# Broadening Data Science Education for the Future Biomanufacturing Workforce

# Workshop Report

December 2023

# Table of Contents

# Executive Summary

Biomanufacturing is a rapidly evolving field and becoming a critical area for innovative drug and therapeutic development. Novel manufacturing processes and equipments yield substantial amounts of data that can support the scale-up of processes and technologies from the lab to commercial manufacturing. Artificial intelligence (AI), machine learning (ML), and predictive modeling can facilitate the analysis and manipulation of this data to inform process development, facility design, and real-time changes in manufacturing processes. **The biomanufacturing industry needs a workforce that possesses both the skills and knowledge of biomanufacturing as well as data analysis and data science concepts. However, a significant portion of the current biomanufacturing workforce does not possess sufficient skills in this area, which is a significant barrier to growing the industry.**

The further growth of the industry will depend on the integration of data science and data analysis into current and new biomanufacturing curriculum to equip the incoming workforce with the skills that industry needs. To assist in addressing this need, the Georgia Institute of Technology (Georgia Tech) is working to establish a Future Manufacturing Network (FMNet) that will support national bio, cell, and tissue manufacturing activities across the field of biopharmaceutical manufacturing.

This report summarizes the work that FMNet has done to date—with a focus on a competency model that maps needed data science skills at different levels of education and different roles in the biopharmaceutical manufacturing workforce. This model is based on the results of a two-day virtual workshop in 2021, a follow-on survey, and a set of focus groups in 2023 to further refine the content.

## Needed Skills for the Biomanufacturing Workforce

Due to the dynamic and interdisciplinary nature of the biomanufacturing field, skills needs vary across the diverse people involved in the workforce. While domain experts may need specialized skills and knowledge, the average worker in a biomanufacturing company would not necessarily require the same level of expertise.

A high-level summary of different skill areas and the level of that skill needed for each worker type is depicted in the table below. The levels of skills are defined as follows:

- **Basic**: an individual who has taken a single high school or collegiate class on the topic, or who has relevant on-the-job training or any equivalent experience.
- **Proficient**: an individual with at least a university minor certificate related to the topic or equivalent experience.
- **Advanced**: an individual with an undergraduate degree or equivalent experience.
- **Expert**: an individual with an advanced degree and industry experience, or equivalent experience.

### Table 1. Rollup of Needed Skills in Areas of Biomanufacturing by Worker Type

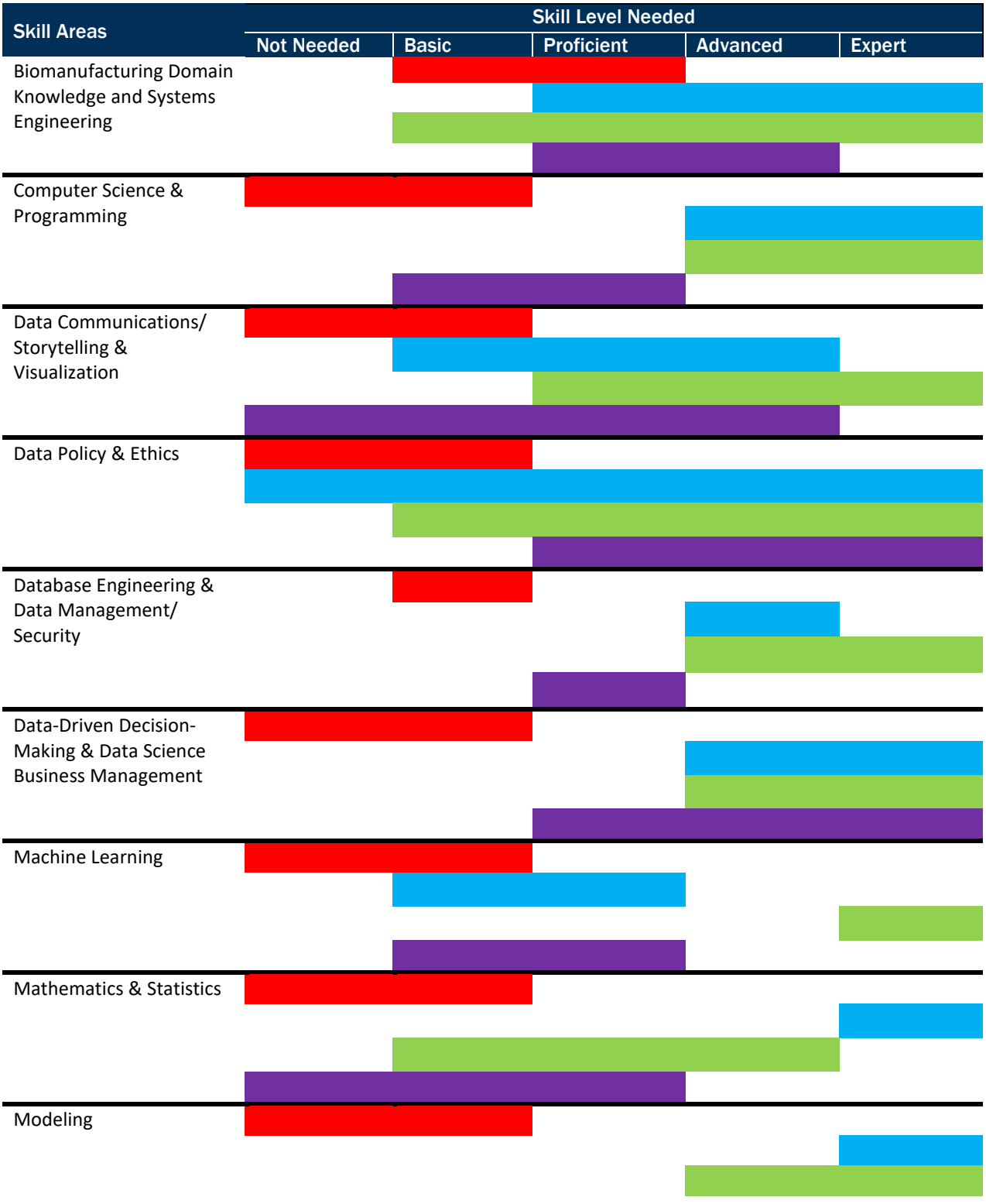
Legend:

= All biomanufacturing workers

= Workers with domain expertise*

= Data science specialists

= Managers and leadership



| Skill Areas | Skill Level Needed | | | | |
|---|---|---|---|---|---|
| | Not Needed | Basic | Proficient | Advanced | Expert |

**The domain for these experts will vary based on the skill area.*

## Educational Programs

In an effort to identify the skills gaps of the current workforce, FMNet developed a landscape assessment of existing education programs that address these critical skill areas. A high-level summary of the coverage of existing programs by skill area and by education level is depicted in the table below.

### Table 2. Rollup of Existing Educational Programs in Areas of Biomanufacturing

| Education Level | Biomanufacturing Domain Knowledge and Systems Engineering | Computer Science & Programming | Data Communications/ Storytelling & Visualization | Data Policy & Ethics | Database Engineering & Data Management/ Security | Data-Driven Decision-Making & Data Science Business Management | Machine Learning | Mathematics & Statistics | Modeling |
|---|---|---|---|---|---|---|---|---|---|
| K-12 | | | | | | | | | |
| Technical Colleges | | | | | | | | | |
| Undergraduate | | | | | | | | | |
| Graduate | | | | | | | | | |
| Continuing Education | | | | | | | | | |

This table shows that there are **significant gaps in existing curricula for biomanufacturing concepts**. Current programs offered at varying education levels do not sufficiently build the skills that the incoming workforce needs to succeed in the biomanufacturing industry. In addition, this list is incomplete; there may be a lack of broad awareness of other programs that may exist which would fill these gaps. Ultimately, the level of coverage of existing programs and the gaps identified make the case for additional curriculum development—the goal of FMNet.

## Activities

Over the next several years, FMNet will be prioritizing activities to address the gaps identified in the landscape assessment. These activities will encompass the following high-level objectives:

- Foundational Knowledge of Data Science Concepts
- Program or Course Design
- Dataset Curation and Access
- Learning and Connection Opportunities

- Governance and Decision Making

# Background

The Georgia Institute of Technology (Georgia Tech) is working to identify concrete actions to enable the future state of biomanufacturing. Their efforts build on the work of the National Cell Manufacturing Consortium (NCMC) and the strategy outlined in the 2016 National Roadmap (and the subsequent 2017 and 2019 updates, which can be found here), which identified the current challenges with advancing cell and gene therapy manufacturing into the future, as well as activities to address these challenges.

