16

Lived Experience

Recall Sospital, the leading (fictional) health insurance company in the United States that tried to transform its care management system in Chapter 10 with fairness as a top concern. Imagine that you are a project manager at Sospital charged with reducing the misuse of opioid pain medications by members. An opioid epidemic began in the United States in the late 1990s and is now at a point that over 81,000 people die per year from opioid overdoses. As a first step, you analyze member data to understand the problem better. Existing machine learning-based opioid overdose risk models trained on data from state prescription drug monitoring programs (which may include attributes originating in law enforcement databases) have severe issues with data quality, consent, bias, interpretability, and transparency.[1] Also, the existing risk models are predictive machine learning models that can easily pick up spurious correlations instead of causal models that do not. So you don't want to take the shortcut of using the existing models. You want to start from scratch and develop a model that is trustworthy. Once you have such a model, you plan to deploy it to help human decision makers intervene in fair, responsible, and supportive ways.

You are starting to put a team together to carry out the machine learning lifecycle for the opioid model. You have heard the refrain that diverse teams are better for business.[2] For example, a 2015 study found that the top quartile of companies in gender and racial/ethnic diversity had 25% better financial performance than other companies.[3] In experiments, diverse teams have focused more on facts and been more innovative.[4] But do diverse teams create better, less biased, and more trustworthy machine

[1] Maia Szalavitz. "The Pain Was Unbearable. So Why Did Doctors Turn Her Away?" In: *Wired* (Aug. 2021).

[2] Among many other points that are used throughout this chapter, Fazelpour and De-Arteaga emphasize that the business case view on diversity is problematic because it takes the *lack* of diversity as a given and burdens people from marginalized groups to justify their presence. Sina Fazelpour and Maria De-Arteaga. "Diversity in Sociotechnical Machine Learning Systems." arXiv:2107.09163, 2021.

[3] The study was of companies in the Americas and the United Kingdom. Vivian Hunt, Dennis Layton, and Sara Prince. "Why Diversity Matters." McKinsey & Company, Jan. 2015.

[4] David Rock and Heidi Grant. "Why Diverse Teams are Smarter." In: *Harvard Business Review* (Nov. 2016).

learning models?[5] How and why? What kind of diversity are we talking about? In which roles and phases of the machine learning lifecycle is diversity a factor?

> "I believe diversity in my profession will lead to better technology and better benefits to humanity from it."
>
> —Andrea Goldsmith, electrical engineer at Stanford University

The first question is whether the team even affects the models that are created. Given the same data, won't all skilled data scientists produce the same model and have the same inferences? A real-world experiment assigned 29 teams of skilled data scientists an open-ended causal inference task of determining whether soccer referees are biased against players with dark skin, all using exactly the same data.[6] Due to different subjective choices the teams made in the problem specification and analysis, the results varied. Twenty teams found a significant bias against dark-skinned players, which means that nine teams did not. In another real-world example, 25 teams of data scientists developed mortality prediction models from exactly the same health data and had quite variable results, especially in terms of fairness with respect to race and gender.[7] In open-ended lifecycles, models and results may depend a lot on the team.

If the team matters, what are the characteristics of the team that matter? What should you be looking for as you construct a team for modeling individual patients' risk of opioid misuse? Let's focus on two team characteristics: (1) *information elaboration*: how do team members work together, and (2) *cognitive*: what do individual team members know. In the first characteristic: information elaboration, socioculturally non-homogeneous teams are more likely to slow down and consider critical and contentious issues; they are less apt to take shortcuts.[8] Such a slowdown is not prevalent in homogeneous teams and importantly, does not depend on the team members having different sets of knowledge. All of the team members could know the critical issues, but still not consider them if the members are socioculturally homogeneous.

You have probably noticed quotations sprinkled throughout the book that raise issues relevant to the topic of a given section. You may have also noticed that the people quoted have different sociocultural backgrounds, which may be different than yours. This is an intentional feature of the book. Even if they are not imparting knowledge that's different from the main text of the book, the goal of the quotes is for you to hear these voices so that you are pushed to slow down and not take shortcuts. (Not taking shortcuts is a primary theme of the book.)

