A Graphical Toolkit for Longitudinal Dataset Maintenance and Predictive Model Training in Health Care

Eric Bai[1,*] Sophia L. Song[1,*] Hamish S. F. Fraser[2] Megan L. Ranney[3]

[1] Warren Alpert Medical School, Brown University, Providence, Rhode Island, United States
[2] Brown University Center for Biomedical Informatics, Providence, Rhode Island, United States
[3] Brown-Lifespan Center for Digital Health, Providence, Rhode Island, United States

Address for correspondence Megan L Ranney, MD, MPH, FACEP, Brown-Lifespan Center for Digital Health, 139 Point Street, Providence, Rhode Island 02903, United States (e-mail: megan_ranney@brown.edu).

Appl Clin Inform 2022;13:56–66.

Abstract

Background Predictive analytic models, including machine learning (ML) models, are increasingly integrated into electronic health record (EHR)-based decision support tools for clinicians. These models have the potential to improve care, but are challenging to internally validate, implement, and maintain over the long term. Principles of ML operations (MLOps) may inform development of infrastructure to support the entire ML lifecycle, from feature selection to long-term model deployment and retraining.

Objectives This study aimed to present the conceptual prototypes for a novel predictive model management system and to evaluate the acceptability of the system among three groups of end users.

Methods Based on principles of user-centered software design, human-computer interaction, and ethical design, we created graphical prototypes of a web-based MLOps interface to support the construction, deployment, and maintenance of models using EHR data. To assess the acceptability of the interface, we conducted semistructured user interviews with three groups of users (health informaticians, clinical and data stakeholders, chief information officers) and evaluated preliminary usability using the System Usability Scale (SUS). We subsequently revised prototypes based on user input and developed user case studies.

Results Our prototypes include design frameworks for feature selection, model training, deployment, long-term maintenance, visualization over time, and cross-functional collaboration. Users were able to complete 71% of prompted tasks without assistance. The average SUS score of the initial prototype was 75.8 out of 100, translating to a percentile range of 70 to 79, a letter grade of B, and an adjective rating of "good." We reviewed persona-based case studies that illustrate functionalities of this novel prototype.

Conclusion The initial graphical prototypes of this MLOps system are preliminarily usable and demonstrate an unmet need within the clinical informatics landscape.

Keywords

- user-centered design [L01.224.900.900]
- software design [L01.224.900.820]
- software validation [L01.224.900.868]
- data management [L01.399.375]
- electronic health records and systems
- machine learning
- data processing
- security
- interfaces and usability

* *These authors contributed equally to this work.*

received
May 26, 2021
accepted after revision
November 9, 2021

© 2022. Thieme. All rights reserved.
Georg Thieme Verlag KG,
Rüdigerstraße 14,
70469 Stuttgart, Germany

DOI https://doi.org/10.1055/s-0041-1740923.
ISSN 1869-0327.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Background and Significance

Big data are rapidly transforming how health care generates new and leverages existing knowledge.[1] Defined by the "four Vs," that is, volume, variety, velocity, and veracity,[2] big data are especially promising in informing health care predictive analytics.[3] Machine learning (ML) models have been developed to accurately predict risk of adverse events including hospital readmission,[4–7] sepsis,[8,9] suicide risk,[10] opioid overdose risk,[11] and postmyocardial infarction mortality risk,[12] as well as coordinate clinical management.[13] An especially important repository of health care big data is the electronic health record (EHR), through which vast stores of information can be automatically processed to create a collection of tools to support clinician decision-making and inform patient care.

A large body of literature exists on the development of predictive models for health care, but best practices for local validation, implementation, and long-term maintenance in clinical settings remain unclear.[14] Coiera has referred to this divide as the "last mile" of implementation, where algorithms must navigate the challenges of differing local contexts and preexisting organizational structures.[15] Challenges specific to the "last mile" include validating accuracy metrics on a local level, connecting accuracy to clinical outcomes, generalizing and calibrating datasets drawn from diverse populations, and embedding models in a dynamic organizational system.[15] Other challenges include infrastructure investment, high maintenance costs, and the need to continuously adapt the system to handle new forms of data.[16–24] The process demands significant time and human capital and thus is frequently limited to a few major academic centers.[20,21,23] Robust validation[22] is particularly difficult, as the heterogeneity of EHR systems and strict privacy regulations preclude widespread interinstitutional sharing. Therefore, it is unsurprising that only 34.6% of hospital readmission models[5] and only 44 models published between 2010 and 2019 are implemented in real-world clinical practice.[21] Even after successful initial implementation, ML systems' development process makes them susceptible to "technical debt," the long-term negative effects of immature, insufficiently tested code created early on in the software development life cycle.[23,25] This technical debt manifests as high maintenance costs, inability to add new features, or, in severe cases, the need to replace the entire system. Designing and building software infrastructure to support the ML development lifecycle and manage technical debt is the basis for ML operations (MLOps), an emerging field focused on applying well-established development and operations (DevOps) practices in software engineering to scaling ML in production.[26–28] Sculley et al observed that only a small fraction of the codebase of ML projects is actual ML code; the rest is infrastructure that can be factored out of individual projects into a reusable framework.[23,29] Reducing codebase size reduces the footprint susceptible to accruing technical debt.[25]

While DevOps tools are robust and widely available, few comprehensive, publicly available end-to-end platforms for MLOps exist.[29] For example, IBM's Runway[30] and Facebook's FBLearner[31] are both proprietary internal tools for managing ML experiments. Google's Kubeflow[32] and TensorFlow Extended[33] are specific to Kubernetes and TensorFlow, respectively. ModelDB is limited to computational neuroscience.[34] MLweb[34] is browser-based and is thus limited by performance constraints and lack of access to prebuilt codebases. MLFlow has emerged as an open-source end-to-end MLOps framework,[35] but it is a command-line driven tool requiring high levels of technical expertise. In addition, issues of data security, standardization, governance, and regulatory compliance specific to health care necessitate specialized solutions for this field.[1]

Objectives

In this paper, we present a conceptual framework, in the form of a graphical prototype, for a programming language and model agnostic MLOps tool designed to support the entire lifecycle of predictive model development using EHR data. We use best practices from the Human–Computer Interaction (HCI)[36,37] to design and assess two prototypes for an end-to-end machine learning workflow. Finally, we present several persona-based case studies demonstrating MLOps' potential functionalities such as data management, feature selection, model training, deployment, long-term maintenance, and visualization over time.