To help address the challenges identified in that roadmap, Georgia Tech is establishing the Future Manufacturing Network, or FMNet, which will complement national bio, cell, and tissue manufacturing activities supported through the National Institute for Innovation in Manufacturing of Biopharmaceuticals (NIIMBL), the Advanced Regenerative Manufacturing Institute's (ARMI) BioFabUSA, and the NSF Engineering Research Centers (ERCs) for Cell Manufacturing Technologies (CMaT) and Cellular Metamaterials (Cell-Met), which will cover Data Science, AI, and Predictive Modeling.

FMNet is focusing early efforts on identifying skills gaps in the current biomanufacturing workforce related to data science. By defining the skills needed to advance data science capabilities in biomanufacturing, determining existing education programs that could help address these gaps, and identifying activities to address any remaining gaps, FMNet will develop a competency model that can be used to inform programs in biomanufacturing and data science. Figure 1 below outlines this process.

### Figure 1. Overview Graphic


One of FMNet's first activities was to convene experts from industry and academia in a **two-day workshop** to define unmet workforce needs for data science, artificial intelligence (AI), and machine learning (ML) as applied to biomanufacturing; to determine existing education programs that could

address these currently unmet needs; and to identify activities and opportunities for collaboration to further address industry needs for data science, AI, and ML applied to biomanufacturing.

Following the workshop, participants were asked via a **survey** to answer questions about existing educational programs and necessary skillsets for advancing the integration of data science and AI in biomanufacturing, which were captured from the workshop. Based on the survey responses, FMNet held **four virtual focus group sessions** in July and August 2023 to validate and refine this content.

This report consolidates all of the outputs from the discussion that took place on the second day of the workshop, the survey results, and the discussion from the focus group sessions.

# Needed Skills

The first step in building a competency model is to identify the skills and knowledge that are critical for applying data science to biomanufacturing. Many of these skills require highly educated or experienced biomanufacturing or data science specialists and experts and would be addressed by different members of a team of workers. Participants additionally identified **priority skill or knowledge areas** which may not be adequately addressed by existing educational programs or training opportunities.

Outlined below is a highlight of the priority skills identified by workshop participants, grouped by competency area; an expanded list of all critical data science skills identified by participants; and a general description of the way that each competency area impacts biomanufacturing process development. Critical skills and knowledge were also assessed for whether they were required primarily for data science specialists, workers with "domain experience"—with the appropriate "domain" defined by each focus group—leadership and management, or all biomanufacturing workers.

★ Skill or knowledge area identified by participants as a **priority need** in the biomanufacturing workforce

## Required Proficiency Levels

The definitions for each proficiency level—Basic, Proficient, Advanced, and Expert—are outlined below:

- **Basic**: an individual who has taken a single high school or collegiate class on the topic, or who has relevant on-the-job training or any equivalent experience.
- **Proficient**: an individual with at least a university minor certificate related to the topic or equivalent experience.
- **Advanced**: an individual with an undergraduate degree or equivalent experience.
- **Expert**: an individual with an advanced degree and industry experience, or equivalent experience.

## Biomanufacturing Domain Knowledge & Systems Engineering (Crosscutting)

Biomanufacturing is a rapidly evolving field that draws on expertise from a variety of more traditional disciplines. Knowledge of how to engineer and model biological systems is particularly crucial for scaling up biomanufacturing processes. This knowledge area represents multidisciplinary skills as well as biomanufacturing-specific skills needed for product development and manufacturing at scale. These skills do not necessarily require specialized instruction or deep knowledge of data science but do involve

some form of data analysis and action based on the data. The other knowledge areas detailed in this report are all applications of data science and data analysis that tie back to biomanufacturing domain knowledge and systems engineering.

## Skills by Worker

All types of workers require at least a basic understanding of the biomanufacturing domain and systems engineering. They should also all be proficient in data management and documentation, as well-curated datasets are critical for transitioning to larger-scale manufacturing operations and monitoring quality, safety, efficacy, and critical quality attributes (CQAs).

### *Average Biomanufacturing Industry Worker*

The average worker in the biomanufacturing industry should possess the following skills:

- An understanding of how to find and access data
- An understanding of how to comply with regulatory guidelines to ensure product approval
- The ability to correlate datasets to identify needed data or analyses

### *Workers with Domain Expertise*

Workers with expertise in a biomanufacturing domain area should possess the following skills:

- The ability to identify CQAs and engineer systems to ensure their consistency
- A deep understanding of cell biology and the concepts behind advanced therapeutics
- Expertise in understanding and correlating data created throughout the manufacturing process

### *Data Science Specialists*

Data science specialists do not need a high level of understanding of actual cell biology and aseptic techniques, as this is where workers involved in the production of the therapeutic products are expected to have more understanding and practice. However, they should possess:

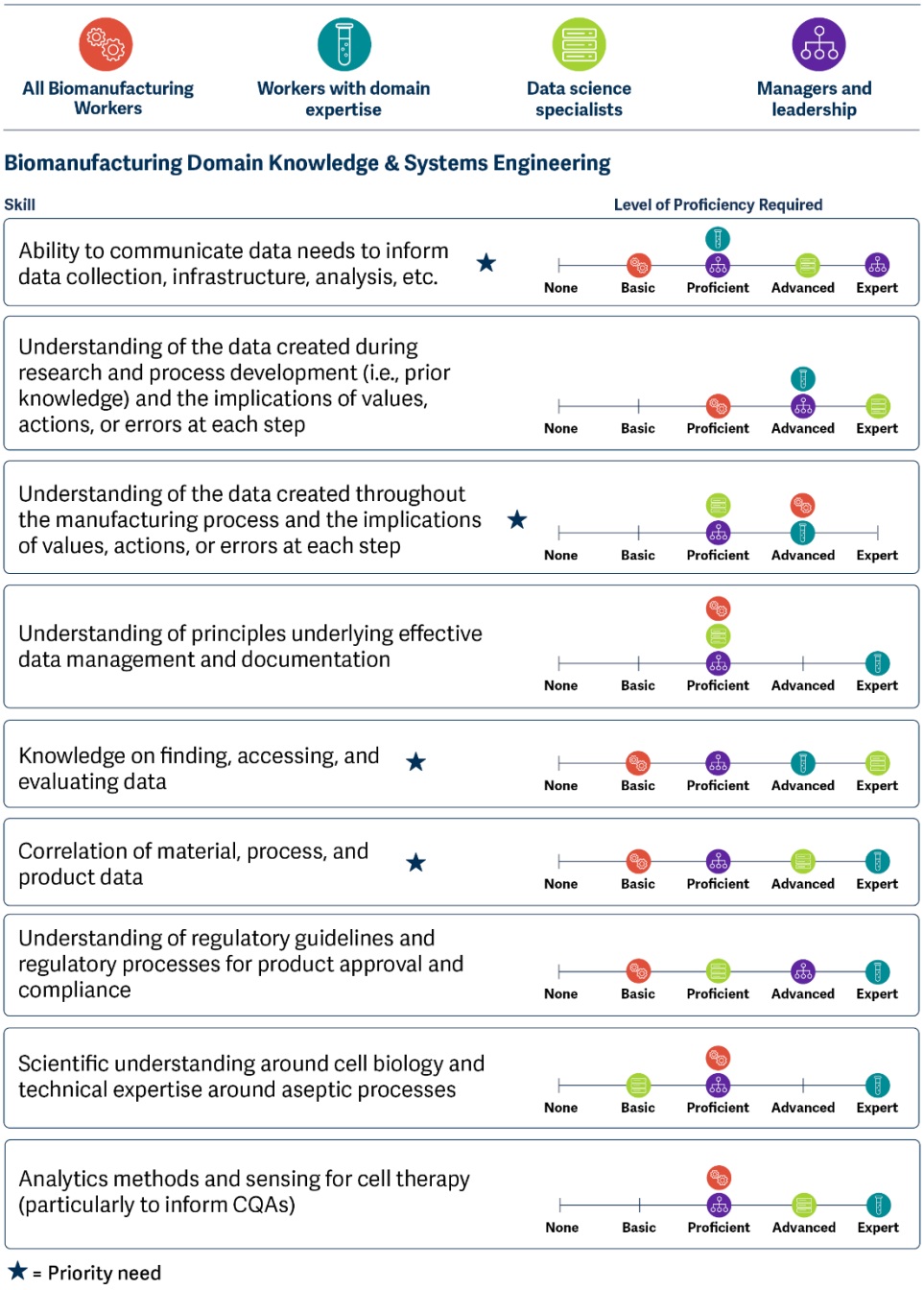
- Expertise in accessing and evaluating the different types of data generated during the R&D phase of product development

### *Managers and Leadership*

Managers and leadership should possess the following skills:

- General proficiency in biomanufacturing domains
- The ability to make decisions and instruct workers on how to follow processes and protocols for maintaining regulatory compliance and ensuring product safety
- An advanced understanding of regulatory requirements
- The ability to communicate with workers and resolve infrastructure needs to collect and process the data involved in R&D and systems engineering (note: leadership would need expertise in this area, but lower-level managers would not; this distinction is shown in the table below with two icons for managers and leadership)

## Figure 2. Needed Biomanufacturing Domain Skills Mapped to Types of Workers Who Need Them



## Computer Science & Programming

Large-scale biomanufacturing relies on automation of both production processes and handling processes for extremely large datasets. This knowledge area represents critical applications of computer science and programming to control biomanufacturing systems. While concepts such as data analysis and simulations that relate to computer science and programming overlap with modeling, machine learning, and data-based engineering and management, computer science and programming specifically refers to software, programming languages, and computer infrastructure.

