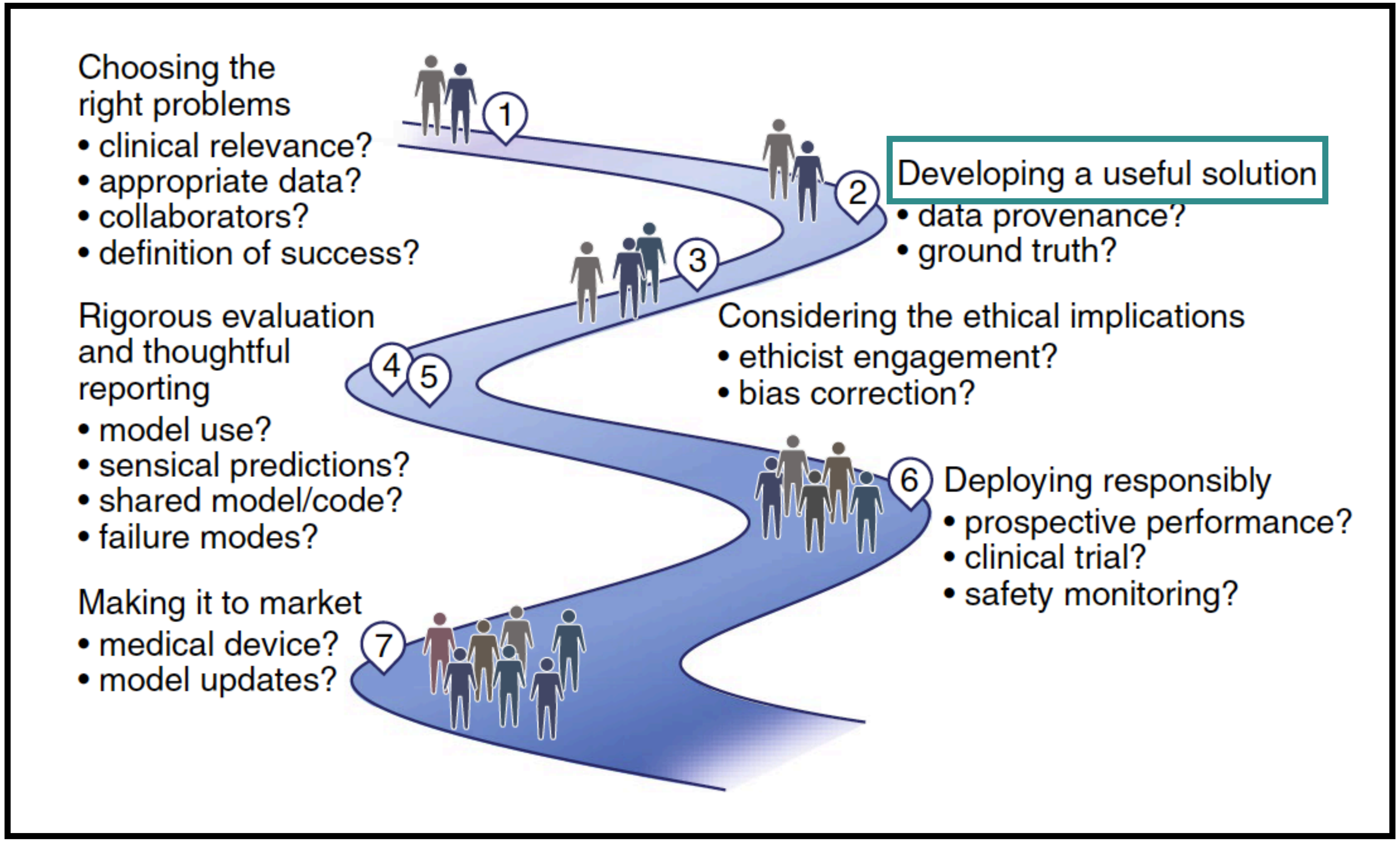
Importance of an Interdisciplinary Team

- **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 • ML researchers • Health information and technology experts • Implementation experts
Decision-makers	• Hospital administrators • Institutional leadership • Regulatory agencies • State and federal government
Users	• Nurses • Physicians • Laboratory technicians • Patients • Friends and family (framily)