Methods

Defining User Personas

From review of health information technology (IT) implementation workflows at health care organizations, including our own,[38,39] we employed the HCI Personas methodology[36,37] as a framework for approaching and categorizing the potential user base of our prototype. Through iterative consultation with content experts, we used this framework to organize health care–based data analysis and management roles into three key "user personas." These are the health informatician, the chief information officer (CIO), and the organization's clinical and data stakeholders (physicians, clinical staff, and health care researchers).

Together, these three personas are involved in the creation, maintenance, and ongoing evaluation of clinically relevant predictive models in the health care setting. The health informatician focuses on exploring datasets to support the development and validation of predictive models for clinical problems identified by the clinical and data stakeholders. At many organizations, this process is conducted with oversight from the CIO, who serves as the liaison between health informaticians and clinical and data stakeholders.[40] Clinical and data stakeholders serve as both the drivers of model development and as the end users. Ongoing evaluation of model performance involves health informaticians preparing performance data for the CIO and other stakeholders to review. These user personas represent a simplified schematic of the larger health care analytics space; their unique roles inform workflow development and customization.

 © 2022. Thieme. All rights reserved.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

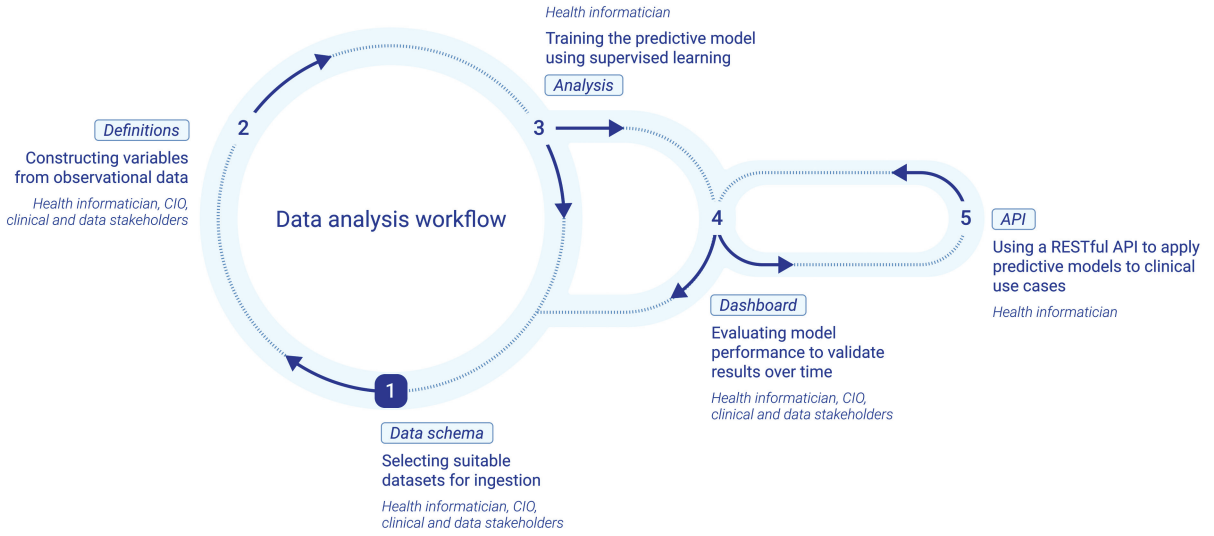

Fig. 1 A visual depiction of our standardized machine learning workflow. The intersecting circular shapes highlight the iterative nature of predictive model creation. The outlined badges indicate the major end-user groups that interact at each step. While the health informatician participates in all stages of the workflow, clinical and data stakeholders and the CIO participate in selecting suitable datasets for ingestion, constructing variables, and evaluating model performance over time. API, application programming interface; CIO, chief information officer; RESTful, representational state transfer.

Defining Workflow Parameters

To define the workflow parameters for predictive model development in clinical settings, we draw on best practices for automated analyses without loss of validity,[20] including principles of continuous integration in software engineering[27] and the framework for a "delivery science" of artificial intelligence in health care described by Li et al.[14] The following workflow parameters, modified from those proposed by Reps et al,[41] were identified for developing and implementing predictive models in clinical practice.

1. Selecting suitable datasets for ingestion.
2. Constructing variables from observational data.
3. Training the predictive model using supervised learning.
4. Evaluating model performance with internal validation using test sets, and, subsequently, with prospectively collected data.
5. Using a representational state transfer (RESTful) application programming interface (API) to apply models to clinical use cases after appropriate clinical sign-off; this includes postdeployment monitoring and periodic model retraining.

In ►Fig. 1, we map our identified user personas onto this ML workflow, highlighting the leading role of the health informatician and the specific points of collaboration between the personas. As noted in the figure, the health informatician is primarily responsible for executing the workflow with support and input from the CIO and clinical and data stakeholders.

Framework Design

We reviewed literature on principles of human-centered design in health care[42–45] and the potential challenges of operationalizing ML at scale[29,46] to determine features to include in the initial prototype designs (►Table 1). To accommodate both novice and experienced health informaticians and support visual representations of data, we designed the framework as a web-based graphical user interface (GUI) rather than a command-line interface. For many tasks in both health care and non-health care settings, GUIs are easier to learn and result in higher user satisfaction than command line interfaces.[47,48] Benefits of GUIs include the ability to provide guidance at each required step, encourage exploration, and support graphical visualizations of data.[49]

Using the user interface design tool Figma (San Francisco, California, United States), we designed two prototypes; these are based on our team's ongoing "Emergency Digital Smart Notifications (EDSN)" project focused on training predictive models using a deidentified longitudinal EHR dataset of more than 190,000 emergency department visits at an academic health center in Rhode Island (in preparation).