### Skills by Worker

#### *Average Biomanufacturing Industry Worker*

The average worker in the biomanufacturing industry does not need knowledge of programming languages, documentation practices, or how computational infrastructure works. They should know:

- How to use the software needed for their role
- How to know when there is a problem with the software so they can notify an expert to fix it

#### *Workers with Domain Expertise*

Workers with expertise in the computer science and programming domain area should possess the following skills:

- The ability to debug code and revise programs
- Advanced knowledge in writing software

#### *Data Science Specialists*

Data science specialists should possess the following skills:

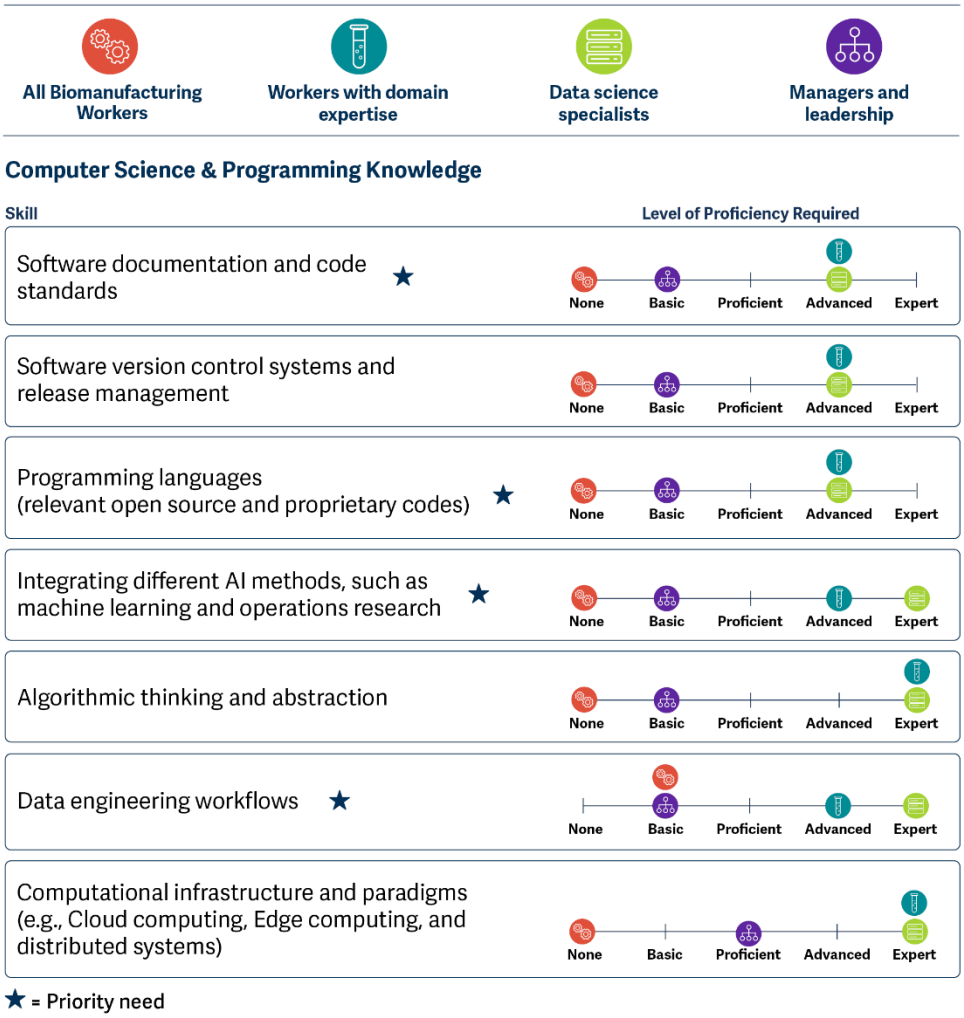
- Advanced understanding of programming languages, software documentation, and software version control systems
- Expertise in integrating different AI methods, algorithmic thinking, and computational infrastructure (e.g., Cloud computing)

#### *Managers and Leadership*

Managers and leadership do not need an ability to write programs or work deeply with languages, but should possess the following skills:

- A basic understanding of how to use software, how programs function, and software documentation standards
- Ability to discuss computational program development with domain experts

## Figure 3. Needed Computer Science & Programming Skills Mapped to Types of Workers Who Need Them



## Data Communications/Storytelling & Visualization

To successfully apply data science concepts and processes to biomanufacturing, teams of data specialists must know how to present and explain these processes to stakeholders as diverse as patients, clinicians, regulators, and business partners. This knowledge area represents critical skills for visualizing and communicating key takeaways and recommendations from available data to internal and external audiences with varying backgrounds and levels of understanding. Communicating effectively requires visualizations that are interactive, intuitive, and meaningful and created within a scientific or engineering context.

### Skills by Worker

#### *Average Biomanufacturing Industry Worker*

The average worker in the biomanufacturing industry should have the following skills in communicating and visualizing data:

- A basic understanding of how storytelling and technical writing can be used to explain data science concepts in biomanufacturing
- Basic skills in extracting key takeaways from data and communicating with data science experts
- Ability to develop meaningful and intuitive visualizations

#### *Workers with Domain Expertise*

In general, workers with expertise in a biomanufacturing domain area should possess a higher level of the same skills that the average biomanufacturing industry worker needs. They should also have the following skills:

- Advanced capability to explain data science concepts to more general audiences
- Ability to explore data and conduct trend analyses
- Ability to create more dynamic visualizations and dashboards

#### *Data Science Specialists*

Data science specialists should possess the following skills:

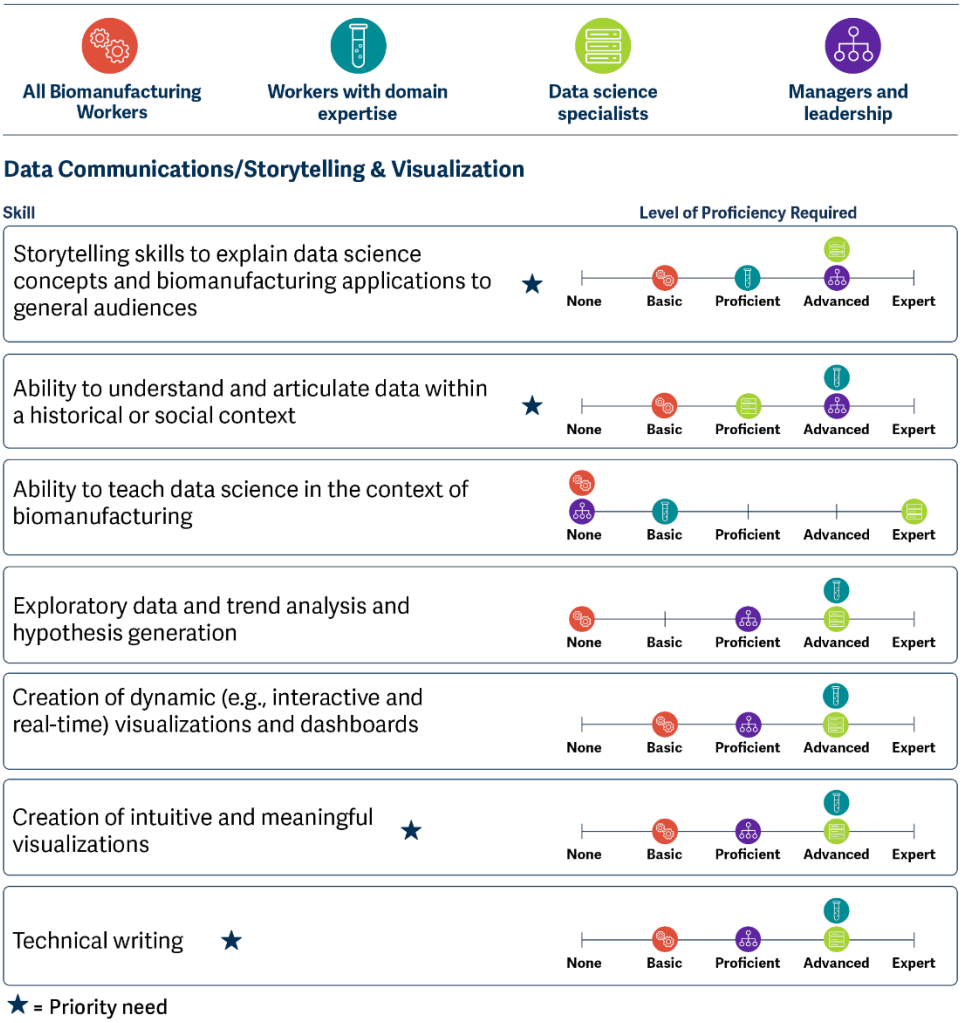
- Advanced ability to work with trend analyses and generate hypotheses based on data
- Advanced capability in communicating through dashboards, technical writing, and visualizations
- Expertise in teaching how data science concepts relate to biomanufacturing