[5] Caitlin Kuhlman, Latifa Jackson, and Rumi Chunara. "No Computation Without Representation: Avoiding Data and Algorithm Biases Through Diversity." arXiv:2002.11836, 2020.

[6] Raphael Silberzahn and Eric L. Uhlmann. "Crowdsourced Research: Many Hands Make Tight Work." In: *Nature* 526 (Oct. 2015), pp. 189–191.

[7] Timothy Bergquist, Thomas Schaffter, Yao Yan, Thomas Yu, Justin Prosser, Jifan Gao, Guanhua Chen, Łukasz Charzewski, Zofia Nawalany, Ivan Brugere, Renata Retkute, Alidivinas Prusokas, Augustinas Prusokas, Yonghwa Choi, Sanghoon Lee, Junseok Choe, Inggeol Lee, Sunkyu Kim, Jaewoo Kang, Patient Mortality Prediction DREAM Challenge Consortium, Sean D. Mooney, and Justin Guinney. "Evaluation of Crowdsourced Mortality Prediction Models as a Framework for Assessing AI in Medicine." medRxiv:2021.01.18.21250072, 2021.

[8] Daniel Steel, Sina Fazelpour, Bianca Crewe, and Kinley Gillette. "Information Elaboration and Epistemic Effects of Diversity." In: *Synthese* 198.2 (Feb. 2021), pp. 1287–1307.

Sociocultural differences are associated with differences in *lived experience* of marginalization.[9] Remember from Chapter 1 that lived experience is the personal knowledge you have gained through direct involvement in things from which you have no option to escape. Related to the second characteristic of the team: what the team members know, one key cognitive theory relevant for trustworthy machine learning is that people with lived experience of marginalization have an *epistemic advantage*: when people reflect on their experience of being oppressed, they are better able to understand all sides of power structures and decision-making systems than people who have not been oppressed.[10] Briefly mentioned in Chapter 4, they have a bifurcated consciousness that allows them to walk in the shoes of both the oppressed and the powerful. In contrast, privileged people tend to have blind spots and can only see their own perspective.

> "People with marginalized characteristics—so people who had experienced discrimination—had a deeper understanding of the kinds of things that could happen to people negatively and the way the world works in a way that was a bit less rosy."
>
> —Margaret Mitchell, research scientist at large
>
> "The lived experiences of those directly harmed by AI systems gives rise to knowledge and expertise that must be valued."
>
> —Emily Denton, research scientist at Google
>
> "Technical know-how cannot substitute for contextual understanding and lived experiences."
>
> —Meredith Whittaker, research scientist at New York University

In modern Western science and engineering, knowledge derived from lived experience is typically seen as invalid; often, only knowledge obtained using the scientific method is seen as valid. This contrasts with critical theory, which has knowledge from the lived experience of marginalized people at its very foundation. Given the many ethics principles founded in critical theory covered in Chapter 15, it makes sense to consider lived experience in informing your development of a model for opioid misuse risk. Toward this end, in the remainder of the chapter, you will:

- map the cognitive benefit of the lived experience of team members to the needs and requirements of different phases of the machine learning lifecycle, and
- formulate lifecycle roles and architectures that take advantage of that mapping.

[9] Neurodiversity is not touched upon in this chapter, but is another important dimension that could be expanded upon.

[10] Natalie Alana Ashton and Robin McKenna. "Situating Feminist Epistemology." In: *Episteme* 17.1 (Mar. 2020), pp. 28–47.

16.1 Lived Experience in Different Phases of the Lifecycle

The first stage in the lifecycle of developing an opioid misuse risk model is problem specification and value alignment. In this phase, there is a clear need for the inclusion of people with different lived experiences to question assumptions and identify critical issues in the four levels of value alignment covered in Chapter 14: whether you should work on the problem, which pillars of trustworthiness are of concern, how to measure performance in those pillars, and acceptable ranges of metrics. The epistemic advantage of these team members is critical in this phase. The blind spots of team members who have not experienced systematic disadvantages will prevent them from noticing all the possible misuses and harms that can arise from the system, such as undue denials of care to traditionally marginalized individuals. This is the phase in which participatory design, also covered in Chapter 14, should be used.[11]