Proposed Framework



Developing a Useful Solution

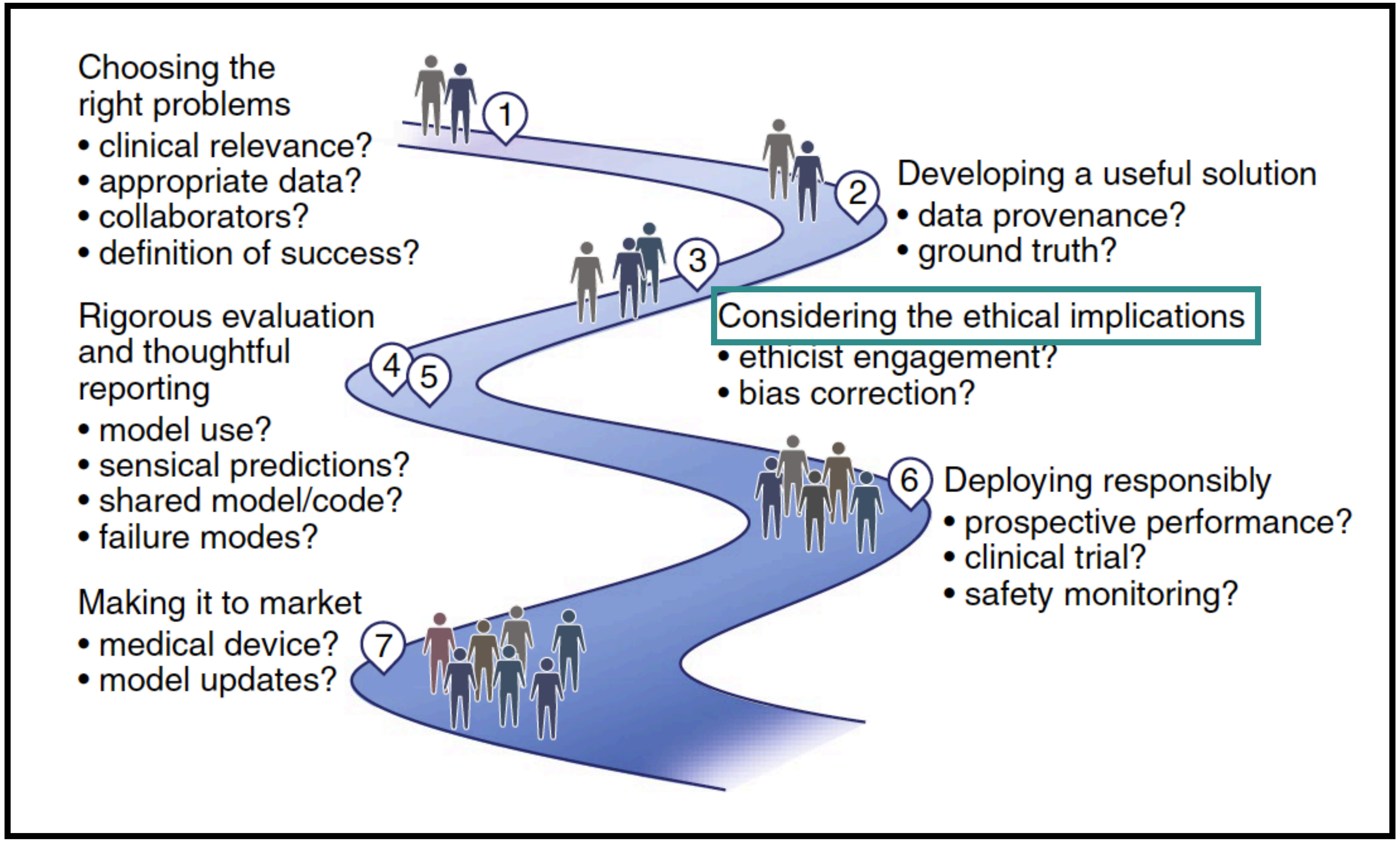
- Understand the data to determine if they are appropriate for the question