#### *Managers and Leadership*

Managers and leadership should be comfortable with using the visualization methods developed by experts on their team to make decisions based on the information being presented. They should have:

- Proficiency in interpreting trends, developing and reading visualizations and dashboards, and technical writing
- An advanced ability to use data to explain data science concepts to general audiences to explain the context behind company decisions

## Figure 4. Needed Data Communications/Storytelling & Visualization Skills Mapped to Types of Workers Who Need Them



## Data Policy & Ethics

This knowledge area represents critical ethical concerns for biomanufacturing and relevant legal, regulatory, and social considerations. The methods by which research is conducted impact the development of data models and the decisions made based on those models—ensuring ethical conduct in research and development is crucial to ensure bias is not introduced that would affect any areas of the biomanufacturing process. Additionally, preserving privacy and confidentiality of patient data is essential throughout the manufacturing process, and this concept touches across other areas such as data management/security.

### Skills by Worker

#### *Average Biomanufacturing Industry Worker*

The average worker in the biomanufacturing industry does not need a high level of understanding of the implications and impacts of AI applications and data models since they are not usually involved in their design. In general, they should have a basic understanding of the following concepts:

- Biomedical ethics, privacy and confidentiality, and data sharing governance and intellectual property protection
- Data integrity, validation, and provenance to inform their data collection process

#### *Workers with Domain Expertise*

Workers with expertise in a biomedical ethics domain area would also not necessarily have proficiency in the impacts of poorly designed AI applications. Their expertise would instead cover the following areas:

- Biomedical ethics and privacy and confidentiality to support compliance with HIPAA and other applicable laws and to lead the interface with regulatory authorities
- Data sharing governance and intellectual property protection to facilitate legal agreements and help maintain a facility's business licenses
- Responsible conduct of clinical research to ensure AI applications and models do not contain biases

#### *Data Science Specialists*

Data science specialists are not usually involved in the conduct of clinical research; while they do not need expertise in responsible research conduct and biomedical ethics, they should have enough knowledge to flag issues in clinical data. They should also have the following skills:

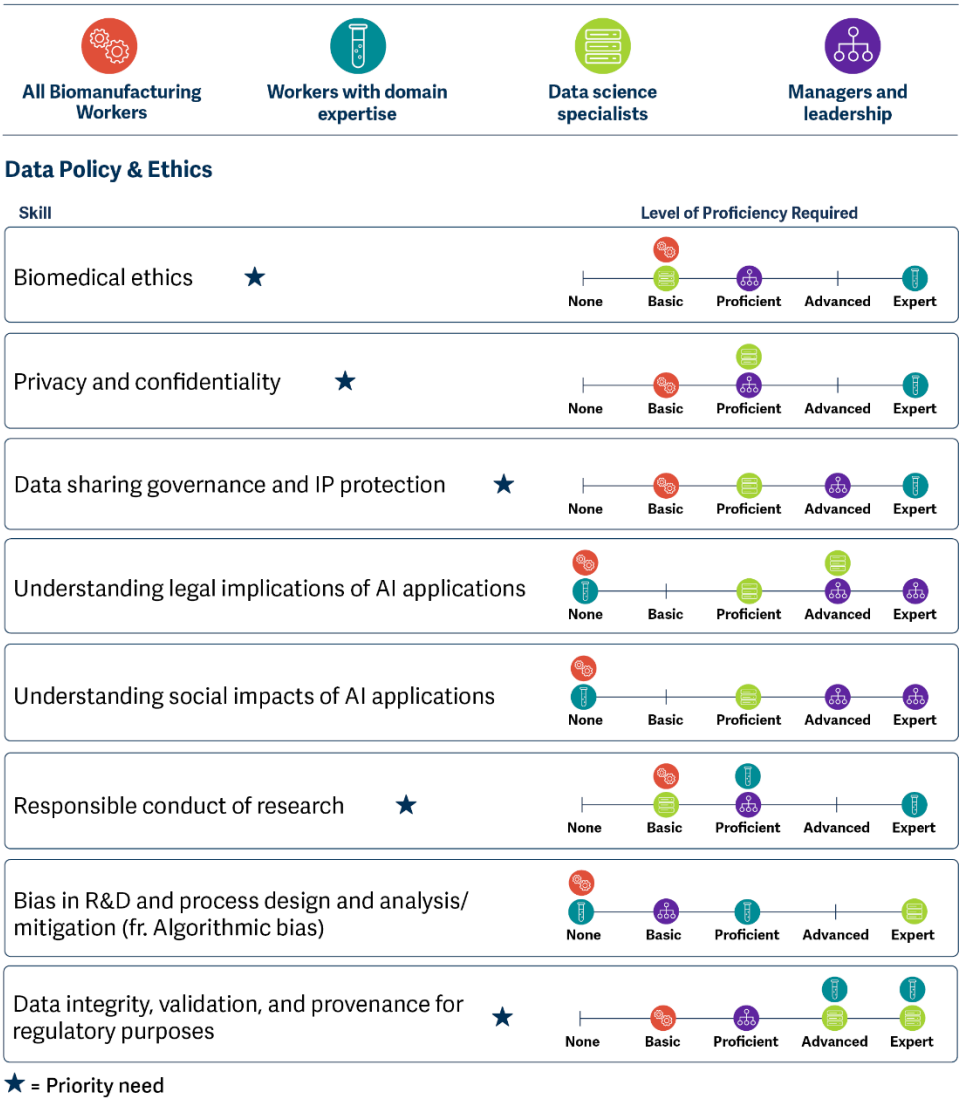
- Proficiency in understanding privacy and confidentiality requirements, as they are the ones implementing data infrastructure
- Advanced understanding of data integrity and validation for regulatory filing purposes, as well as the ability to interface with regulators during inspections if questions arise
- Expertise in data provenance both within a batch and across batches/chains

#### *Managers and Leadership*

Managers and leadership by and large should have proficiency or more advanced expertise in this skill area due to their position. They are responsible for ensuring that policies and laws are adhered to, lead companies/teams of people, and make decisions that impact R&D processes, which all require a deep understanding of the potential ethical and policy concerns that could arise in product development. More specifically, they should possess:

- Proficiency in biomedical ethics and the ability to teach/explain the concept to employees
- Proficiency in responsible conduct of research and the ability to review research prior to publication to ensure it was conducted responsibly
- Expertise in the legal implications and social impacts of AI applications that are poorly designed or contain biases to explain it to the developers and help with identifying issues

### Figure 5. Needed Data Policy & Ethics Skills Mapped to Types of Workers Who Need Them




## Database Engineering & Data Management/Security

Effectively leveraging the large amounts of data generated from biomanufacturing requires special considerations for data infrastructure and management, especially with respect to patient privacy concerns. This knowledge area represents critical data management skills and competencies with special relevance to biomanufacturing. Along with the data policy and ethics area, this area is foundational to ensuring successful and secure product development. This skill area also covers the entire product

development and data management lifecycle, for which security and management of the data being generated, handled, and stored is critical for regulatory approval.

