(DRAFT) Bridges to Data Science: The K-12 Math of AI

This resource outlines modern data science and artificial intelligence (AI) concepts and maps them to their corresponding roots in K-12 Common Core Math Standards [1] and the international PISA 2021 Mathematics Framework [2].

Our aim is to build bridges between classroom content and industry knowledge, with the goal of creating K-12 learning environments grounded in culturally relevant real-world applications.

Use Cases

We designed *Bridges to Data Science* to be an actionable resource for audiences that may include:

- Teachers planning a lesson on the real-world applications of math in data science / AI
- Instructional designers building data science / AI learning activities
- Data scientists or industry mentors sharing knowledge accessible at the K-12 level

About Us

The National AI Circle is non-profit organization that empowers K-12 students with the artificial intelligence education they need to succeed in tomorrow's economy. We are a community of students, educators, data scientists, and academic researchers in AI. Our mission is to provide every student with the pathways and motivation to shape the growing AI industry in diverse, creative, and ethical ways.

To accomplish our mission, the National AI Circle builds communities, content, and programs bridging K-12 education and the data-driven workplace. Our aim is to provide all students – especially those in underserved communities – the opportunities to connect what they learn in school to real-world, culturally relevant applications in artificial intelligence.

Guide to Contents

This document organizes concepts by mathematical topic area and their corresponding applications in data science and AI. Many of these concepts are shared in broader fields of applied mathematics such as statistics, computer science, information theory, and even broader techniques used in areas such as psychology research. We encourage these cross-disciplinary examples as they present holistic and historical perspectives on artificial intelligence.

For readers who are new to the field of AI, we recommend reading through the introduction followed by each section organized by mathematical topic area. These will provide an introduction to big ideas and concepts in AI intended for a general audience. Each section contains:

- Real-world, culturally relevant applications of AI built on K-12 math
- A practical teaching guide for use by teachers and industry mentors
- Links to supplementary resources and lessons
- Standards alignment tables mapping AI concepts to K-12 mathematical standards

For readers who are more familiar with AI and are hoping to build their own learning content, you may choose to refer directly to the detailed content alignment tables in each section.

What is Artificial Intelligence?

Artificial intelligence (AI) is a technology that is shaping our society, economy, and everyday lives. Research conducted by the World Economic Forum estimates that by 2030, AI will fundamentally change 30% of future jobs – automating old jobs, and creating perhaps as many new jobs as it replaces through overall economic growth [3].

These rapid economic advancements have also come with a spread in hype and misinformation about what AI actually is. The Brookings Institution noted in 2018 that “few concepts are as poorly understood as artificial intelligence …. the lack of clarity around the term enables technology pessimists to warn AI will conquer humans, suppress individual freedom, and destroy digital privacy” [4].

Against this backdrop, our nation’s students are paying attention as they prepare for their own careers. A 2019 study estimates that 47% of young adults believe they will work in jobs that do not currently exist, yet only 18% believe they are prepared with the 21st century skills they need. Most importantly, 74% feel that they are not getting enough information on careers that will be available [5].

We see a clear call to action for a democratic K-12 education in artificial intelligence that is grounded in science, ethics, and cultural relevance. This guide gathers examples from our own research and backgrounds in education, industry, and academia to illustrate what modern AI is: **a collection of computer models that use mathematical techniques to automate “intelligent” tasks such as prediction or classification**. Modern AI (or “machine learning”) is closely tied to statistics and data mining and can be considered as a branch of data science.

By drawing connections from AI to its roots in K-12 mathematics along with examples from everyday life, we hope to demystify and empower AI learning for all students and their educators.

Section I: Vector Math in AI

AI models "understand the real-world" by using **feature vectors.** Feature vectors are numeric representations of information that capture multiple aspects (or dimensions) of collected data.

For example, suppose we're building an AI model to identify similar basketball players. We may choose to provide the model information such as player experience, minutes played, or the number of points a player scores per game on average. In this scenario, we could represent every basketball player by a three-dimensional feature vector: [years of experience, minutes played, points per game].