> "New perspectives ask new questions and that's a fact. This is exactly why inclusion matters!"
>
> —Deb Raji, fellow at Mozilla

The second phase, data understanding, requires digging into the available data and its provenance to identify the possible bias and consent issues detailed in Chapter 4 and Chapter 5. This is another phase in which it is important for the team to be critical, and it is useful to have members with epistemic advantage. In Chapter 10, we already saw that the team developing the Sospital care management system needed to recognize the bias against African Americans when using health cost as a proxy for health need. Similarly, a diagnosis for opioid addiction in a patient's data implies that the patient actually interacted with Sospital for treatment, which will also be biased against groups that are less likely to utilize the health care system. Problem owners, stakeholders, and data scientists from marginalized groups are more likely to recognize this issue. Furthermore, a variety of lived experiences will help discover that large dosage opioid prescriptions from veterinarians in a patient's record are for their pets, not for them; prescription claims for naltrexone, an opioid itself, represent treatment for opioid addiction, not evidence of further misuse; and so on.

The third phase in developing an opioid misuse model is data preparation. You can think of data preparation in two parts: (1) data integration and (2) feature engineering. Critique stemming from lived experience has little role to play in data integration because of its mechanical and rote nature. Is this also the case in the more creative feature engineering part? Remember from Chapter 10 that biases may be introduced in feature engineering, such as by adding together different health costs to create a single column. Such biases may be spotted by team members who are advantaged in looking for them. However, if dataset constraints, such as dataset fairness metric constraints, have already been included in the problem specification of the opioid misuse model in anticipation of possible harms, then no additional epistemic advantage is needed to spot the issues. Thus, there is less usefulness of lived experience of marginalization among team members in the data preparation stage of the lifecycle.

[11] Vinodkumar Prabhakaran and Donald Martin Jr. "Participatory Machine Learning Using Community-Based System Dynamics." In: *Health and Human Rights Journal* 22.2 (Dec. 2020), pp. 71–74.

In the fourth phase of the lifecycle, the team will take the prepared data and develop an individualized causal model of factors that lead to opioid misuse.[12] Coming after the problem specification phase that sets forth the modeling task and the performance metrics, and after the data understanding and data preparation phases that finalize the dataset, the modeling phase is not open-ended like the soccer referee and mortality prediction tasks described in the previous section. The modeling is quite constrained from the perspective of the data scientist.

A recent study tasked 399 data scientists, each working alone, with developing models of the mathematical literacy of people based on approximately five hundred of their biographical features; the dataset and basic performance metrics were clearly specified (no fairness metric was specified).[13] Importantly, the dataset had many data points and was purposefully and carefully collected as a representative sample without population biases. Thus, the dataset had negligible epistemic uncertainty. The study analyzed the 399 models that were created and found no significant relationship between the unwanted bias of the models and the sociocultural characteristics of the data scientists that produced them.

In this example and other similar regimented and low-epistemic uncertainty modeling tasks, the lived experience of the team is seemingly of low importance. In contrast, when there is great epistemic uncertainty like you may have in analyzing opioid abuse, the inductive bias of the model chosen by the data scientist has a great role to play and the lived experience of the data scientist can become important. However, mirroring the argument made earlier about an explicit problem specification lessening the epistemic advantage for members of marginalized groups in feature engineering, a clear specification of all relevant trust metric dimensions also lessens the usefulness of lived experience in modeling.

Evaluating the opioid risk model once it has been created is not as straightforward as simply testing it for the specified allowable trust metric ranges in the ways described in Chapter 14. Once a model is tangible, you can manipulate it in various ways and better imagine the harms it could lead to. Thus, being critical of the model during evaluation is also a job better done by a team that has members who have experienced systematic disadvantage and are attuned to the negative impacts it may have if it is deployed within Sospital's operations.

Finally, if the model has passed the evaluation stage, the ML operations engineers on the team carry out the deployment and monitoring phase of the lifecycle. Their role is primarily to ensure technical integration with Sospital's other systems and noting when the trust metric ranges elicited during value alignment are violated over time. This is another phase of the lifecycle in which there is not much epistemic advantage to be had by a team containing engineers with lived experience of marginalization.