## Skills by Worker

### *Average Biomanufacturing Industry Worker*

The average worker in the biomanufacturing industry will need proficiency in the underlying principles of effective data management, as data is at the core of their work. They will need at least a basic understanding in other areas such as:

- Chain of identity and chain of custody and data provenance
- Ensuring data integrity through processes such as cleaning and transforming data as well as addressing missing or conflicting data
- Lifecycle management and record retention policies
- Data security and privacy management principles

### *Workers with Domain Expertise*

Workers with expertise in a biomanufacturing domain area will need a higher overall level of knowledge on database engineering and data management practices than the average worker. They should possess:

- Proficiency in chain of identity and chain of custody and data provenance
- A deeper knowledge of how to ensure quality of data being generated and handled in manufacturing processes
- The ability to work with data science specialists to ensure data lifecycle management and record retention policies are being adhered to in their work
- Expertise in the principles underlying effective data management and documentation

### *Data Science Specialists*

Data science specialists are the experts in this skill area, with the most background and familiarity in developing and maintaining datasets and databases in biomanufacturing. These specialists will need:

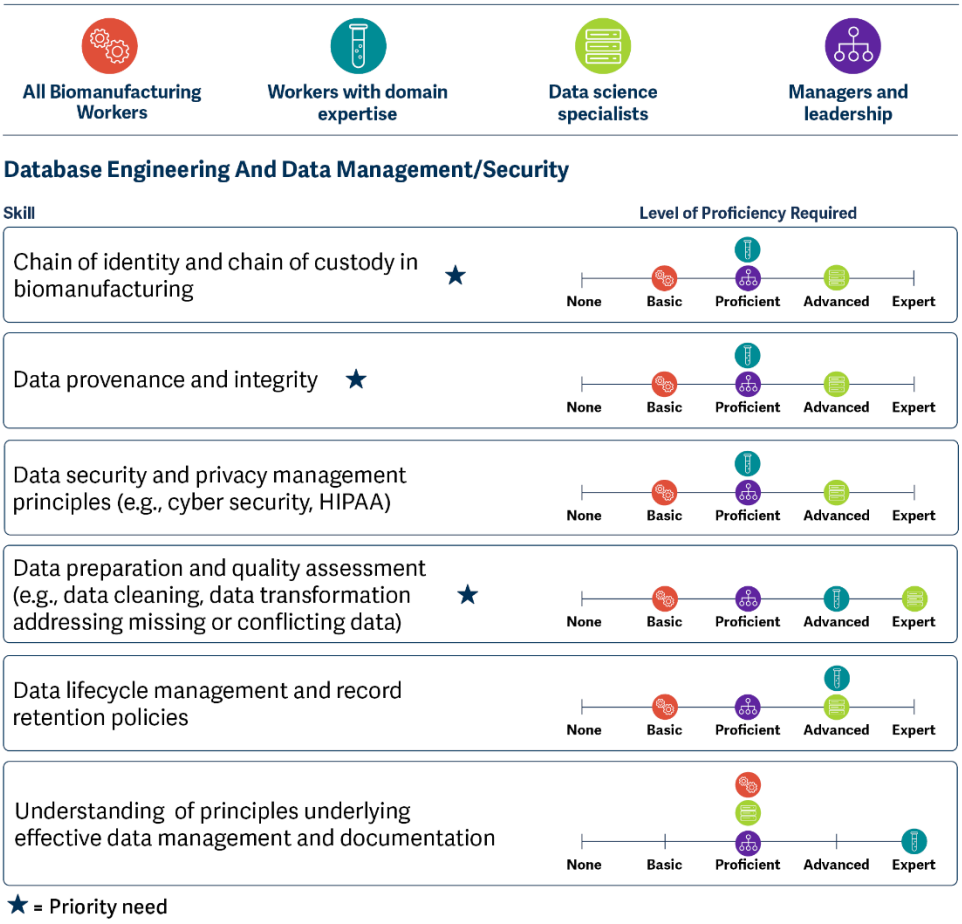
- Advanced understanding of how chain of identity, chain of custody, and data integrity are to be maintained and reported
- Expertise in maintaining data quality
- Expertise in handling and transforming vast amounts of data
- An understanding of the principles behind effective data management and documentation and the ability to work with the biomanufacturing domain experts

### *Managers and Leadership*

Managers and leadership should be proficient in all of the following areas of database engineering and data management to communicate with the experts and direct changes as needed to ensure compliance:

- The handling of potentially sensitive or private patient data
- Chain of identity and chain of custody
- Data cleaning and quality management
- The principles underlying these concepts

## Figure 6. Needed Database Engineering & Data Management/Security Skills Mapped to Types of Workers Who Need Them




## Data-Driven Decision-Making and Data Science Business Management

A major challenge to biomanufacturing is being able to improve process efficiency, which allows manufacturing processes to scale up to ultimately move more products to market. This knowledge area represents skills critical to applying data science to biomanufacturing in ways that increase efficiency and value in business settings. Because decisions are being made about how the business operates and how the manufacturing process is carried out, the biomanufacturing workforce should be able to understand how data is used to make decisions and what these decisions impact.

### Skills by Worker

#### *Average Biomanufacturing Industry Worker*

The average worker in the biomanufacturing industry will need to be able to understand the importance of data and its role in the decisions made by management and process designers. They will need to know:

- How to help assess data quality as it is generated, stored, and transformed in their work
- Basic applications for data science in biomanufacturing and why certain approaches are used

- How the data they are working with is used for real-time decision-making

### *Workers with Domain Expertise*

Workers with expertise in a biomanufacturing domain area will work closely with data science specialists and leadership to ensure that decision-making is based on valid data. They will need advanced experience in:

- Ensuring the availability of quality of data for decision-making
- Working with data science experts to appropriately incorporate data science applications into the biomanufacturing process and identify how to incorporate AI as appropriate
- Testing these biomanufacturing processes to ensure they are resilient

### *Data Science Specialists*

Data science specialists are crucial for monitoring and incorporating advanced data science and AI concepts into biomanufacturing to advance the field. They will need deep expertise in:

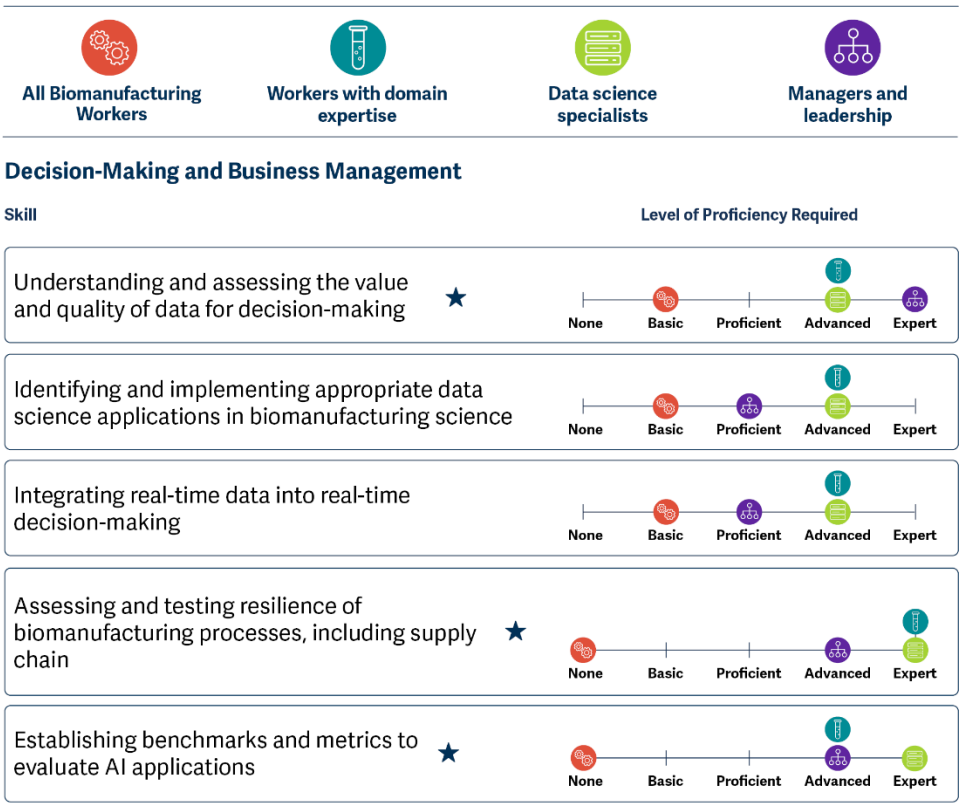
- Using data to assess the resilience of biomanufacturing processes and supply chains and report the findings to leadership
- Identifying and establishing benchmarks for ways to evaluate potential AI applications in the manufacturing process