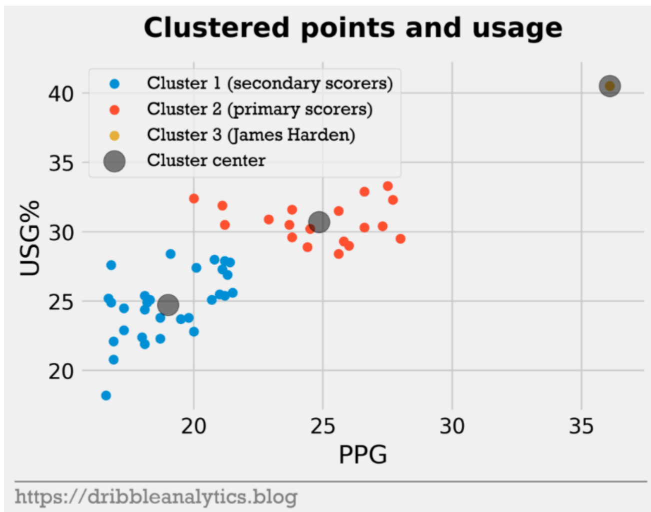

Fig 1. A visualization of two-dimensional feature vectors for NBA players using features such as usage and points per game. Source: DribbleAnalytics [6]

Additional examples of feature vectors include:

- Fashion products, where each dimension is a product attribute such as weight, cost, or material
- Images, where each dimension is the color value of an individual pixel

Once data is collected in the form of feature vectors, AI models use mathematical techniques to extract **knowledge** from the data - predictions, trends, clusters, correlations, etc. These cover a broad range of applications, including:

- Recommending online content, by comparing posts on TikTok, Instagram, YouTube, Netflix represented by feature vectors with information such as views, likes, or tags.

- AI-assisted medical diagnosis, by classifying images of tissue samples to provide signals as input to doctors [7]. Feature vectors in this setting may be crafted with help from medical domain experts and computer scientists [8], or even just provided as raw pixel values.

There are meaningful connections from K-12 mathematics to the techniques that modern AI systems apply to learn from feature vectors. One notable example is the **Pythagorean Theorem**, which is commonly applied to estimate the **similarity** of two data points by measuring the distance between their feature vectors. The smaller the distance, the greater the similarity.

Teaching Guide for Vector Math in AI

Framework	Main Alignments to Vector Math in AI
Common Core Mathematics Standards (CCSS)	[CC8.G.B.8] Apply the Pythagorean Theorem to find the distance between two points in a coordinate system [CCHSN.VM.A] Represent and model with vector quantities [CCHSN.VM.B] Perform operations on vectors
PISA2021 Mathematics Framework – Processes	[36] Formulating situations mathematically [57] Appreciating the power of abstraction and symbolic representation [69 – 70] Using mathematical modelling as a lens onto the real world
PISA2021 Mathematics Framework – Content Area	[98 – 101] Quantity

Both the Common Core Standards and PISA2021 contain a strong call to action to teach K-12 students how to represent and model real-world quantities. The Common Core introduction to number and quantity provides examples such as "social science rates such as per-capita income, and rates in everyday life such as points scored per game or batting averages" as quantities that high school students are expected to encounter in modelling [1].

PISA2021 further expands this call to action by explicitly highlighting the role that computer models play in addressing real-world problems. The following quote is worth noting:

"Computer Simulations: Both in mathematics and statistics there are problems that are not so easily addressed because the required mathematics are complex or involve a large number of factors all operating in the same system or because of ethical issues relating to the impact on living beings or their environment. Increasingly in today's world such problems are being approached using computer simulations driven by algorithms." (PISA2021 Mathematics Framework, Paragraph 100) [2].

