

Viewpoint

The Last Mile: Where Artificial Intelligence Meets Reality

Enrico Coiera, MBBS, PhD

Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

Corresponding Author:

Enrico Coiera, MBBS, PhD

Australian Institute of Health Innovation

Macquarie University

Level 6 75 Talavera Rd

Sydney, 2109

Australia

Phone: 61 298502403

Email: enrico.coiera@mq.edu.au

Related Article:

Comment on: <http://jmir.org/2019/10/e16194/>

Abstract

Although much effort is focused on improving the technical performance of artificial intelligence, there are compelling reasons to focus more on the implementation of this technology class to solve real-world applications. In this “last mile” of implementation lie many complex challenges that may make technically high-performing systems perform poorly. Instead of viewing artificial intelligence development as a linear one of algorithm development through to eventual deployment, there are strong reasons to take a more agile approach, iteratively developing and testing artificial intelligence within the context in which it finally will be used.

(*J Med Internet Res* 2019;21(11):e16323) doi: [10.2196/16323](https://doi.org/10.2196/16323)

KEYWORDS

artificial intelligence; implementation science; sociotechnical systems

Trying to separate information technology from the way it is used by people is like trying to separate a breath from the lungs that took it. We are inextricably bound to technology, and it shapes us as much as we shape it. This idea that humans and machines together constitute a larger sociotechnical system is foundational to our understanding of the modern world [1].

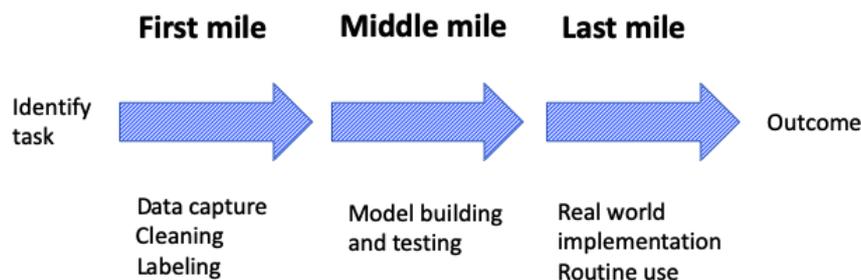
The sociotechnical lessons that the past hold for emerging technologies such as artificial intelligence (AI) should come as no surprise [2]. Amongst these old lessons is the maxim that the application of technology should be shaped by the problem at hand, and not the technology [3]. Although innovation in technology opens up new classes of solution, only the real world can tell us which problems are most worth solving and which solution class is most appropriate. The corollary is that a technology-driven demonstration of success with a problem does not necessarily mean it is the best solution for that problem, nor even that the problem is important.

There are thus two broad agendas for AI research [4]. The first is a technical one, pursuing new methods, architectures, and technologies that push the boundary of machine intelligence. The second is an applied agenda, where we focus on questions

of implementation, extending our understanding of how best to exploit technology within the sociotechnical world it must operate.

The risk we pay for technology neophilia, our love of the new, is that we may not see the system for the technology. With AI, much research focuses on demonstrating technological success on clinical tasks such as image interpretation, diagnosis, or treatment options [5]. There are many compelling reasons for us to shift the balance from such technical demonstrations to real-world applications, not least because many technical leaps are still needed to solve challenging real-world problems.

There are three broad stages in the development of data-driven technologies like machine learning (Figure 1). Once a task such as diagnosis has been selected for algorithm development, in the “first mile,” data are acquired, possibly labelled, and preprocessed or “cleaned.” Next is the middle mile, which focusses on developing and testing the technical performance of different algorithms built using these data. Only in the last mile are algorithms embedded in real-world processes and tested for impact on real-world outcomes.

Figure 1. Three stages in the development of artificial intelligence technologies.

Each “mile” has its own challenges. “First mile problems” include the foundational challenges of gathering and curating high-quality data. For technologies such as machine learning, which are often dependent on large amounts of high-quality data, a bottleneck in data acquisition translates into a roadblock to technology application. The middle mile is home to all the challenges of data-driven algorithm development, including managing biases, replicability, causal inference, avoiding overfitting on training data, and enhancing the generalizability of any models and algorithms developed.

In the last mile, we face the reality that AI does not do anything on its own. It must be connecting somehow to real-world processes, and its impact on those processes needs to be consequential. It is at this point that technology developed for its own sake quickly comes to grief. For example, it is one thing to demonstrate machine learning can interpret thyroid scans for cancer as well or better than humans—a technical feat [6]; it is another for that feat to be meaningful. In the current setting, where thyroid cancer is both overdiagnosed and overtreated, we do not necessarily need better diagnoses. Instead, we need more nuanced and less aggressive approaches to management [7].

Last mile challenges are thus ones of implementation, and for researchers, of implementation science. These challenges exist at many levels, and include the following:

- **Measurement:** Standard metrics of AI performance relate to how well it completes its assigned task. We traditionally use performance measures like sensitivity, specificity, and area under the receiver operating characteristic curve (AUC). There is, however, a long chain of events that must occur between high technical performance and actual impact on clinical outcome. Success at an early stage in this information value chain is necessary, but not sufficient, to guarantee impact in the real world [8]. Evaluating real-world performance thus requires a shift from measuring technical accuracy to evaluating impact on processes and people. Measures like AUC can also be misleading in clinical settings [9]. AUC measures the overall performance of an algorithm across the entire receiver operating characteristic curve, while ideal real-world operation may best be limited to one segment of it [10].
- **Generalization and Calibration:** When an AI is trained on historic data, its future performance is dependent on how well new data match the historic data. It is a common implementation challenge to discover that a high-performing algorithm developed on data from one population deteriorates when used on a different population, reflecting

underlying differences in the frequency and nature of events within the data sets. For this reason, AIs may need to be tuned to a specific end population. In many settings, this final population will also be dynamic, varying because of recurrent events like seasonal disease shifts, changes in population characteristics, and unexpected new events such as disease outbreaks. This means that AIs may need to be recalibrated, either periodically or dynamically, to reflect population changes. We also will need to closely monitor AI performance to detect shifts in its behavior that indicate that recalibration is needed [11]. More challenging is the fact that the more effectively an AI improves outcomes, the faster its performance may appear to degrade, as its very success can alter the association between predictors in its model and outcomes [12].