### *Managers and Leadership*

Managers and leadership will need to be proficient in this area due to the nature of the decisions being made and how they need to interface with teams of people with differing expertise in these concepts. They should be able to:

- Recognize the quality of the datasets being used to inform decision-making and work with the data scientists and domain experts to make changes as needed
- Understand how data is being gathered and integrated into real-time processes that inform decision-making
- Make decisions on biomanufacturing process or supply chain changes based on resilience assessments
- Discuss the application of AI into existing processes and understand the business case for/against doing so

### Figure 7. Data-Driven Decision-Making and Data Science Business Management Skills Mapped to Types of Workers Who Need Them




## Machine Learning

Machine learning (ML) holds promise for tackling the highly complex nature of biomanufacturing processes and products, but the level of understanding of various ML concepts varies widely among the biomanufacturing workforce. Although ML is still an emerging field, the biomanufacturing industry can leverage it for advanced engineering and modeling purposes, which will enable further scale-up and technology transfer for manufacturing processes. This knowledge area represents both the fundamental concepts of ML and the considerations for its use in biomanufacturing. Due to ML's unique applications for biomanufacturing and relative newness, most workers will not need deep understanding of it.

### Skills by Worker

#### *Average Biomanufacturing Industry Worker*

The average worker in the biomanufacturing industry does not need to understand more advanced ML features, but should understand some of the underlying concepts related to ML, such as:

- How data privacy, security, and bias could impact ML models and the importance of ensuring data quality
- How to collaborate on feature engineering applications with domain experts as needed

### *Workers with Domain Expertise*

Workers with expertise in a biomanufacturing domain will need more proficiency with ML concepts and applications to be able to work with ML in their domain. They should be able to:

- Work with managers and leadership to develop biomanufacturing feature engineering applications with ML experts
- Know the difference between ML methods and how to interpret their results
- Identify and resolve issues with data bias and issues that impact privacy and security

### *Data Science Specialists*

Data science specialists serve as the ML experts for an organization and will be the ones designing and deploying ML models. They need to be completely comfortable with all aspects of ML and how to leverage it for biomanufacturing process improvement and modeling/forecasting. These skills include:

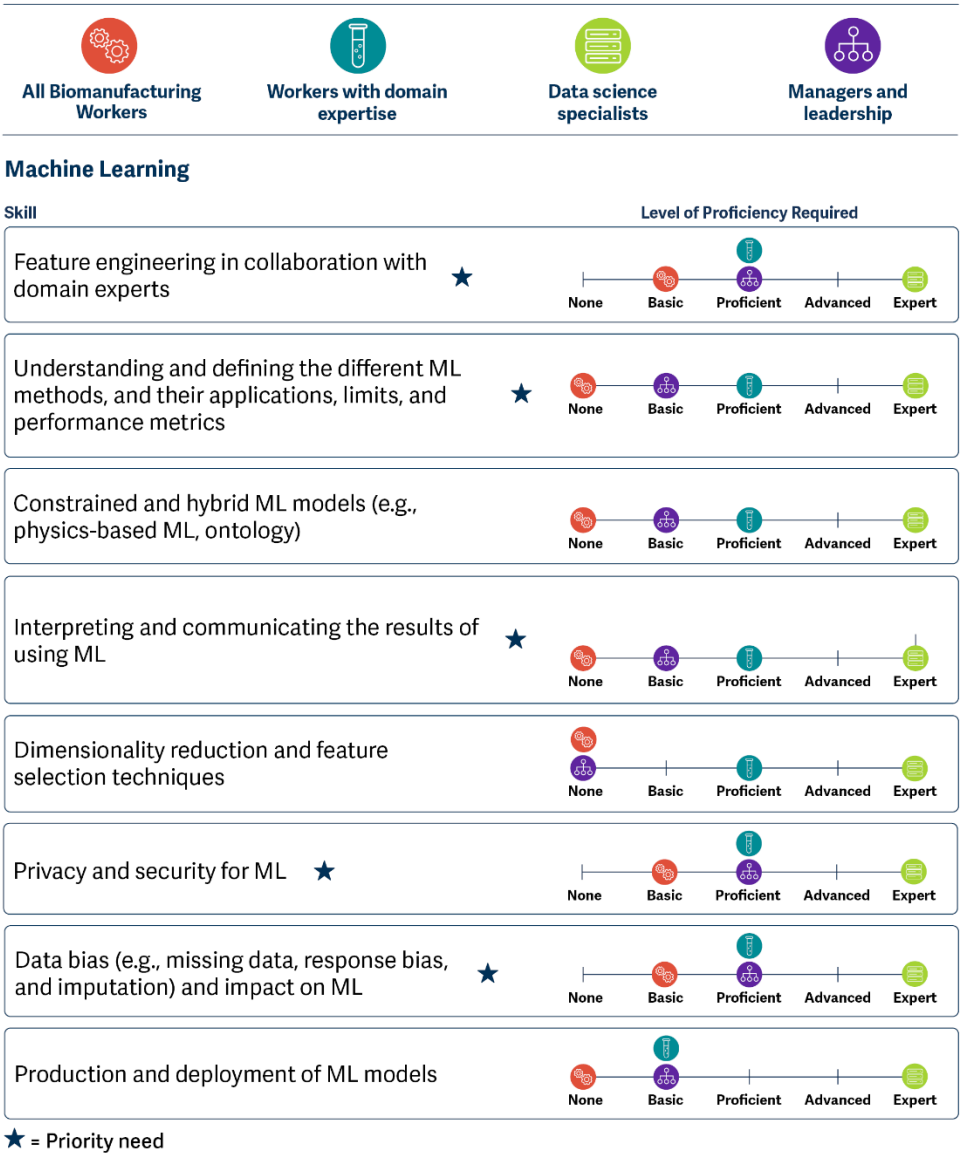
- Feature engineering applications for ML
- The full extent of the limits and applications of different ML methods
- Constrained and hybrid ML model applications
- How to produce, deploy, and maintain privacy and security in ML models
- Dimensionality reduction and feature selection techniques

### *Managers and Leadership*

Managers and leadership will need to possess some familiarity with ML to be able to work with the data scientists and domain experts and understand the business case for using ML to improve processes. Because they are the decision-makers, they need this understanding to be able to approve process changes and any other equipment or personnel adjustments to incorporate ML models. They should know:

- The purpose of ML and how to evaluate and apply it to a company's strategic goals
- When potential improvements are needed to existing models and how the changes would be deployed
- Ways to mitigate potential risks with implementing ML (e.g., policy changes)

## Figure 8. Machine Learning Skills Mapped to Types of Workers Who Need Them




## Mathematics & Statistics

Biomanufacturers must engineer efficient, commercially viable biomanufacturing processes that allow for product consistency, efficacy, and safety. This knowledge area represents the foundational mathematics and statistics skills necessary for developing and understanding quantitative assessments of biomanufacturing processes and products. Although most if not all workers would need to take a math or statistics course as part of their major to work in this industry, this area also covers more in-depth concepts that would require more specialized skills.