Framework Usability Testing

We identified a convenience sample of six individuals representing the three potential user personas: two health informaticians, two CIOs, and two clinical and data stakeholders (health care researchers). While the workflow is primarily targeted toward the health informatician, we included CIO and clinical and data stakeholder personas in our user testing because they participate in decisions regarding dataset selection, variable definition, and model evaluation. Additionally, feedback on acceptability from multiple groups encourages discussion on organization-wide "data culture" in successful implementation of models.[50]

Two facilitators (E.B. and S.S.) following a semistructured script (►Supplementary Appendix A1, available in the online

 © 2022. Thieme. All rights reserved.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Table 1 Design features by core design tenet

Core tenet	Mission statement	Proposed design features
Ethical	The system must be ethical, adhering to the highest standards of data security and preservation of patient anonymity through deidentified data	Accepts only deidentified PHI and includes robust configuration and administrative controls for enforcing security best practices
Auditable	The system must be auditable so that all events are viewable for security, monitoring, and debugging purposes	All security, monitoring, and debugging events are viewable and downloadable. Qualified users can opt-in to notifications of key system events
Adaptable	The system must be adaptable so that the system continues to grow in value over time to justify the up-front cost of implementation	Model performance is trended over time and all features, outputs, and models have edits tracked over time for easy viewing and rollback
Automated	The system must be automated so that the models can continue to incorporate new input within user-defined parameters at all steps of the workflow without immediate user intervention	New data automatically triggers a cascade of updates to features, outputs and models without manual user intervention. These asynchronous running tasks are centralized for easy monitoring
Accessible	The system must be accessible so that the results of the models are available to all those who need them	Validated models can be made publicly accessible via a RESTful API for application to clinical settings after appropriate clinical sign-off with post-deployment monitoring using the dashboard and events views. Performance reports can be generated for communicating higher order trends in performance

Abbreviations: API, application programming interface; PHI, protected health information; RESTful, representational state transfer.

version) explored the acceptability of both prototypes (►**Supplementary Appendix A2**, available in the online version) with potential users who were naïve to the design. One facilitator moderated each session while the other recorded user quotes and observations. After navigating through each prototype, interviewees were asked to fill out the System Usability Scale,[51] a reliable (Cronbach's $\alpha > 0.90$)[52–54] survey instrument validated in several contexts for evaluating usability of software interfaces.[55]

Following these evaluative interviews, we tallied completion rates for each task and iteratively analyzed interview notes containing both verbatim quotes and interviewer observations to reveal recurrent themes.[56] We then revised designs to correct usability shortcomings (identified, for example, by low completion rates during testing) and to better support the workflow described by each user persona during the interview.

Results

Initial Prototypes

We created two prototypes focusing on our primary use cases for the health informatician (►**Supplementary Appendix A2**, available in the online version). The first prototype highlights the workflow of creating and implementing a new model, including primary data schema definition, trialing feature and outcome definitions, and debugging initial versions of ML scripts. The second prototype explores the experience of maintaining production models deployed via the API over the span of a few months.

Usability Testing Results

As expected, participants from each of our three identified user personas (health informatician, the CIO, and the organization's clinical and data stakeholders) varied widely in both familiarity with ML principles and day-to-day job descriptions. On average, users were able to complete 71% of prompted tasks, with the most challenging being discovering the events page, configuring API settings, and interacting with the dashboard visualizations without prior guidance. ►**Supplementary Appendix B** (available in the online version) describes detailed feedback, including task completion counts grouped by thematic category.

We obtained SUS scores for four out of six user testing sessions for an average score of 75.8 out of 100. This is slightly higher than the reported average of 67.6 for enterprise software prototypes.[57] Using the Curved Grading Scale (CGS) interpretation for SUS scores,[58] this translates to a percentile range of 70 to 79, a letter grade of B, and an adjective rating of "good."

Users found the overall value of the system to be streamlining the continued monitoring and upkeep of production models. Users challenged the team to provide better explanations for each section to expand the number of programming languages supported when specifying models and to consider more sophisticated performance visualizations

 © 2022. Thieme. All rights reserved.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Table 2 Key user testing sentiments by user persona

User persona	Key sentiments
Health informatician	Emphasized the usefulness of bringing together a "live model that keeps updating on some kind of scheduled basis" and being "able to monitor [the model]" after deployment
	Wanted "a little bit more of an introduction" and "a little more description about what each section does" for the "users who are coming to use these tools less frequently and from different data science education levels"
Health care researcher (a clinical and data stakeholder)	"Once [she] finished developing [a model], this would be a really good interface especially if there are multiple people working on this"
	Setting up the system would "need lots of prework [but] once complete … would be really helpful" over the long term
Chief information officer (CIO)	Remarked on the possible "reduce[d] maintenance work" that would come with manual retraining and monitoring
	Regarding writing structured query language queries in the configuration pages, a CIO concluded that "this is the kind of thing [for which] you'd want an analyst or data scientist … rather than a clinician"

rather than summary scores trended over time (►Table 2). Based on user feedback, we revised the designs (►Fig. 2: overview of revised designs; ►Supplementary Appendix C [available in the online version]: annotated design documents), guided by core design tenets (►Supplementary Appendix D, available in the online version).

Case Study of Potential Health Informatician Workflow

Here, we outline the potential workflow of this MLOps program for a new health informatician user. The health informatician first configures mapping rules to transform data stored in comma separated value (CSV) files to structured query language (SQL) tables, specifying how CSV file names map onto SQL table names (►Fig. 3A) and how CSV columns map onto SQL columns (►Fig. 3B). We chose the widely used SQL language as a database management tool because its fixed data schema is well-suited for the structure of most clinical datasets.[59] Future data exports from the EHR will automatically apply these rules to ingest data. Once data are ingested, the informatician codes SQL definitions for calculating all inputs and outcomes from the ingested data schema. The built-in SQL editor prioritizes usability by listing out query requirements, providing a preview of the available data schema, and showing query sample output (►Fig. 3D). As input and outcome definitions are updated, all edits to these definitions are tracked to form a project changelog for future records (►Fig. 3C).[60]

The informatician is now ready to define models, first specifying SQL definitions for inputs and outcomes and then the model environment before implementing each of the four scripts required for a complete model definition (►Fig. 4B) and then running the training function. The interface does not have "prepackaged" models; rather, users may write or import their own custom scripts. While the model environment must be one of the supported languages (currently Julia,[61] Python,[62] and R[63]), the editor also allows users to import project files from their computers or a third-party code-hosting service such as Github (San Francisco, California, United States). Users can select shell scripts from these project files to run either before or after the training function is executed (►Fig. 4C). This approach of selecting a fixed-model environment while supporting shell scripts enables support for polyglot projects and represents a compromise between language-specific optimizations and the infinite flexibility of shell scripting. As model definitions are updated, all prior iterations of the model remain viewable (►Fig. 4A).

Models are automatically refreshed each time new data are detected; any currently executing tasks are summarized within a top-level menu for easy tracking (►Fig. 5A). Performance of each of these iterations may be trended graphically using standard or custom metrics to determine promising candidate models (►Fig. 5B). Aggregate performance across multiple outcomes—perhaps each an "experiment"[41]—may also be compared (►Fig. 5A). Model candidates can be made publicly accessible via an access-controlled RESTful API (►Fig. 5C) to either deploy to production or integrate with additional tools. Documentation for each API endpoint is automatically generated to further streamline access (►Fig. 5D).