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**

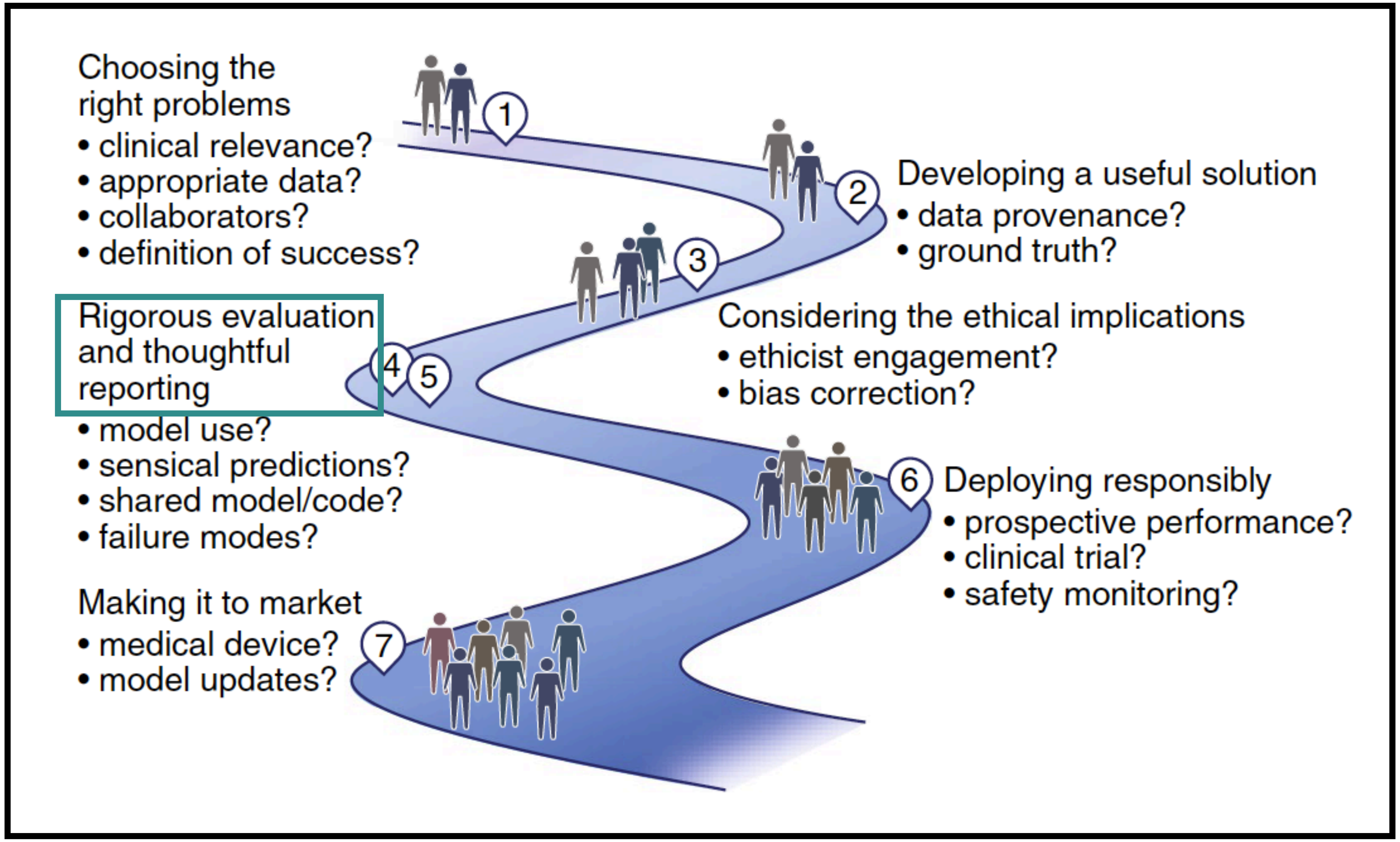
Proposed Framework



Considering Ethical Implications

- Health data is prone to race, sex, and other biases
  - Ex. 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
  - Ex. Imputation of missing data such as HIV and smoking status
- Requires that ethicists, ML experts, and other stakeholders work together