## Skills by Worker

### *Average Biomanufacturing Industry Worker*

In general, the average worker in the biomanufacturing industry does not need much knowledge or expertise in this area aside from what they might receive in a general education degree program. They should have a basic knowledge of:

- Selecting statistical methods
- Describing patterns and trends in the data they work with
- The concepts of causality and correlation

### *Workers with Domain Expertise*

In this area, these workers are experts in the mathematics and statistics domain. They should possess the highest level of skills and knowledge of all the areas needed for mathematics and statistics, including:

- Statistical inference
- Causality and correlation
- Multivariate statistical analyses
- Matrices, linear algebra, and differential equations
- Networks and graph theory
- Mathematical optimization

### *Data Science Specialists*

Data science specialists will work with the math and statistics experts to analyze the data that they are working with, and will need more proficiency in this area than others in the workforce. They should possess:

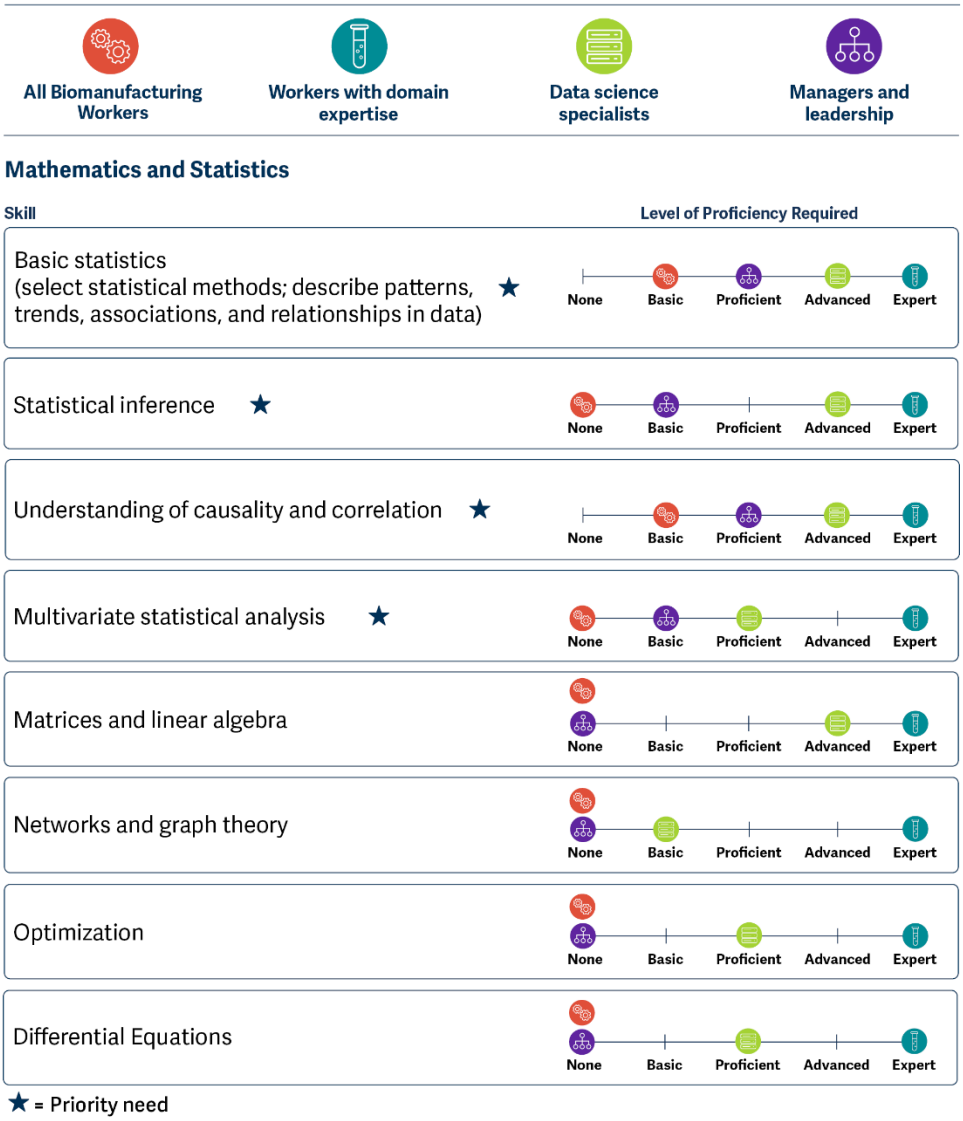
- Advanced ability in areas that most impact their work in data science, including finding associations/relationships between data, drawing conclusions from that data, and using linear algebra to build algorithms
- Proficiency in areas that inform modeling, including differential equations and multivariate analysis
- A basic understanding of networks and graph theory

### *Managers and Leadership*

Managers and leadership will need more familiarity with math and statistics concepts than an average worker to be able to communicate with and interpret information from the domain experts and data scientists, but will not need a high level of skill. They should be able to:

- Understand basic statistical principles to be able to see and understand the patterns and trends in datasets and the correlation/causality of different data points or datasets
- Make inferences from the data to inform their decision-making

## Figure 9. Mathematics and Statistics Skills Mapped to Types of Workers Who Need Them




## Modeling

Biomanufacturing processes and products are uniquely complex. This knowledge area represents critical modeling skills or applications needed to understand and develop complex systems in biomanufacturing. Advanced modeling has the potential to transform the biomanufacturing industry and enable more advanced scale-up processes and the ability to predict, mitigate, or react to risk, but some of the advanced modeling systems still lack consensus on how to define them and how exactly they would apply to biomanufacturing.

### Skills by Worker

#### *Average Biomanufacturing Industry Worker*

The average worker in the biomanufacturing industry should have a basic understanding of what modeling is used for and how to use models, with skills in:

- Assessing models for accuracy
- How systems change over time and based on different variables

### *Workers with Domain Expertise*

These workers will be experts in the mathematics and statistics domain rather than a biomanufacturing domain. They should have the highest level of expertise in all areas related to modeling, including:

- A deep understanding of the different types of models, how to use them, how to interpret their results, and how to assess them for accuracy
- Expertise in developing and interpreting results from advanced models like digital twins, hybrid physical-statistical models, multivariate modeling, and dynamical systems

### *Data Science Specialists*

Data science specialists will need the same level of expertise as the domain experts, including:

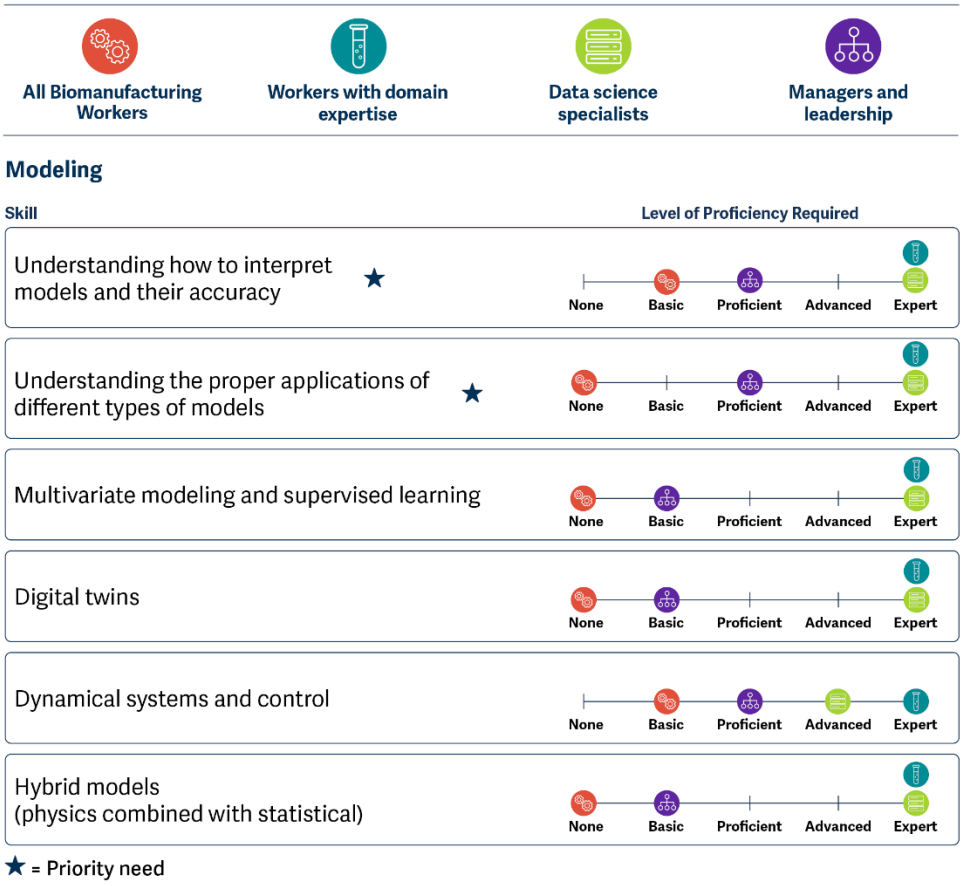
- The ability to use their data in advanced models and work with the domain experts to interpret the models
- Understanding which models would be most appropriate to use with different datasets and how to combine datasets into hybrid models

### *Managers and Leadership*

Managers and leadership should be able to discuss impediments and timelines with workers who build models and communicate any decisions that they make based on these models. Doing so requires the following skills:

- A basic ability to interpret the results of multivariate modeling, digital twins, and hybrid models
- The ability to determine the accuracy of the model and work with the experts to make changes if needed
- An understanding of dynamical systems and control as applied to biological processes and how those impact the product development processes

## Figure 10. Modeling Skills Mapped to Types of Workers Who Need Them



## Existing Educational Programs

This section captures the preliminary inputs of participants regarding existing programs and resources that address the data science training needs described above either generally or in contexts specific to biomanufacturing across all educational levels and relevant competency areas.