The flexibility of our script-based approach to model configuration supports external validation. During model creation, the informatician may select inputs and outcomes corresponding to both the internal and external validation datasets (►Fig. 4B). While the internal validation data are used in the first script to train the model, both datasets can be used in the fourth script to score the model with respect to both internal and external validation metrics (►Fig. 5A). Also, the API functionality allows models to integrate with other external validation packages (►Fig. 5C).

Health Care–Specific Considerations

Our approach provides tools to address health care–specific considerations, including maintaining patient privacy as

 © 2022. Thieme. All rights reserved.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Fig. 2 Design overview schematic. An overview of the version 2 of our proposed designs after incorporating user feedback. Each tile represents a major section of the design. The directional arrows represent our hypothesized workflow, starting from set-up of the data analysis pipeline to tracking model performance on the dashboard to launching high-performing models via the API. The labels correspond to the specific design document that provides an in-depth overview of that specific functionality. API, application programming interface.

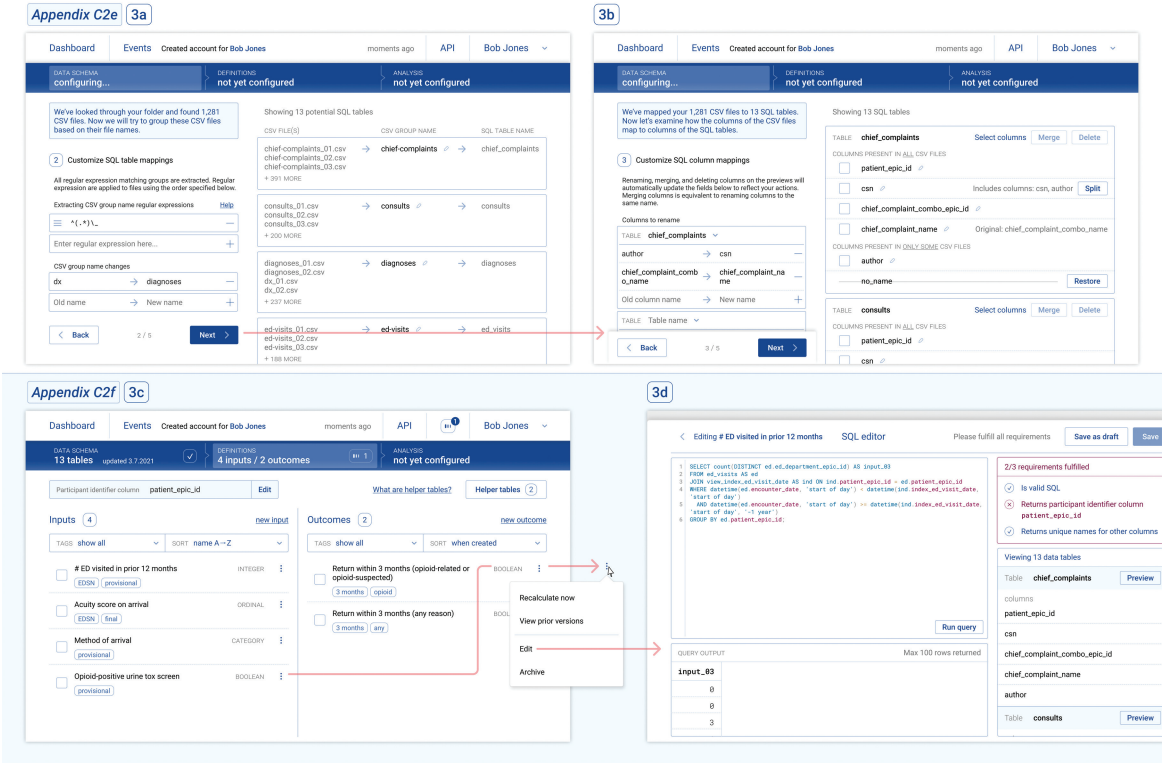

Fig. 3 Design walkthrough, specifying data schema and definitions. Walkthrough of specifying the data schema during data ingestion and input and outcome definitions; (**A**) specifying rules for mapping CSV file names to SQL table names; (**B**) specifying rules for mapping CSV column names to SQL column names; (**C**) overview of all inputs and outcomes, prior versions tracked and accessible via menu; (**D**) SQL editor for specifying definitions with requirements list and query output preview. CSV, comma separated value.

 © 2022. Thieme. All rights reserved.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