- **Local Context:** It is a fundamental tenet of implementation science that differences in the context in which a technology is embedded are associated with changes in performance. If we see an organization as a network of people, processes, and technologies, it is clear that the network underpinning any two organizations will be different. Implementation can be seen as the act of fitting a new technology or process into a pre-existing organizational network, and the goodness of fit of technology to network will shape any impact on organizational performance [13]. This is as true for AI as it is for digital health, in general, or for any new process or technology. To complicate matters, organizational networks are themselves dynamic. The impact of a technology will thus change with time, as the way it “fits” an organizational network changes and possibly because of its own presence—past connections will disappear or be replaced by new ones.

The software world has made a fundamental shift from seeing software development as a linear process that starts with user requirements and ends with end-user testing, to an agile one where users are embedded in a rapid and iterative process that adaptively fits software to users. Implementation science must travel the same path and especially so with dynamically changing AI.

AI development should not be seen as a linear journey stretching from the first to the last mile. Doing so risks the end product not meeting real-world needs, just as it does with software. Instead, implementation should be seen as an agile, iterative, and lightweight process of obtaining training data, developing algorithms, and crafting these into tools and workflows. Finding the right balance between reusing general technology and building to meet local needs will be crucial [14]. Either way,

AI should not be created far from the place they will be used. will ultimately live in.
Ideally, they should be born deep within the network that they

Conflicts of Interest

None declared.

References

1. Berg M, Aarts J, van der Lei J. ICT in health care: sociotechnical approaches. *Methods Inf Med* 2003;42(4):297-301. [Medline: [14534625](#)]
2. Stefanelli M. The socio-organizational age of artificial intelligence in medicine. *Artif Intell Med* 2001 Aug;23(1):25-47. [Medline: [11470215](#)]
3. Coiera E. Medical informatics. *BMJ* 1995 May 27;310(6991):1381-1387. [doi: [10.1136/bmj.310.6991.1381](#)] [Medline: [7787545](#)]
4. Coiera EW. Artificial intelligence in medicine: the challenges ahead. *J Am Med Inform Assoc* 1996;3(6):363-366 [FREE Full text] [doi: [10.1136/jamia.1996.97084510](#)] [Medline: [8930853](#)]
5. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019 Jan;25(1):44-56. [doi: [10.1038/s41591-018-0300-7](#)] [Medline: [30617339](#)]
6. Li X, Zhang S, Zhang Q, Wei X, Pan Y, Zhao J, et al. Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. *The Lancet Oncology* 2019 Feb;20(2):193-201. [doi: [10.1016/S1470-2045\(18\)30762-9](#)]
7. Coiera E. On algorithms, machines, and medicine. *Lancet Oncol* 2019 Feb;20(2):166-167. [doi: [10.1016/S1470-2045\(18\)30835-0](#)] [Medline: [30583847](#)]
8. Coiera E. Assessing Technology Success and Failure Using Information Value Chain Theory. *Stud Health Technol Inform* 2019 Jul 30;263:35-48. [doi: [10.3233/SHTI190109](#)] [Medline: [31411151](#)]
9. Cook NR. Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction. *Circulation* 2007 Feb 20;115(7):928-935. [doi: [10.1161/CIRCULATIONAHA.106.672402](#)]
10. Thomas G, Kenny LC, Baker PN, Tuytten R. A novel method for interrogating receiver operating characteristic curves for assessing prognostic tests. *Diagn Progn Res* 2017 Nov 15;1(1):17. [doi: [10.1186/s41512-017-0017-y](#)]
11. Ong M, Magrabi F, Coiera E. Syndromic surveillance for health information system failures: a feasibility study. *J Am Med Inform Assoc* 2013 May 01;20(3):506-512 [FREE Full text] [doi: [10.1136/amiajnl-2012-001144](#)] [Medline: [23184193](#)]
12. Lenert M, Matheny ME, Walsh CG. Prognostic models will be victims of their own success, unless.... *J Am Med Inform Assoc* 2019 Sep 03;ocz145. [doi: [10.1093/jamia/ocz145](#)] [Medline: [31504588](#)]
13. Coiera E. Chapter 12: Implementation, in *Guide to Health Informatics*. Boca Raton: CRC Press; 2016:173-194.
14. Johannessen L, Ellingsen G. Integration and Generification—Agile Software Development in the Healthcare Market. *Comput Supported Coop Work* 2009 Sep 16;18(5-6):607-634. [doi: [10.1007/s10606-009-9097-8](#)]

Abbreviations

AI: artificial intelligence

AUC: area under the receiver operating characteristic curve

Edited by G Eysenbach; submitted 18.09.19; peer-reviewed by D Mendes; accepted 20.10.19; published 08.11.19

Please cite as:

Coiera E

The Last Mile: Where Artificial Intelligence Meets Reality

J Med Internet Res 2019;21(11):e16323

URL: <http://www.jmir.org/2019/11/e16323/>

doi: [10.2196/16323](#)

PMID: [31702559](#)

©Enrico Coiera. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 08.11.2019. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic

information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.