Proposed Framework



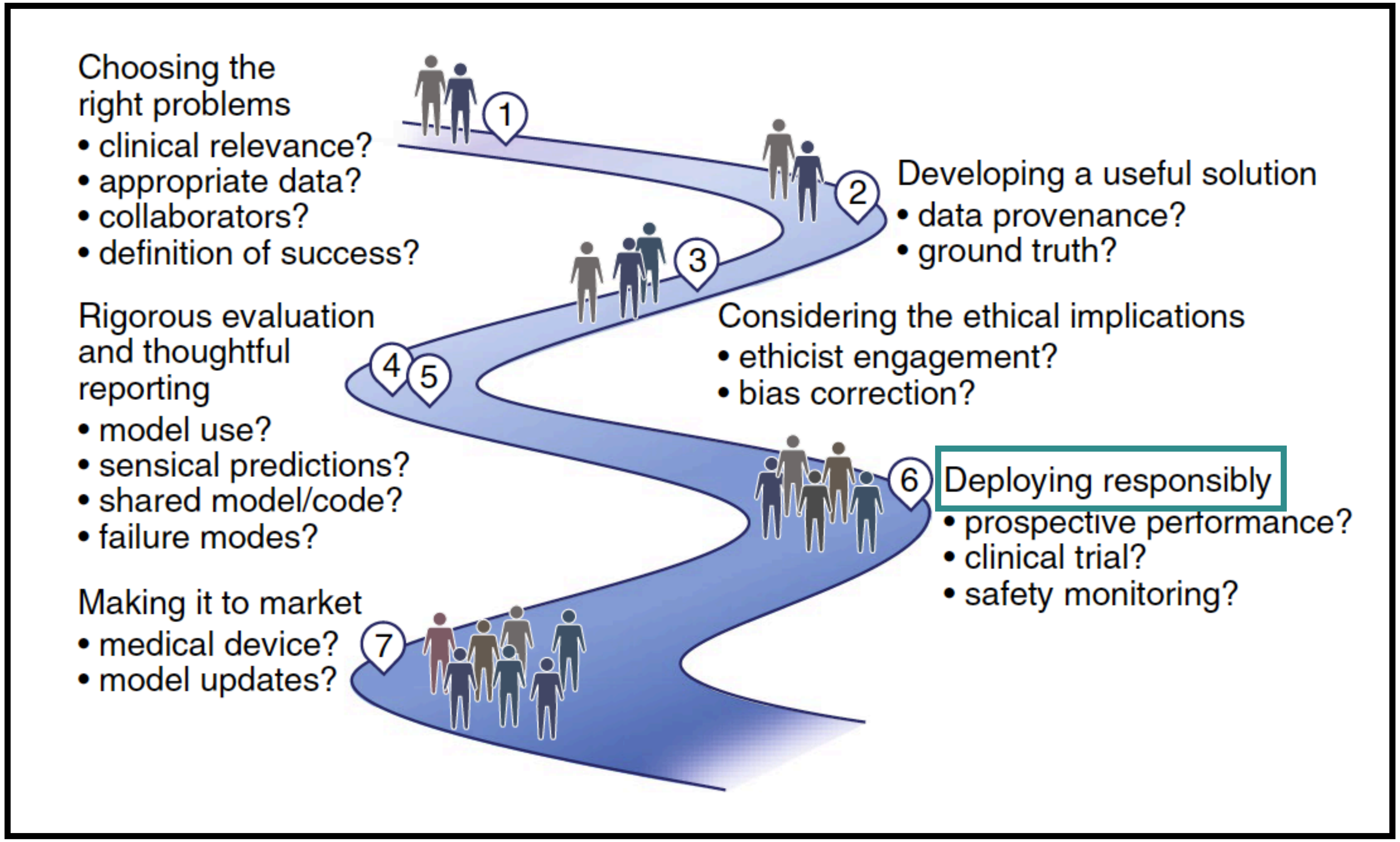
Rigorously Evaluating the Model

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