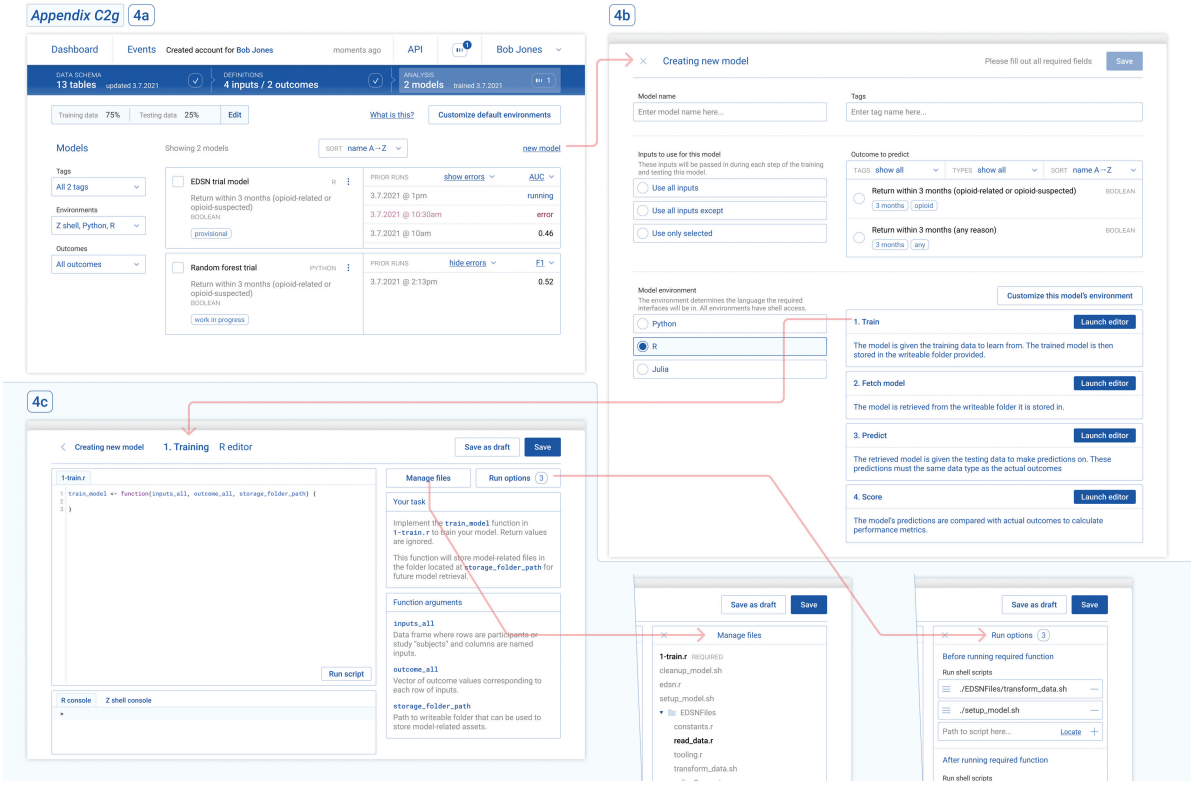

Fig. 4 Design walkthrough, configuring models. Walkthrough of configuring models; (**A**) overview of all active models showing prior runs; (**B**) create a new model by specifying inputs, outcomes, and language environment; (**C**) dedicated code editor for each step of defining a model with ability to add additional project files and specify bash scripts to run before or after function execution

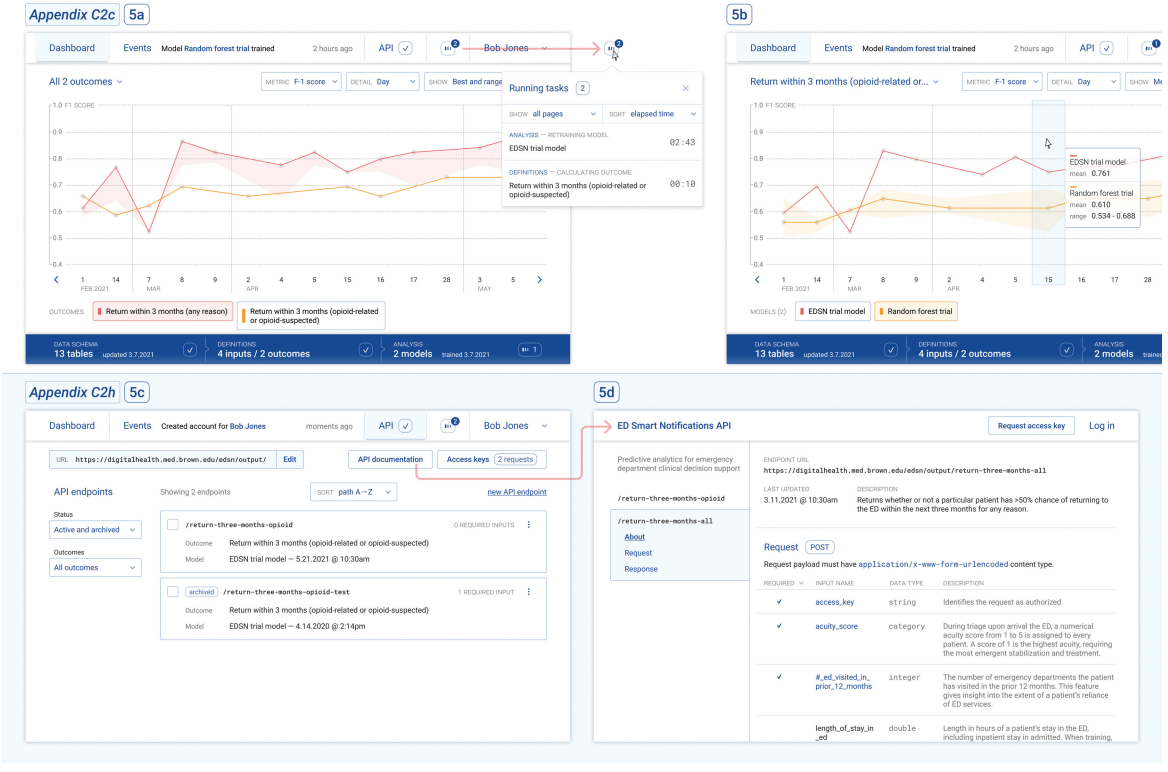

Fig. 5 Design walkthrough, tracking performance and deploying models. Walkthrough of tracking performance and deploying models; (**A**) compare aggregated performance across different outcomes, keep track of ongoing tasks; (**B**) compare model performance for a single outcome; (**C**) configure RESTful API endpoints to allow authorized applications to access trained models; (**D**) send automatically generated documentation to developers of authorized applications. API, application programming interface; RESTful, representational state transfer.

 © 2022. Thieme. All rights reserved.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