Recommended Minimum Grade: 8

Although vectors and matrices are generally introduced in high school courses (and expanded upon in college via linear algebra), there are proven examples of middle-school aligned curricula that teach feature vectors in AI in ways that are relevant to students' lives:

- MIT Media Lab's AI Ethics Curriculum introduces the concept of features and datasets through a game - AI Bingo – that relates AI to common interactions middle school students may have with technology platforms, such as suggested emojis in a mobile chat or an application that is capable of recognizing songs playing from audio [9]

- The National AI Circle's HeadFirst AI Workshop introduces feature vectors through a middle school activity where students explore how to build their own YouTube recommender system. By quantifying the similarity in attributes such as Title, Tags, or Views, students learn to customize the behavior of video recommendations and implications for users [10]

Both examples above introduce features and datasets in a real-world context. This is especially important for grade 8, where most students are not yet familiar with vectors and are experienced only with analyses that are two-dimensional at most. Bivariate data is explored further at the high school level. Industry mentors will also likely find it useful to introduce practical ways to encode categorical data such as text (see [10] for examples).

Detailed Alignment Table by Concept for Vector Math in AI

AI Concept(s)	Common Core Standard(s)	PISA2021 Component(s)
Clustering (k-means) / Unsupervised Learning / Cohort Analysis	[CC6.SP.B] Summarize and describe distributions [CC8.G.B.8] Apply the Pythagorean Theorem to find the distance between two points in a coordinate system	[63] … Statistical analysis is often a matter of imposing a structure on a set of data [96] Geometric approximations
Dimensionality Reduction (PCA) / Contextual Embeddings / Vector Space Models	[CCHSS.ID.B.5] Summarize categorical data for two categories in two-way frequency tables [CCHSN.VM.C.11] Multiply a vector (regarded as a matrix with one column) by a matrix of suitable dimensions to produce another vector. Work with matrices as transformations of vectors	[73] … modelling variation as measured by the variance or in the case of multiple variables the covariance matrix
Semantic Similarity	[CCHSN.VM.A.1] Recognize vector quantities as having both magnitude and direction [CCHSF.TF.A] Extend the domain of trigonometric functions using the unit circle	[57] Students use representations – whether text-based, symbolic, graphical, numerical, geometric or in programming code – to organize and communicate their mathematical thinking [112, 115 – 118] Representations [are] often dependent on the context in which a problem arises … personal, occupational, societal, scientific
Convolutional Neural Networks / Matrix Filters	[CCHSN.VM.C.11] Multiply a vector (regarded as a matrix with one column) by a matrix of suitable dimensions to produce another vector. Work with matrices as transformations of vectors [CCHSN.VM.C.12] Work with 2x2 matrices as transformations of the plane	

Section II: Probability and Statistics in AI

AI models must contend with the inherent **uncertainty** behind tasks such as prediction or decision making in the real world. From technical design to success metrics, AI models in industry are communicated using the language of probability and statistics.

For example, suppose we've built an AI model to assist doctors in diagnosing COVID-19, given patient features such as current symptoms and medical history. Our model may learn to recognize probabilistic associations between these factors and COVID-19, and use these to produce a "positive" or "negative" classification for each patient.

