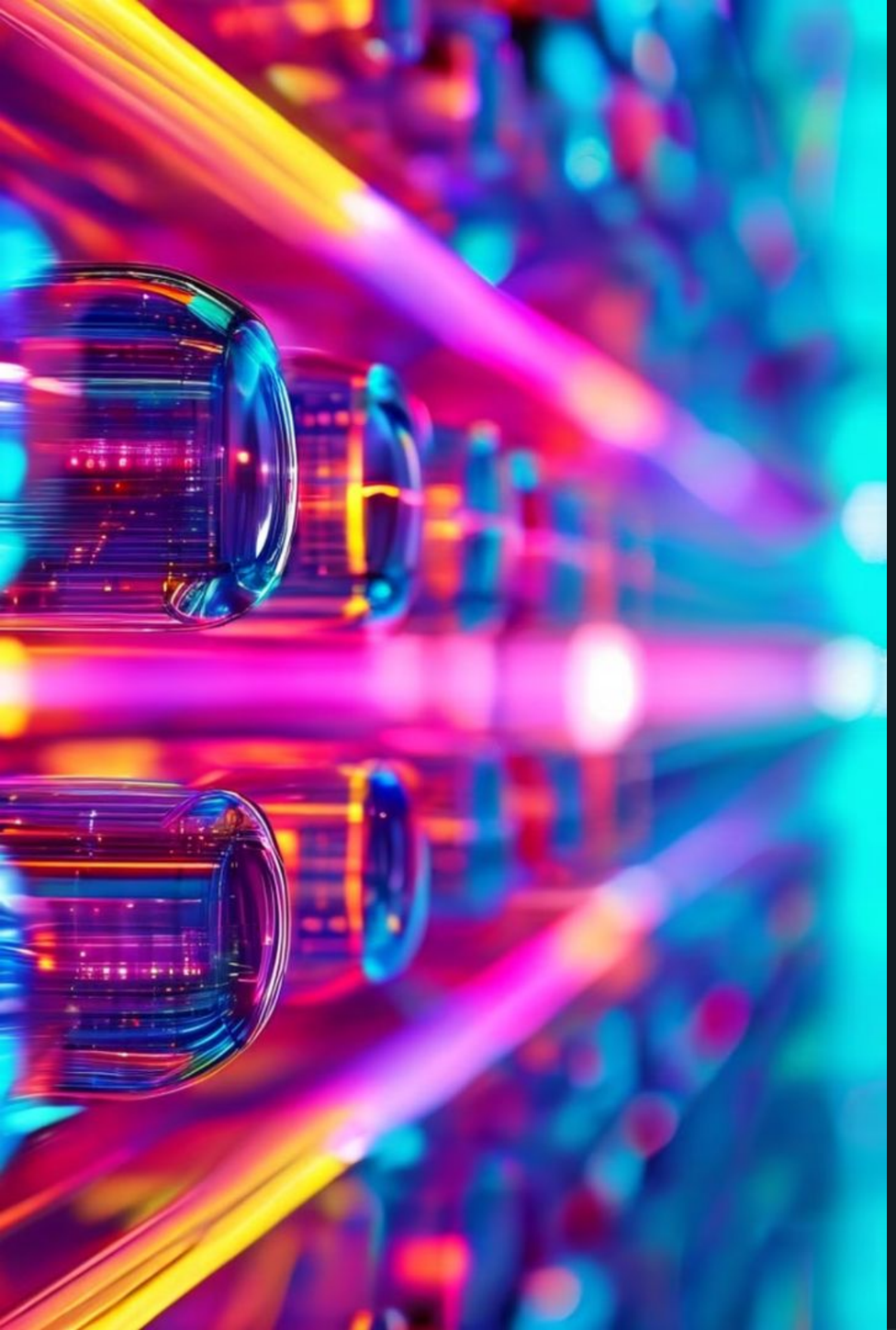
NumPy: The Foundation of Data Science

NumPy is a powerful Python library for numerical computing. It provides high-performance multidimensional arrays and a wide range of mathematical functions to operate on them.

by Adamya Neupane



Creating Arrays

NumPy arrays are the core data structure. They can be created from lists, tuples, or other arrays.

1

Array Creation

Arrays can be initialized with various methods, including using the `array()` function, `arange()` for ranges, and `zeros()` or `ones()` for pre-filled arrays.

2

Multidimensional Arrays

NumPy supports arrays with any number of dimensions, allowing you to work with complex data structures.

3

Data Types

Specify the data type of elements in the array for optimal performance and memory management.

Array Operations

NumPy enables efficient mathematical operations on arrays, working element-wise.

Arithmetic Operations

Perform addition, subtraction, multiplication, division, and other arithmetic operations on entire arrays or specific elements.