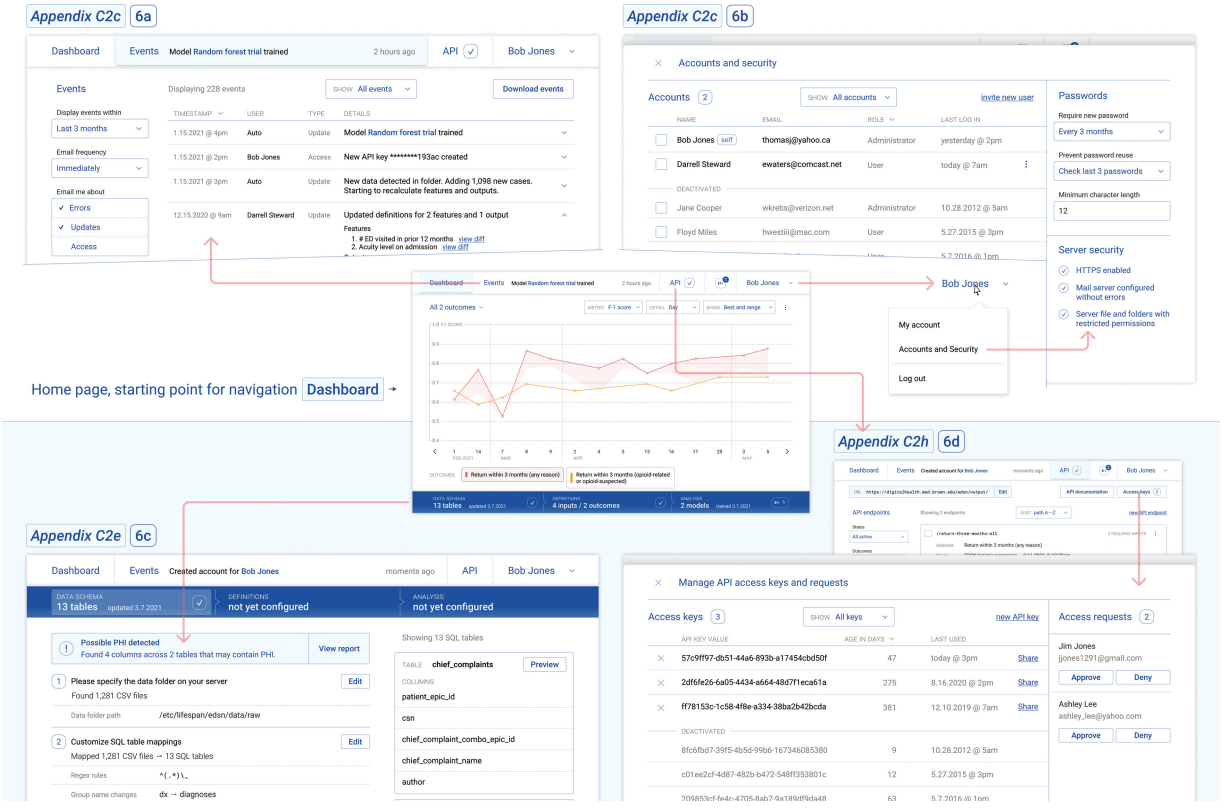

Fig. 6 Healthcare-specific design concerns. Feature designs for healthcare-specific data security and privacy concerns; (**A**) track all events across all users, configure notifications for key events; (**B**) manage access permissions, set password policies, and check server security status; (**C**) flag possible protected health information leaks for remediation; (**D**) create API access keys and manage access requests. API, application programming interface.

specified by the Health Insurance Portability and Accountability Act (HIPAA).[64–66] Technical provisions are maintained by properly configuring server and communication protocols; we increase the visibility of key technical configuration parameters by displaying security status on a prominent status panel (►**Fig. 6B**). Administrative provisions are maintained by a centralized event log with configurable notifications (►**Fig. 6A**), a centralized security dashboard for managing user access and specifying minimum password standards (►**Fig. 6B**), a built-in protected health information (PHI) checker that uses regular expressions to search for possible instances of the 18 PHI identifiers (►**Fig. 6C**), and an access key mechanism to ensure only authorized applications are able to access models via the API (►**Fig. 6D**).

Discussion

We developed and evaluated a prototype for an MLOps GUI to automate data ingestion, model retraining, and deployment for ML integration into clinical tools. We found that the initial designs were acceptable to our three user personas. We then revised the designs to address shortcomings identified by our formative user testing and to incorporate suggestions identified through qualitative thematic analysis of each user-testing session.

Robust literature exists on potential applications for predictive models[4–6,8,10–12] and on the challenges of implementing models in real-world settings.[14,16–24] Our conceptual prototypes extend this work by proposing an integrated MLOps framework for the entire predictive model development lifecycle, including a core panel of features that would otherwise require configuring multiple services into a custom pipeline.[29,67] Unlike efforts by Google,[32,33] Facebook,[31] IBM,[30,46] and others, our work complements existing open-source MLOps frameworks, such as MLFlow,[35] by providing a potential GUI for its primarily command-line driven functionality.

We recognize that graphical prototypes do not easily translate to finished systems. However, our initial mapping of essential features and our designs for an intuitive graphical framework can provide the groundwork for developers of scalable MLOps systems. Although our work was conducted at a single academic health care system, we address the operational challenges of set-up, maintenance, and monitoring common to ML research conducted across a variety of health care institutions. National ML models, such as the Epic Sepsis Model,[68] have poor external validity in local contexts; thus, it is imperative that health care organizations develop in-house capacity for adapting and validating ML models for local use. Of course, the clinical efficacy of any predictive model trained on EHR data are predicated on the quality of the input data and on recognizing the impact of human factors and clinical practice standards on model use.[69] Adequate data governance by

 © 2022. Thieme. All rights reserved.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

hospital systems and widespread, robust testing of the clinical impact of these tools is necessary to successfully traverse the "last mile" of implementation.[50]

Limitations and Future Work

There are several limitations to our study. Given our focus on evaluative rather than summative user testing, this work focused on the detection and elimination of usability issues rather than formal summative usability testing. Also, while our user testing subjects were chosen to minimize prior knowledge of our work and to maximize coverage of user personas, the small number of people interviewed is prone to bias and incomplete thematic saturation. Although we integrated periodic model retraining into our workflow and prototype, we did not account for other methods of mitigating model degradation such as dynamic model updating through detection of calibration drift.[70]

Future work includes conducting summative user research on our design and refining the core feature set and graphical interface through successive iterations of design and evaluation. Additionally, we will explore mobile interfaces, trial custom performance metrics, and visualizations to better support convergent and divergent external validation approaches,[71] and integrate with third-party performance visualization tools. Full development of this tool will pave the way for effective application of advanced techniques such as transfer learning[72,73] and encourage further advancements in harnessing big data to improve the safety and efficacy of clinical care.

Conclusion

We presented a conceptual framework, in the form of a graphical prototype, for an MLOps tool designed to support the entire lifecycle of predictive model development using EHR data. We evaluated the acceptability of this tool among three "user personas" of end users: the health informatician, CIO, and organization's clinical and data stakeholders. Users were able to complete a majority of prompted tasks and agreed that this tool, if built, would fill a niche in the health care analytics landscape. This prototype extends current work in MLOps and predictive modeling to offer concrete design solutions for the implementation and ongoing maintenance of predictive models in healthcare settings.

Clinical Relevance Statement

This MLOps framework can serve as the basis for the development of EHR-integrated predictive model management software. With full deployment and productization, health care systems of any size may be able to use EHR data more efficiently to develop questions, visualize connections, conceptualize and test models, validate external models, track performance over time, and integrate insights into a powerful suite of clinical decision support tools.

Multiple Choice Questions

1. Which of the following is an accurate statement about the use of predictive models in clinical settings?
   a. due to the low cost of computing and storage, most healthcare systems have the financial capacity to develop and implement predictive models
   b. standards for data security and governance in healthcare are more relaxed than those of technology companies that handle private user data, such as Facebook
   c. there is little role for software to support predictive model development as most healthcare systems have experienced analysts
   d. it is important to externally validate predictive models before implementing them as clinical decision support tools, which can be an expensive process

   Correct Answer: The correct answer is option d. It is important to establish the generalizability of predictive models through external validation; this can be an expensive process as data from electronic health record (EHR) systems is heterogeneous and model development and validation frequently requires experienced analysts. Due to Health Insurance Portability and Accountability Act (HIPAA) laws, regulations regarding healthcare data governance are generally more stringent than those of technology companies.

