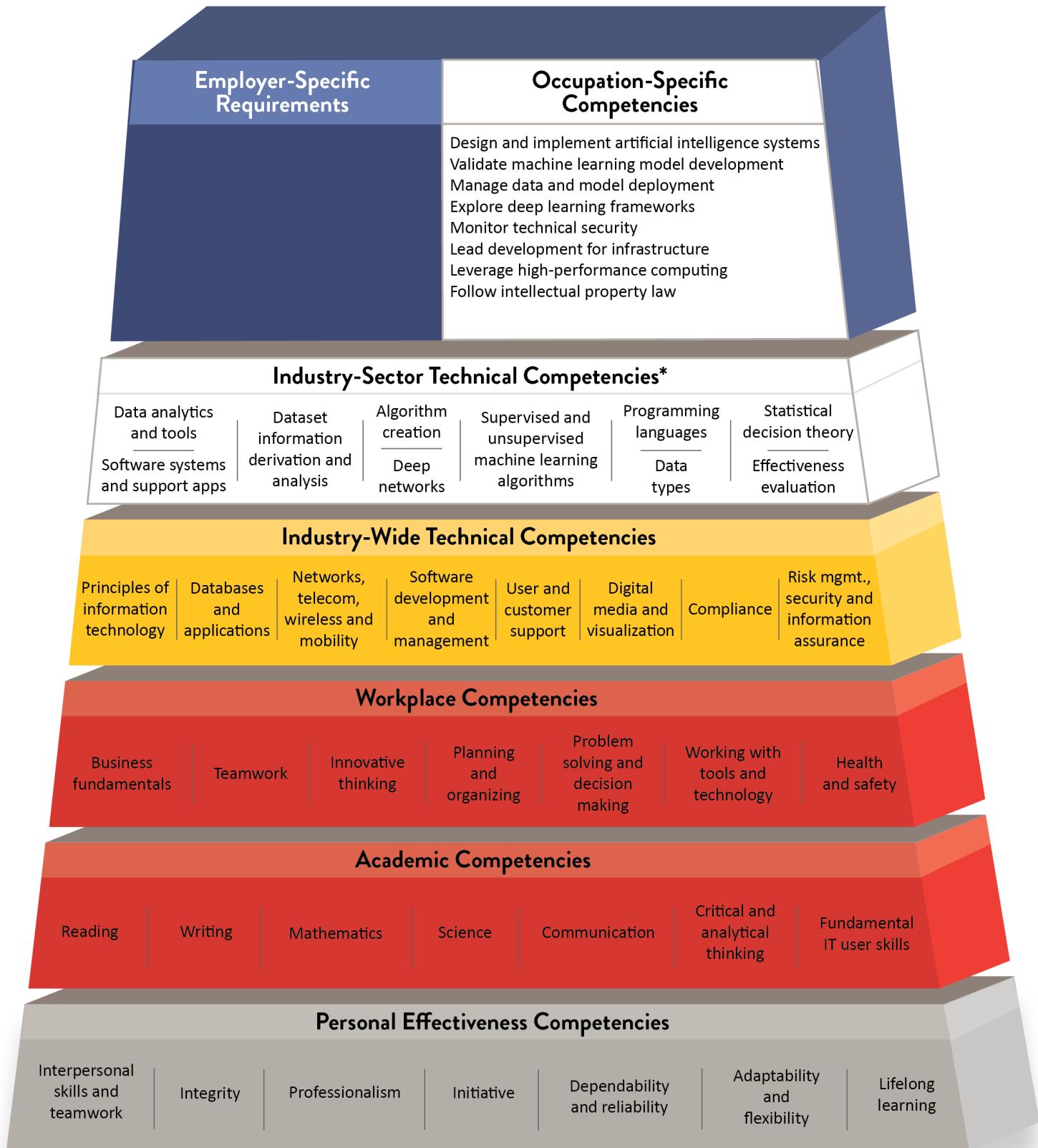


Minnesota Dual-Training Pipeline

Competency Model for Information Technology

Occupation: Data Science/Artificial Intelligence Machine Learning Specialist



Based on: Information Technology Competency Model, Employment and Training Administration, United States Department of Labor, February 2025. For more detailed information about competency model creation and sources, visit dli.mn.gov/business/workforce/information-technology.



Competency Model for Data Science/Artificial Intelligence Machine Learning Specialist

Data Science/Artificial Intelligence Machine Learning Specialist – A professional that sources, cleans and processes data to extract meaning for analytical purposes. Establishes and achieves objectives using techniques associate with artificial intelligence reasoning and uncertainty. Applies logic, probability analysis, and machine-learning concepts to problem-solving initiatives.

*Pipeline recommends the Industry-Sector Technical Competencies as formal training opportunities (provided through related instruction) and the Occupation-Specific Competencies as on-the-job (OJT) training opportunities.

Industry-Sector Technical Competencies

Related Instruction for dual training means the organized and systematic form of education resulting in the enhancement of skills and competencies related to the dual trainee's current or intended occupation.

- **Data analytics and tools** – Understand the science of examining raw data with the purpose of discovering knowledge. Use of tools such as Power BI, Tableau, etc.
- **Software systems and support apps** – Ability to design, develop, and modify software systems. Knowledge of system support apps such as Jupyter notebooks, etc.
- **Dataset information derivation and analysis** – Understand the relevant methods for performing data collection, representation, transformation, and data-driven decision making.
- **Algorithm creation** – Be able to design a set of instructions to perform a specific task.
- **Deep networks** – Understand the artificial neural network (ANN) with multiple layers between the input and output layer consisting of neurons, synapses, weights, biases, and functions.
- **Supervised and unsupervised machine learning algorithms** – Be able to discuss both the theoretical underpinnings of machine learning techniques and experience in implementing them.
- **Programming languages** – Understand high-level general-purpose programming languages such as Python, Sequel, R, etc. These aim to help programmers write clear, logical code for small and large-scale projects.

- **Data types** – Understand kinds of data items, as defined by the values it can take, the programming languages used, or the operations that can be performed on it.
- **Statistical decision theory** – Understand how this theory brings together psychology, statistics, philosophy, and mathematics to analyze the decision-making process.
- **Effectiveness evaluation** – Know how to apply machine learning to a given problem.

Occupation-Specific Competencies

On-the-Job Training is hands-on instruction completed at work to learn the core competencies necessary to succeed in an occupation. Common types of OJT include job shadowing, mentorship, cohort-based training, assignment-based project evaluation and discussion-based training.

- **Design and implement artificial intelligence systems** – Understand how to identify the problem, prepare data, choose algorithms, train the algorithms, choose a particular programming language and run-on selected platform.
- **Validate machine learning model development** – Be able to define the objective, collect, and prepare data, choose the model, train the machine model, evaluate, prediction or inference.
- **Manage data and model deployment** – Know how to make the model available for other collaborators to quality test. Once validated, deployment is the process of configuring an analytic asset for integration with other applications or access by business users to serve the production workload at scale.
- **Explore deep learning frameworks** – Understand PyTorch or TensorFlow which offer building blocks for designing, training and validated deep neural networks.
- **Monitor technical security** – Understand security for storage and network service delivery.
- **Lead development for infrastructure** – Know how to code for multiple clouds.
- **Leverage high-performance computing** – Understand the practice of aggregating, computing power in a way that delivers much higher performance than one could get out of a typical desktop computer or workstation to solve large problems in science, engineering, or business.
- **Follow intellectual property law** – Know the multiple areas of law that govern the ownership and rights to digital “products of the mind”.

Updated December 2025