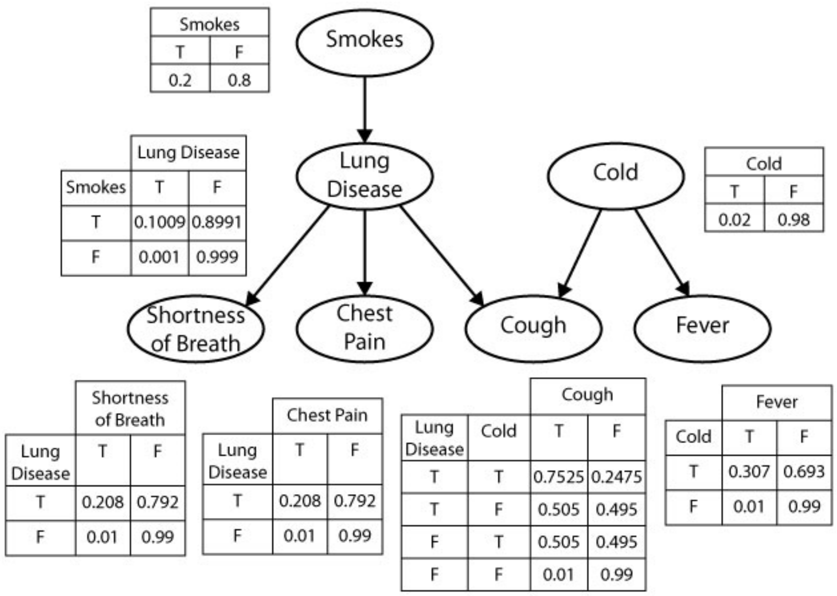

Fig 2. A depiction of a probabilistic graphical model (PGM) as an example of AI associating patient factors in medical diagnosis. Source: ProbMods [11]

However, the results are likely to face uncertainty from hidden factors that are difficult to measure – for example, how frequently did the patient touch his or her face? Other factors may be hidden rightfully due to privacy reasons. Any of these factors may influence the actual prognosis of the patient.

To account for this uncertainty, data scientists hold AI systems accountable to statistical success metrics such as **precision** or **recall**. Precision is a ratio that measures the likelihood that any "positive" classification produced by the AI model is correct. Recall is a ratio that measures the percentage of "positive" prognoses identified by the AI model. Such statistical metrics provide a picture of how AI models are likely to perform once they are applied in society, and ultimately drive researchers to develop reliable and unbiased AI.

Both probabilistic AI models and statistical metrics are applications of bivariate categorical tables that students learn starting in grade 8. The foundation of statistical AI is the more general concept of **conditional probability** introduced in high school statistics, which has been applied broadly in systems ranging from simple email spam classifiers to the deep learning models that power online content platforms such as YouTube, Facebook, Amazon, and Netflix.

Teaching Guide for Probability and Statistics in AI

Framework	Main Alignments to Probability and Statistics in AI
Common Core Mathematics Standards (CCSS)	[CC6.SP.A] Develop understanding of statistical variability [CC8.SP.A] Investigate patterns of association in bivariate data [CCHSS.CP.A] Understand independence and conditional probability and use them to interpret data
PISA2021 Mathematics Framework – Processes	[69 – 70] Using mathematical modelling as a lens onto the real world [72 - 74] Understanding variation as the heart of statistics
PISA2021 Mathematics Framework – Content Area	[102 – 106] Uncertainty and data

PISA2021 highlights the connection between uncertainty and data as one of four major content areas of mathematical knowledge that every fifteen-year-old should learn. Particularly relevant to AI and modern data science methods is the encouragement for students to participate in "conditional decision making… a measure of the variation characteristic of much of what people encounter in their daily lives … The expectation for the PISA test items is that students will be able to read the relevant data from the table with a deep understanding for the meaning of the data they are extracting" [2].

Emphasizing real-world connections to statistics and data will allow students to build an intuition for the conditional probability that AI models perform on larger-scale datasets with thousands of factors. It will also demonstrate many of the pitfalls in building and interpreting the results of AI models, with **Simpson's paradox** and **causation versus correlation** being two prominent examples.

Recommended Minimum Grade: 6

Students are introduced to variability and uncertainty in Common Core standards as early as grade 6. Asking students to collect and evaluate data in culturally relevant, real-world contexts may provide a more intuitive introduction to probability than a purely symbolic introduction through the abstractions of probability. For example, an exercise asking students to survey the percentage of classmates who prefer ice cream to cookies (and why?) may spark discussion on causation, uncertainty, and sampling.

One example that successfully introduces data science and statistical models in middle school is YouCubed's Data Talks and Data Science Lessons [12]: a series of resources ranging from 5 to 10 minute classroom conversations on data visualizations to a lesson unit that introduces data analysis and modeling in the context of students' home communities. AI models perform an analogous role within our society, with the only primary difference being computational ability to automate data processing. Both human-driven data models and AI-driven data models and are subject to the same challenges and biases concerning variability.