Logical Operations

Use logical operators like `and`, `or`, and `not` to filter and manipulate arrays based on conditions.

Universal Functions

NumPy provides a collection of universal functions (ufuncs) that operate element-wise on arrays, enabling efficient calculations like trigonometric functions, exponents, and logarithms.

Matrix Operations

NumPy efficiently handles matrix operations, which are fundamental in linear algebra and data analysis.

Operation	Description
Transpose	Swaps rows and columns of a matrix.
Inverse	Calculates the inverse of a square matrix.
Dot Product	Performs matrix multiplication, essential for linear transformations and solving systems of equations.

Statistical Operations

NumPy provides powerful functions to analyze and summarize data within arrays.

1

Mean

Calculates the average of elements in an array.

2

Median

Finds the middle value in a sorted array.

3

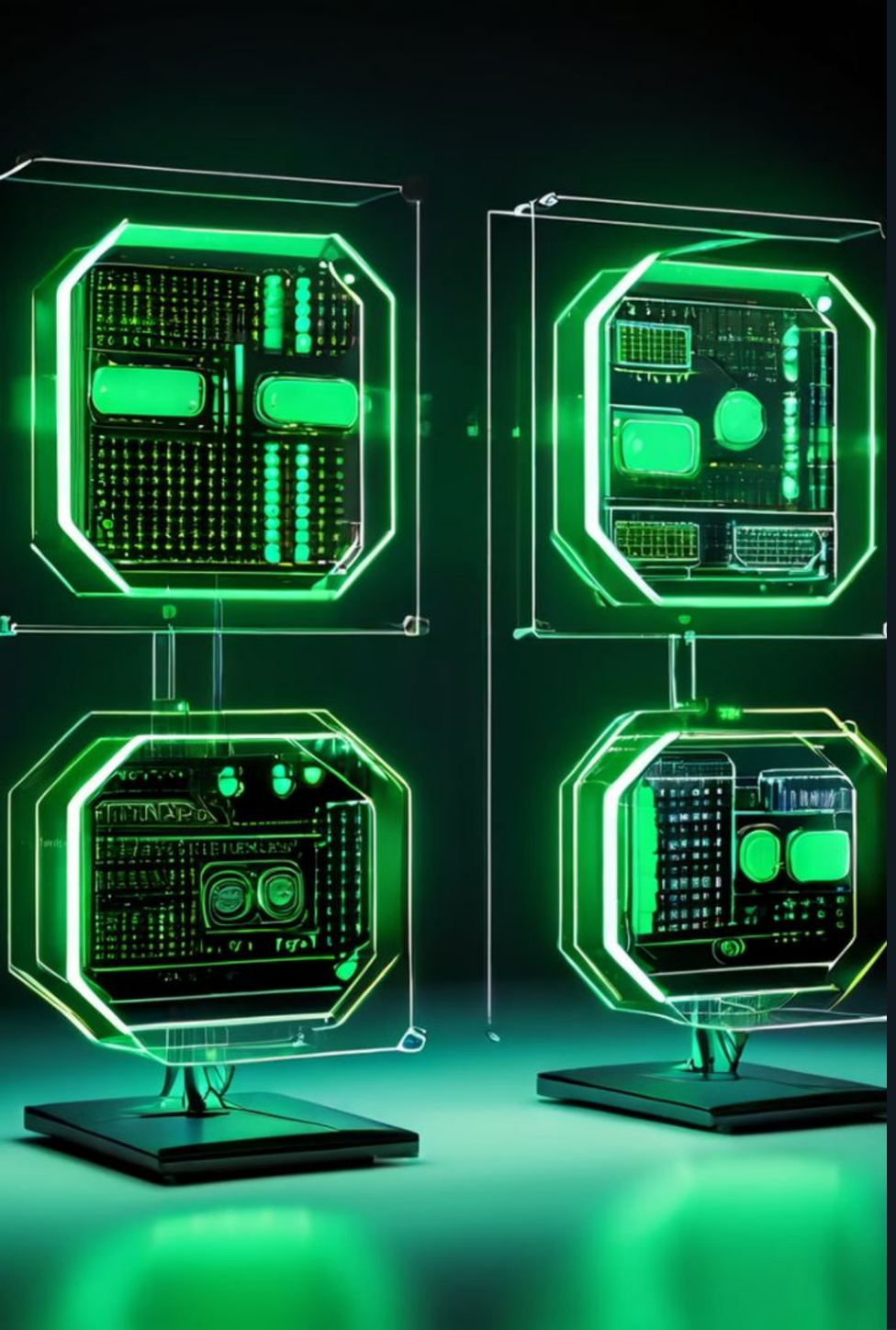
Standard Deviation

Measures the spread of data points around the mean.