The goal of this effort will be to assess educational programs and resources to determine which data science training needs are adequately addressed by existing programs, which existing programs can be adapted or expanded to address unmet needs, and which needs require development or implementation of new programs or resources.

## Figure 10. Existing Educational Programs that Address Needed Skills

| | Biomanufacturing Domain Knowledge & Systems Engineering (crosscutting) | Computer Science & Programming | Data Communications/ Storytelling & Visualization | Data Policy & Ethics Needs | Data-Driven Decision-Making and Data Science Business Management | Database Engineering & Data Management/Security | Machine Learning | Mathematics & Statistics | Modeling |
|---|---|---|---|---|---|---|---|---|---|
| **K-12** | | | | | | | | | |
| AI4K12 | | ● | | | | | | | |
| AP courses in high school | | | | | | | | ● | |
| ARMI programs for workforce development | ● | | | | | | | | |
| CTAE/Vocation Biotech pathways | ● | | | | | | | | |
| UCLA Rob Gould high school program for staistics and data science | | | ● | | | | | ● | ● |
| **Technical Colleges** | | | | | | | | | |
| Athens Technical College | | | | | | | | | |
| DSC-WAV* | | | ● | | | | | | |
| **Undergraduate** | | | | | | | | | |
| Boston University/MIT Capstone programs* | | | | | | | | | |
| DSC-WAV* | | | ● | | | | | | |
| GA Tech and Emory University joint program on bioengineering | ● | | | | | | | | |
| Project TIER | | | | ● | | | | ● | |
| UCLA Math/Stat Data Theory major* | | | | | | | | ● | ● |
| GA Tech Computational Science and Engineering (interdisciplinary program requirements) | | ● | | | | | | | ● |

*Experiential learning

| | Biomanufacturing Domain Knowledge & Systems Engineering (crosscutting) | Computer Science & Programming | Data Communications/ Storytelling & Visualization | Data Policy & Ethics Needs | Data-Driven Decision-Making and Data Science Business Management | Database Engineering & Data Management/Security | Machine Learning | Mathematics & Statistics | Modeling |
|---|---|---|---|---|---|---|---|---|---|
| **Undergraduate (cont.)** | | | | | | | | | |
| U of Washington BS program in Applied & Computational Mathematical Sciences with a bio/life science option | ● | | | | | | | ● | |
| **Graduate** | | | | | | | | | |
| GT ISYE MS program in analytics (online and regular) | | | | | | | | ● | ● |
| GA Tech and Emory University joint program on bioengineering | ● | | | | | | | | |
| UGA MS/PhD in Bioinformatics | ● | | | | ● | | | | ● |
| GA Tech online MS in CS and ML courses | | ● | | | | | ● | | |
| GA Tech Computational Science and Engineering (interdisciplinary program requirements) | | ● | | | | | | | ● |
| GA Tech interdisciplinary ML PhD program | | | | | | | ● | | |
| UVA School of Data Science Collaboratory for Applied Data Science in Business (DCADS) | | | | ● | ● | | | | |
| UCLA programs in Bioinformatics, Computational Medicine, Institute for Personalized Medicine, Masters in Business Analytics | | | | | | | | | |

*Experiential learning

| | Biomanufacturing Domain Knowledge & Systems Engineering (crosscutting) | Computer Science & Programming | Data Communications/ Storytelling & Visualization | Data Policy & Ethics Needs | Data-Driven Decision-Making and Data Science Business Management | Database Engineering & Data Management/Security | Machine Learning | Mathematics & Statistics | Modeling |
|---|---|---|---|---|---|---|---|---|---|
| **Continuing Education** | | | | | | | | | |
| tinyML (edX) | | | | | | | ● | | |
| Coursera | | ● | | | | | | | |
| Continuing education/ training programs (very short duration) for current workforce | | | | | | | | | |
| fast.ai (with Jeremy Howard) | | ● | | ● | | | ● | ● | ● |
| Insight, Data Science programs | | ● | ● | | ● | ● | ● | ● | |
| The Data Incubator | | ● | | | ● | ● | | ● | |
| UVA School of Data Science Collaboratory for Applied Data Science in Business (DCADS) | | | | ● | ● | | | | |
| Google Data Analytics certificate | | ● | ● | | | ● | | | |
| Project Management (PMI) | | | | | ● | | | | |
| Columbia university DS certification program | | | ● | | | | ● | ● | |
| Google Colaboratory | | ● | | | | | | | |

# Appendix A. Participant List

Participants with * next to their name participated in the focus group sessions only.

Participants with ** next to their name participated in both the workshop and the focus group sessions.

**Berry Bates** Atlanta Technical College

bbates@atlantatech.edu

***Fani Boukouvala**
Georgia Tech
Fani.boukouvala@chbe.gatech.edu

**Carl Burke**
Century Therapeutics
carl@centurytx.com

**Andy Campbell**
ThermoFisher Scientific
Andrew.Campbell@thermofisher.com

**Jon Chesnut**
ThermoFisher Scientific
Jon.Chesnut@thermofisher.com

**Pedro Cruz**
Resilience
pedro.cruz@resilience.com

**Thomas Ewing**
Virginia Tech
etewing@vt.edu

**Marc Ferrer**
National Institutes of Health (NIH)
National Center for Advancing Translational Sciences (NCATS)
marc.ferrer@nih.gov

**Stefan France**
Georgia Institute of Technology/
BioMADE
stefan.france@chemistry.gatech.edu

****Greg Gibson**
Georgia Institute of Technology
greg.gibson@biology.gatech.edu

**Sheng Lin-Gibson**
National Institute of Standards and Technology
sheng.lin-gibson@nist.gov

**Mark Green**
University of California, Los Angeles
mlgucla@gmail.com

**Stan Harrison**
Morgan County High School, GA
stan.harrison@morgan.k12.ga.us

****Roger Hart**
National Institute for Innovation in Manufacturing Biopharmaceuticals (NIIMBL)
rogerhart@alumni.caltech.edu

**Nick Horton**
Amherst College
nhorton@amherst.edu

**Melissa Kemp**
Georgia Institute of Technology
melissa.kemp@bme.gatech.edu

**Arthur Kordon**
Kordon Consulting
Arthur@kordon-consulting.com

**Theresa Kotanchek**
Evolved Analytics
theresa@evolved-analytics.com

****Rachel Lanspa**
Nexight Group
rlanspa@nexightgroup.com

**Rachel Levy**
North Carolina State University
rachel.levy@gmail.com

****Sarah Lichtner**
Nexight Group
slichtner@nexightgroup.com

***Anna Liu**
ThermoFisher Scientific
anna.liu@thermofisher.com

**Jonathan Lowe**
Pfizer
Jonathan.Lowe@pfizer.com

***Catherine Mysliwiec**
Nexight Group
cmysliwiec@nexightgroup.com

**Tom Nelson**
Advanced Regenerative Manufacturing Institute (ARMI) | BioFabUSA
tnelson@armiusa.org

**Rebecca Nugent**
Carnegie Mellon University
rnugent@stat.cmu.edu

****Tony Pan**
Georgia Institute of Technology
tony.pan@gatech.edu

**Pamela Peralta-Yahya**
Georgia Institute of Technology/
BioMADE
pperalta-yahya@chemistry.gatech.edu

**Peng Qiu**
Georgia Institute of Technology
peng.qiu@bme.gatech.edu

****Aditya Radhakrishnan**
Shape Therapeutics
aditya@shapetx.com

**Jeff Rapp**
Athens Technical College
jrapp@athenstech.edu

**Renata Rawling-Goss**
Georgia Institute of Technology
rrawlings.goss@gmail.com

**Greg Russotti**
Century Therapeutics
grussotti@centurytx.com

**Rachel Saidi**
Montgomery College
rachel.saidi@montgomerycollege.edu

***Ryan Shorr**
Century Therapeutics, Inc
rshorr@centurytx.com

***Saurabh Sinha**
Georgia Tech
Saurabh.sinha@bme.gatech.edu

**Josh Speer**
ThermoFisher Scientific
josh.speer@thermofisher.com

***Shalu Suri**
Georgia Tech
shalu.suri@bme.gatech.edu

****Johnna Temenoff**
Georgia Tech
Johnna.temenoff@bme.gatech.edu

**Tom Tubon**
BioMADE
tubon@madisoncollege.edu

***James Warnock**
University of Georgia
James.Warnock@uga.edu