2. Which of the following is an advantage of graphical user interfaces (GUI) compared with a command-line interface?
   a. GUIs have higher user satisfaction than command-line interfaces
   b. GUIs require users to have some technical background to navigate the complex interface
   c. GUIs typically do not require any software installation or setup
   d. GUIs are generally harder to learn for novice users compared with command line interfaces

   Correct Answer: The correct answer is option a. Due to their interactive interface, ability to guide users through different steps of usage, and ability to display visualizations, GUIs have been shown to have higher user satisfaction than command-line interfaces.

Protection of Human and Animal Subjects

This project was deemed exempt from IRB review according to federal and university regulations.

Funding

This study was funded by Advance-CTR Grant (National Institute of Health; grant no.: U54GM115677), 2020 Brown University Summer Research Assistantship Fund.

Conflict of Interest

M.L.R. reports all support from NIGMS (grant no.: U54GM115677) for early phases of this work. S.L.S.

 © 2022. Thieme. All rights reserved.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

reports all support from Brown University Summer Assistantship Fund.

Acknowledgment

We would like to thank Brown Physicians Inc. and Lifespan staff for their valuable feedback during the design and user testing phases of this project.

References

1 Kruse CS, Goswamy R, Raval Y, Marawi S. Challenges and opportunities of big data in health care: a systematic review. JMIR Med Inform 2016;4(04):e38
2 Roski J, Bo-Linn GW, Andrews TA. Creating value in health care through big data: opportunities and policy implications. Health Aff (Millwood) 2014;33(07):1115–1122
3 Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019;20(05):e262–e273
4 Rojas JC, Carey KA, Edelson DP, Venable LR, Howell MD, Churpek MM. Predicting intensive care unit readmission with machine learning using electronic health record data. Ann Am Thorac Soc 2018;15(07):
5 Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. JAMA 2011;306(15):1688–1698
6 Hao S, Wang Y, Jin B, et al. Development, validation and deployment of a real time 30 day hospital readmission risk assessment tool in the maine healthcare information exchange. PLoS One 2015;10(10):e0140271
7 Wu CX, Suresh E, Phng FWL, et al. Effect of a real-time risk score on 30-day readmission reduction in Singapore. Appl Clin Inform 2021;12(02):372–382
8 Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. Crit Care Med 2018;46(04):
9 Teng AK, Wilcox AB. A review of predictive analytics solutions for sepsis patients. Appl Clin Inform 2020;11(03):387–398
10 Machado CDS, Ballester PL, Cao B, et al. Prediction of suicide attempts in a prospective cohort study with a nationally representative sample of the US population. Psychol Med 2021:1–12
11 Oliva EM, Bowe T, Tavakoli S, et al. Development and applications of the Veterans Health Administration's Stratification Tool for Opioid Risk Mitigation (STORM) to improve opioid safety and prevent overdose and suicide. Psychol Serv 2017;14(01):34–49
12 Kwon JM, Jeon KH, Kim HM, et al. Deep-learning-based risk stratification for mortality of patients with acute myocardial infarction. PLoS One 2019;14(10):e0224502
13 Bala W, Steinkamp J, Feeney T, et al. A web application for adrenal incidentaloma identification, tracking, and management using machine learning. Appl Clin Inform 2020;11(04):606–616
14 Li RC, Asch SM, Shah NH. Developing a delivery science for artificial intelligence in healthcare. NPJ Digit Med 2020;3:107
15 Coiera E. The last mile: where artificial intelligence meets reality. J Med Internet Res 2019;21(11):e16323
16 Collins GS, Mallett S, Omar O, Yu L-M. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. BMC Med 2011;9(01):103
17 Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med 2015;13(01):1–10
18 Collins GS, Omar O, Shanyinde M, Yu L-M. A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. J Clin Epidemiol 2013;66(03):268–277
19 Sendak MP, Ratliff W, Sarro D, et al. Real-world integration of a sepsis deep learning technology into routine clinical care: implementation study. JMIR Med Inform 2020;8(07):e15182
20 Schneeweiss S. Learning from big health care data. N Engl J Med 2014;370(23):2161–2163
21 Lee TC, Shah NU, Haack A, Baxter SL. Clinical implementation of predictive models embedded within electronic health record systems: a systematic review. Informatics (MDPI) 2020;7(03):25
22 Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc 2017;24(01):198–208
23 Sculley D, Holt G, Golovin D, et al. Hidden technical debt in machine learning systems. In: Proceedings of the 28th International Conference on Neural Information Processing Systems;2;2503–2511;2015
24 Alahdab M, Çalıklı G Empirical analysis of hidden technical debt patterns in machine learning software. In: Franch X, Männistö T, Martínez-Fernández S, eds. Barcelona, Spain: 20th International Conference, PROFES 2019;2019
25 Cunningham W. The WyCash portfolio management system. In: OOPSLA '92: Addendum to the proceedings on Object-oriented programming systems, languages, and applications (Addendum);29–301992
26 Makinen S, Skogstrom H, Laaksonen E, Mikkonen T. Who needs MLOps: What data scientists seek to accomplish and how can MLOps help? In: 2021 IEEE/ACM 1st Workshop on AI Engineering - Software Engineering for AI (WAIN);2021
27 Karamitsos I, Albarhami S, Apostolopoulos C. Applying DevOps practices of continuous automation for machine learning. Information (Basel) 2020;11(07):363
28 IEEE. Frontiers of data-intensive compute algorithms: sustainable MLOps and beyond. In: 2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC);2020
29 Sato D, Wider A, Windheuser C. Continuous Delivery for Machine Learning. Accessed July 31, 2021 at: https://martinfowler.com/articles/cd4ml.html
30 Tsay J, Mummert T, Bobroff N, Braz A, Westerink P, Hirzel M. Runway: machine learning model experiment management tool. Accessed December 3, 2021: https://mlsys.org/Conferences/doc/2018/26.pdf
31 Hazelwood K, Bird S, Brooks D, Chintala S, Diril U, Dzhulgakov D. Applied Machine Learning at Facebook: A Datacenter Infrastructure Perspective. In: 2018 IEEE International Symposium on High Performance Computer Architecture (HPCA);2018
32 Lamkin T. Kubeflow 1.0: Cloud-Native ML for Everyone - kubeflow - Medium. 2020
33 Katsiapis K, Haas K. Towards ML Engineering with TensorFlow Extended (TFX). In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining;2019
34 Vartak M, Subramanyam H, Lee W-E, et al. ModelDB: a system for machine learning model management. In: HILDA '16: Proceedings of the Workshop on Human-In-the-Loop Data Analytics;2016
35 Chen A, Chow A, Davidson A, et al. Developments in MLflow. In: Proceedings of the Fourth International Workshop on Data Management for End-to-End Machine Learning;2020
36 Kamthan P. Using Personas to Support the Goals in User Stories. In: 2015 12th International Conference on Information Technology - New Generations;2015
37 Negru S, Buraga S. Towards a conceptual model for describing the personas methodology. In: 2012 IEEE 8th International Conference on Intelligent Computer Communication and Processing;2012
38 Liebe JD, Hüsers J, Hübner U. Investigating the roots of successful IT adoption processes - an empirical study exploring the shared