Overall, as shown in Figure 16.1, three lifecycle phases (problem specification, data understanding, and evaluation) can take advantage of having a diverse team containing members that have lived experience of marginalization. The other three phases (data preparation, modeling, and deployment and monitoring) benefit less from the epistemic advantage of team members with lived experience of

[12] Chirag Nagpal, Dennis Wei, Bhanukiran Vinzamuri, Monica Shekhar, Sara E. Berger, Subhro Das, and Kush R. Varshney. "Interpretable Subgroup Discovery in Treatment Effect Estimation with Application to Opioid Prescribing Guidelines." In: *Proceedings of the ACM Conference on Health, Inference, and Learning*. Apr. 2020, pp. 19–29.

[13] Bo Cowgill, Fabrizio Dell'Acqua, Samuel Deng, Daniel Hsu, Nakul Verma, and Augustin Chaintreau. "Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics." In: *Proceedings of the ACM Conference on Economics and Computation*. Jul. 2020, pp. 679–681.

systematic harm. This conclusion suggests a particular lifecycle architecture for developing your opioid risk model, discussed in the next section.

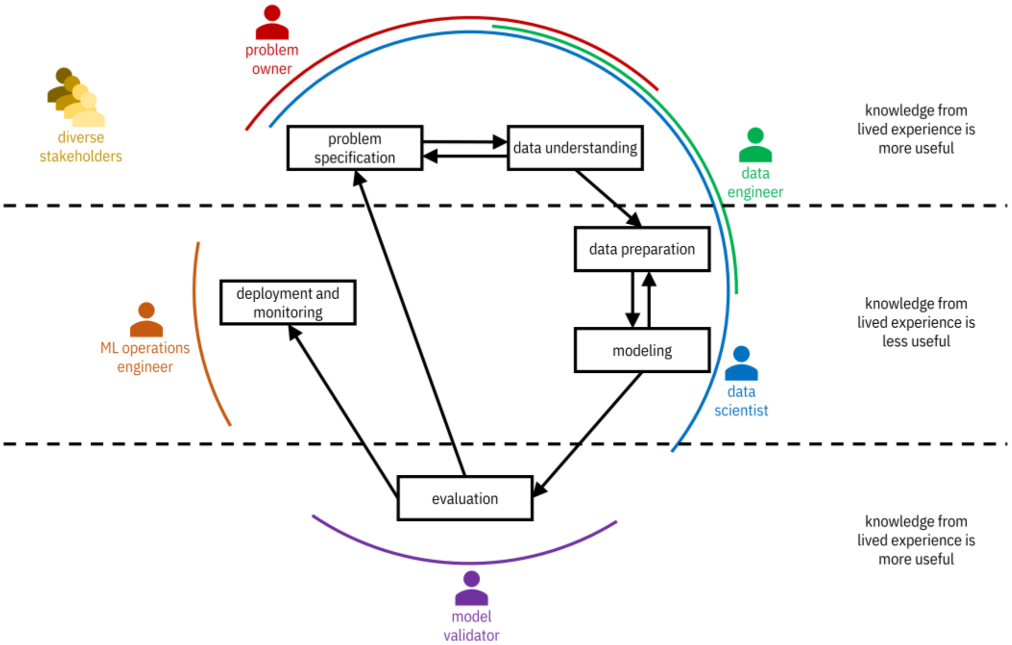

Figure 16.1. *The different phases of the machine learning lifecycle delineated by how useful knowledge from lived experience is. Knowledge from lived experience is more useful in problem specification, data understanding, and evaluation. Knowledge from lived experience is less useful in data preparation, modeling, and deployment and monitoring.* Accessible caption. A diagram of the development lifecycle is marked according to which phases find lived experience more useful and less useful.

16.2 Inclusive Lifecycle Architectures

From the previous section, you have learned that having a diverse team with lived experience of systematic harms is most important in the problem specification, data understanding, and evaluation phases. These phases coincide with the phases in which the problem owner and model validator personas are most prominent. Problem owners often tend to be subject matter experts about the application and are not usually skilled at the technical aspects of data engineering and modeling. They may or may not come from marginalized backgrounds. Given the power structures in place at many corporations, including Sospital, the problem owners often come from privileged backgrounds. Even if that is not true at Sospital, it is strongly suggested to have a panel of diverse voices, including those from marginalized groups, participate and be given a voice in these phases of the lifecycle.