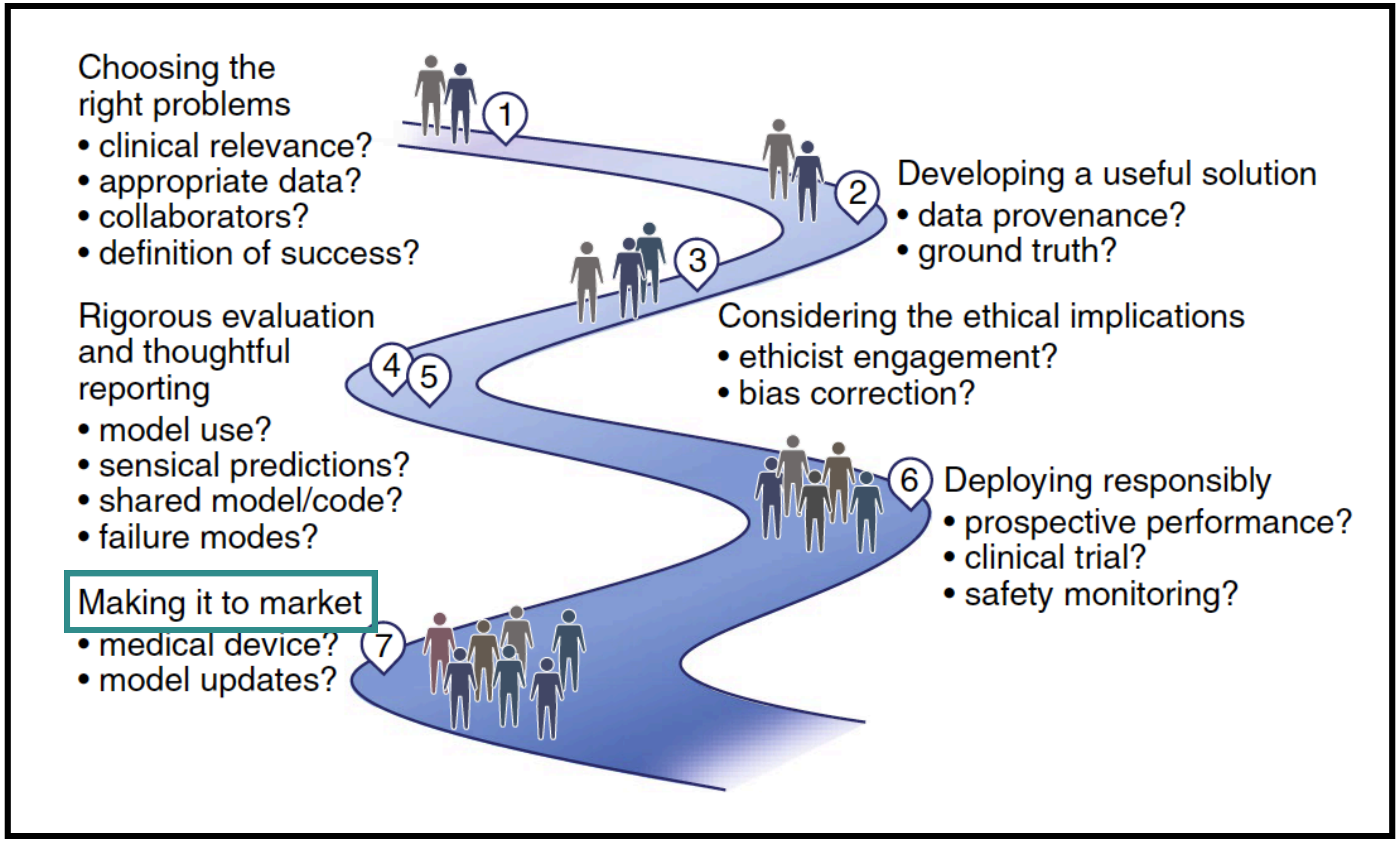
Proposed Framework



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
  - <u>Ex</u>. 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?