 © 2022. Thieme. All rights reserved.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

awareness-knowledge of Directors of Nursing and Chief Information Officers. BMC Med Inform Decis Mak 2016;16(01):10
39 Benda NC, Das LT, Abramson EL, et al. "How did you get to this number?" Stakeholder needs for implementing predictive analytics: a pre-implementation qualitative study J Am Med Inform Assoc 2020;27(05):709–716
40 Grilo A, Lapao LV, Jardim-Goncalves R, Cruz-Machado V. Challenges for the Development of Interoperable Information Systems in Healthcare Organizations. In: 2009 International Conference on Interoperability for Enterprise Software and Applications China;2009
41 Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Inform Assoc 2018;25(08):969–975
42 Zhu L, Zhang S, Lu Z. Respect for autonomy: seeking the roles of healthcare design from the principle of biomedical ethics. HERD 2020;13(03):230–244
43 Brennan MD, Duncan AK, Armbruster RR, Montori VM, Feyereisn WL, LaRusso NF. The application of design principles to innovate clinical care delivery. J Healthc Qual 2009;31(01):5–9
44 Walden A, Garvin L, Smerek M, Johnson C. User-centered design principles in the development of clinical research tools. Clin Trials 2020;17(06):703–711
45 Jensen TB. Design principles for achieving integrated healthcare information systems. Health Informatics J 2013;19(01):29–45
46 Hummer W, Muthusamy V, Rausch T, et al. ModelOps: cloud-based lifecycle management for reliable and trusted AI. In: 2019 IEEE International Conference on Cloud Engineering (IC2E);2019
47 Feizi A, Wong CY. Usability of user interface styles for learning a graphical software application. In:2012 International Conference on Computer & Information Science (ICCIS);2012
48 Shneiderman B, Plaisant C, Cohen MS, Jacobs SM, Elmqvist N. Designing the User Interface: Strategies for Effective Human-Computer Interaction. Boston, MA: Pearson; 2017
49 Nithya B, Ilango V. Predictive analytics in health care using machine learning tools and techniques. In: 2017 International Conference on Intelligent Computing and Control Systems (ICICCS);2018
50 Cabitza F, Campagner A, Balsano C. Bridging the "last mile" gap between AI implementation and operation: "data awareness" that matters. Ann Transl Med 2020;8(07):501
51 Brooke J. SUS: A quick and dirty usability scale. Accessed December 3, 2021: https://hell.meiert.org/core/pdf/sus.pdf
52 Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008;24(06): 574–594
53 Lewis JR, Brown J, Mayes DK. Psychometric evaluation of the EMO and the SUS in the context of a large-sample unmoderated usability study. Int J Hum Comput Interact 2015;31(08):545–553
54 Lewis JR, Sauro J. The Factor Structure of the System Usability Scale. Accessed December 3, 2021: https://measuringu.-com/papers/Lewis_Sauro_HCII2009.pdf
55 Peres SC, Pham T, Phillips R. Validation of the system usability scale (SUS): SUS in the wild. Proc Hum Fact Ergon Soc Annu Meet 2013;57(01):192–196
56 Blandford A, Furniss D, Makri S. Qualitative HCI research: Going behind the scenes. Synth lect hum-centered inform. 2016 Doi: 10.2200/S00706ED1V01Y201602HCI034
57 Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. Denver, CO: CreateSpace Independent Publishing Platform; 2011
58 Sauro J, Lewis JR. Quantifying the User Experience: Practical Statistics for User Research. Waltham, MA: Morgan Kaufmann; 2016
59 Lee KK-Y, Tang W-C, Choi K-S. Alternatives to relational database: comparison of NoSQL and XML approaches for clinical data storage. Comput Methods Programs Biomed 2013;110(01): 99–109
60 Barrak A, Eghan EE, Adams B. On the co-evolution of ML pipelines and source code - empirical study of DVC projects. In: 2021 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER); 2021
61 Bezanson J, Edelman A, and Karpinski S, and Shah VB. Julia: a fresh approach to numerical computing. SIAM Rev 2017;59(01): 65–98
62 Rossum GDrake FL. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace; 2009
63 R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Accessed December 3, 2021: https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing.
64 Mandl KD, Perakslis ED. HIPAA and the leak of "Deidentified" EHR data. N Engl J Med 2021;384(23):2171–2173
65 Peregrin T. Managing HIPAA compliance includes legal and ethical considerations. J Acad Nutr Diet 2021;121(02):327–329
66 Choi YB, Capitan KE, Krause JS, Streeper MM. Challenges associated with privacy in health care industry: implementation of HIPAA and the security rules. J Med Syst 2006;30(01):57–64
67 Zhou Y, Yu Y, Ding B. Towards MLOps: a case study of ML pipeline platform. In: 2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE); Beijing, China;2020
68 Wong A, Otles E, Donnelly JP, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Intern Med 2021;181(08):1065–1070
69 Purkayastha S, Trivedi H, Gichoya JW. Failures hiding in success for artificial intelligence in radiology. J Am Coll Radiol 2021;18(3, pt. B, 3, pt. B):517–519
70 Davis SE, Greevy RA Jr., Lasko TA, Walsh CG, Matheny ME. Detection of calibration drift in clinical prediction models to inform model updating. J Biomed Inform 2020;112:103611
71 Ho SY, Phua K, Wong L, Bin Goh WW. Extensions of the external validation for checking learned model interpretability and generalizability. Patterns (N Y) 2020;1(08):100129
72 Zhuang F, Qi Z, Duan K, et al. A comprehensive survey on transfer learning. Proc IEEE 2021;109(01):43–76
73 Gupta P, Malhotra P, Narwariya J, Vig L, Shroff G. Transfer learning for clinical time series analysis using deep neural networks. J Healthc Inform Res 2020;4(02):112–137

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.