That leaves the other three phases. What about them? The analysis suggests that as long as the specification and validation are done with the inclusion of team members and panelists with lived experience of oppression,[14] then any competent, reliable, communicative, and selfless data engineers, data scientists, and ML operations engineers equipped with the tools and training in trustworthy machine learning will create a trustworthy opioid misuse risk model irrespective of their lived experience. The pool of skilled data scientists at Sospital does not include many individuals with lived experience, and you also don't want to levy a 'minority tax'—the burden of extra responsibilities placed on minority employees in the name of diversity—on the ones there are. So you go with the best folks available, and that is perfectly fine. (Machine learning researchers creating the tools for practitioners should have a variety of lived experiences because researchers have to both pose and answer the questions. Fortuitously, though their numbers are small overall, researchers from groups traditionally underrepresented in machine learning and associated with marginalization are seemingly overrepresented in research on *trustworthy* machine learning, as opposed to other areas of machine learning research.[15])

If the lived experience of the data scientists and engineers on the team is less relevant for building trustworthy machine learning systems, what if the data scientists and engineers are not living beings at all? Technology advances are leading to a near-future state in which feature engineering and modeling will be mostly automated, using so-called auto ML. Algorithms will construct derived features, select hypothesis classes, tune hyperparameters of machine learning algorithms, and so on. As long as these auto ML algorithms are themselves trustworthy,[16] then it seems as though they will seamlessly enter the lifecycle, interact with problem owners and model validators, and successfully create a trustworthy model for opioid misuse.

Shown in Figure 16.2, in this near-future, auto ML instead of data scientists is the controller in the control theory perspective on governance introduced in Chapter 14 and Chapter 15. And this is a-okay. Such an architecture involving auto ML empowers problem owners and marginalized communities to pursue their goals without having to rely on scarce and expensive data scientists. This architecture enables more democratized and accessible machine learning for Sospital problem owners when paired with *low-code/no-code* interfaces (visual software development environments that allow users to create applications with little or no knowledge of traditional computer programming).

> "It's about humans at the center, it's about those unnecessary barriers, where people have domain expertise but have difficulty teaching the machine about it."
>
> —Christopher Re, computer scientist at Stanford University

[14] Those specifications and validations must also be given true power. This point is discussed later using the terminology 'participation washing'.

[15] Yu Tao and Kush R. Varshney. "Insiders and Outsiders in Research on Machine Learning and Society." arXiv:2102.02279, 2021.

[16] Jaimie Drozdal, Justin Weisz, Dakuo Wang, Gaurav Dass, Bingsheng Yao, Changruo Zhao, Michael Muller, Lin Ju, and Hui Su. "Trust in AutoML: Exploring Information Needs for Establishing Trust in Automated Machine Learning Systems." In: *Proceedings of the International Conference on Intelligent User Interfaces*. Mar. 2020, pp. 297–307.

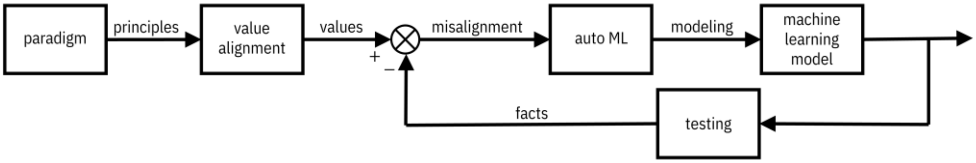

Figure 16.2. *The control theory perspective of AI governance with auto ML technologies serving as the controller instead of data scientists.* Accessible caption. A block diagram that starts with a paradigm block with output principles. Principles are input to a value alignment block with output values. Facts are subtracted from values to yield misalignment. Misalignment is input to an auto ML block with modeling as output. Modeling is input to a machine learning model with output that is fed into a testing block. The output of testing is the same facts that were subtracted from values, creating a feedback loop.