Detailed Alignment Table by Concept for Probability and Statistics in AI

AI Concept(s)	Common Core Standard(s)	PISA2021 Component(s)
Generative Models / Naïve Bayes / Probabilistic Graphical Models Causation vs. Correlation Simpson's Paradox	[CC7.SP.C.8] Find probabilities of compound events using organized lists, tables, tree diagrams, and simulation. [CC8.SP.A] Investigate patterns of association in bivariate data [CCHSS.CP.A] Understand independence and conditional probability and use them to interpret data [CCHSS.ID.B.5] Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies) [CCHSS.ID.C.9] Distinguish between correlation and causation	[104] Conditional decision making … When there is more than one variable, there is variation in each of the variables as well as co-variation characterising the relationships among the variables [103] Economic predictions, poll results, and weather forecasts all include measures of variation and uncertainty … knowledge of number and of aspects of algebra such as graphs and symbolic representation contribute to engaging in problem solving in [making stochastic inferences]
Statistical Significance / Bias and Variance / Overfitting Selection Bias	[CC6.SP.A] Develop understanding of statistical variability [CCSS7.SP.A] Use random sampling to draw inferences about a population. [CCHSS.IC.B] Make inferences and justify conclusions from sample surveys, experiments, and observational studies	[102] In science, technology and everyday life, variation and its associated uncertainty is a given… recognising the place of variation in the real world including, having a sense of the quantification of that variation, and acknowledging its uncertainty and error in related inferences
Precision, Recall, and Accuracy	[CCHSS.CP.A.4] Construct and interpret two-way frequency tables of data when two categories are associated with each object being classified. [CCHSS.CP.B] Use the rules of probability to compute probabilities of compound events	[106] Identifying conditional decisions making as a focal point of the uncertainty and data content category… appreciate how the formulation of the analysis in a model impacts the conclusions that can be drawn and that different assumptions/relationships may well result in different conclusions
Information Theory	[CCHSF.BF.B.5] Understand the inverse relationship between exponents and logarithms	

Sources

[1] National Governors Association Center for Best Practices, Council of Chief State School Officers. “Common Core State Standards for Mathematics”. *National Governors Association Center for Best Practices, Council of Chief State School Officers*. Washington D.C., (2010).

[2] Organisation for Economic Co-operation and Development. “PISA 2021 Mathematics Framework (Second Draft)”. *Programme for International Student Assessment*, (2018-19).

[3] Hawksworth, John. “AI and robots could create as many jobs as they displace”. *World Economic Forum: Annual Meeting of the New Champions*, (2018).

[4] West, Darrell M. “What is artificial intelligence?”. *Brookings Institution: A Blueprint for the Future of AI*, (2018).

[5] BAE Systems. “Wanted: AI Translator (to start in 2040)”. (2019)

[6] Boger, Tal. “Defining NBA players by role with k-means clustering”. *Dribble Analytics*, (2019).

[7] Greenfield, Daniel. “Artificial Intelligence in Medicine: Applications, implications, and limitations”. *Harvard Graduate School of the Arts and Sciences: Science in the News,* (2019).

[8] Beck, Andrew & Koller, Daphne, et al. “Systematic Analysis of Breast Cancer Morphology Uncovers Features Associated with Survival”. *American Association for the Advancement of Science,* (2011).

[9] Payne, Blakeley H. “An Ethics of Artificial Intelligence Curriculum for Middle School Students”. *MIT Media Lab: Personal Robotics* Group, (2019).

[10] National AI Circle. “Building a YouTube Video Recommender”. *National AI Circle: HeadFirst AI Summer Workshops*, (2020).

[11] Goodman, Noah D., Tenenbaum, Joshua B. & The ProbMods Contributors. “Probabilistic Models of Cognition (2nd ed.)” https://prodmods.org, (2016).

[12] Boaler, Jo. “youcubed K-12 Data Science” and “21st Century Teaching and Learning”. *Stanford Graduate School of Education: youcubed*. Retrieved 10/16/2020.