A recent survey of professionals working within the machine learning lifecycle asked respondents their preference for auto ML in different lifecycle phases.[17] The respondents held different lifecycle personas. The preferred lifecycle phases for automation were precisely those in which lived experience is less important: data preparation, modeling, and deployment and monitoring. The phases of the lifecycle that respondents did not care to see automation take hold were the ones where lived experience is more important: problem specification, data understanding, and evaluation. Moreover, respondents from the problem owner persona desired the greatest amount of automation, probably because of the empowerment it provides them. These results lend further credence to an architecture for machine learning development that emphasizes inclusive human involvement in the 'takeoff' (problem specification and data understanding) and 'landing' (evaluation) phases of the lifecycle while permitting 'auto pilot' (auto ML) in the 'cruising' (data preparation and modeling) phase.

Another recent survey showed that machine learning experts were more likely to call for strong governance than machine learning non-experts.[18] This result suggests that problem owners may not realize the need for explicit value alignment in an automated lifecycle. Therefore, the empowerment of problem owners should be only enabled in architectures that place the elicitation of paradigms and values at the forefront.

> "Participation-washing could be the next dangerous fad in machine learning."
>
> —Mona Sloane, sociologist at New York University

Before concluding the discussion on inclusive lifecycle architectures, it is important to bring up *participation washing*—uncredited and uncompensated work by members of marginalized groups.[19]

[17] Dakuo Wang, Q. Vera Liao, Yunfeng Zhang, Udayan Khurana, Horst Samulowitz, Soya Park, Michael Muller, and Lisa Amini. "How Much Automation Does a Data Scientist Want?" arXiv:2101.03970, 2021.

[18] Matthew O'Shaughnessy, Daniel Schiff, Lav R. Varshney, Christopher Rozell, and Mark Davenport. "What Governs Attitudes Toward Artificial Intelligence Adoption and Governance?" osf.io/pkeb8, 2021.

[19] Mona Sloane, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. "Participation is Not a Design Fix for Machine Learning." arXiv:2007.02423, 2020. Bas Hofstra, Vivek V. Kulkarni, Sebastian Munoz-Najar Galvez, Bryan He, Dan Jurafsky, and Daniel A. McFarland. "The Diversity–Innovation Paradox in Science." In: *Proceedings of the National Academy of Sciences of the United States of America* 117.17 (Apr. 2020), pp. 9284–9291.

Participatory design sessions that include diverse voices, especially those with lived experience of marginalization, have to be credited and compensated. The sessions are not enough if they turn out to be merely for show. The outcomes of those sessions have to be backed by power and upheld throughout the lifecycle of developing the opioid abuse model. Otherwise, the entire architecture falls apart and the need for team members with lived experience returns to all phases of the lifecycle.

Leaving aside the difficult task of backing the inputs of marginalized people with the power they need to be given, how should you even go about bringing together a diverse panel? From a practical perspective, what if you are working under constraints?[20] Broad advertising and solicitations from entities that vulnerable people don't know may not yield many candidates. More targeted recruitment in specific social media groups and job listing sites may be somewhat better, but will still miss certain groups. Unfortunately, there are no real shortcuts. You have to develop relationships with institutions serving different communities and with members of those communities. Only then will you be able to recruit people to participate in the problem specification, data understanding, and evaluation phases (either as employees or simply as one-time panelists) and be able to do what you know that you should.

16.3 Summary

- The model produced in a machine learning lifecycle depends on characteristics of the team.
- Teams that are socioculturally heterogeneous tend to slow down and not take shortcuts.
- Team members with lived experience of marginalization have an epistemic advantage in noticing potential harms.
- This epistemic advantage from lived experience is most important in the problem specification, data understanding, and evaluation stages of the lifecycle. It is less important in the data preparation, modeling, and deployment and monitoring stages.
- A sensible architecture for the lifecycle focuses on inclusion of team members with lived experience of systematic harm in the three phases in which they have epistemic advantage.
- The other three phases may be sensibly carried out by trustworthy data scientists and engineers, or even trustworthy auto ML algorithms, which may be empowering for problem owners.

[20] Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. "Stakeholder Participation in AI: Beyond 'Add Diverse Stakeholders and Stir.'" In: *Proceedings of the NeurIPS Human-Centered AI Workshop*. Dec. 2021.