4

Variance

Calculates the average squared deviation of data points from the mean.



Broadcasting

Broadcasting simplifies operations between arrays of different shapes by automatically expanding the smaller array to match the larger one.

Shape Compatibility

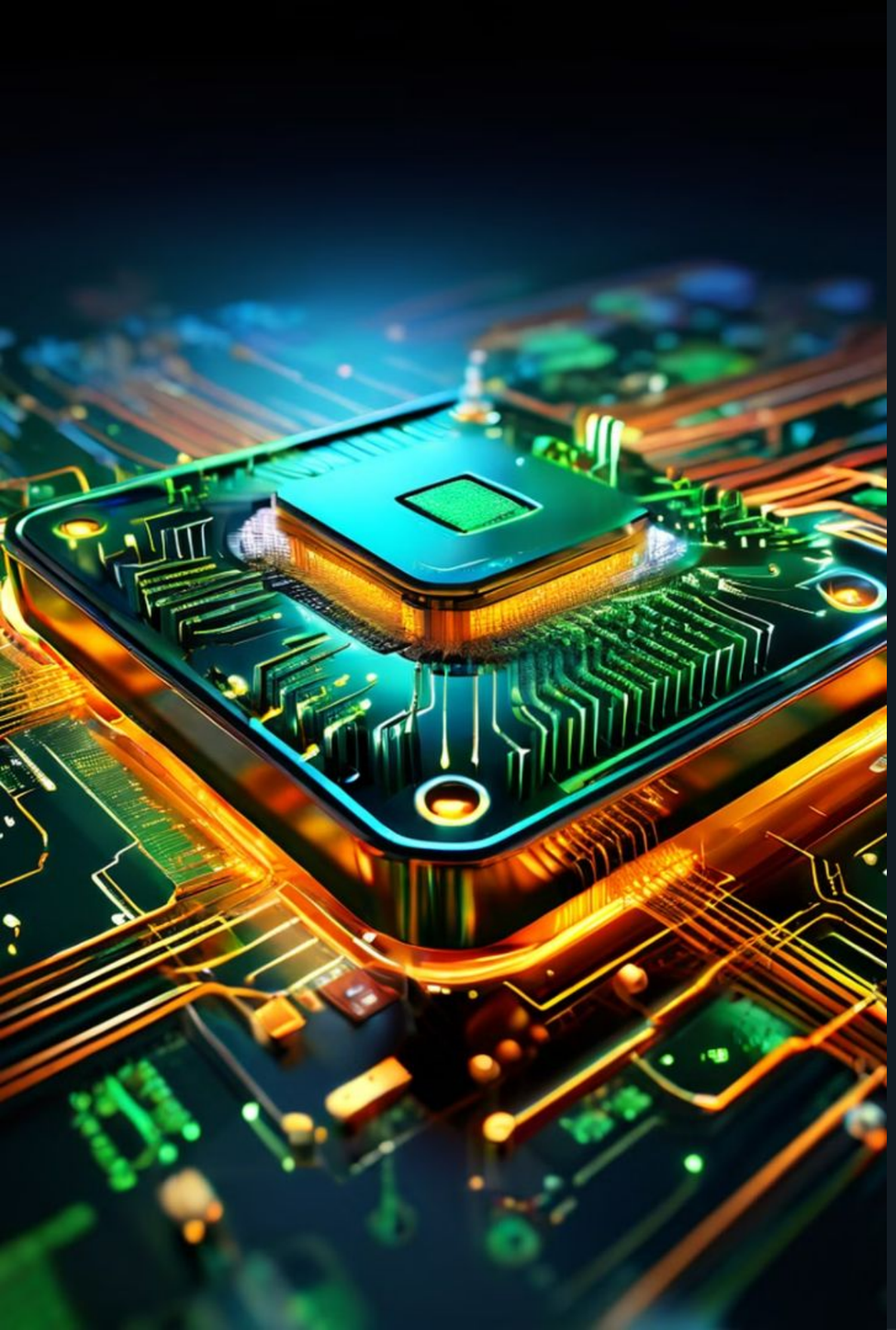
Broadcasting works when the arrays have compatible shapes, either matching dimensions or one dimension being 1.

Efficiency

Broadcasting avoids unnecessary memory copies and computations, making operations on differently shaped arrays efficient.

Element-wise Operations

After broadcasting, element-wise operations can be applied to the arrays.



Optimization and Performance

NumPy's performance is optimized for efficient numerical operations.

Vectorized Operations

NumPy's vectorized operations eliminate Python loops, resulting in significantly faster execution times.

Memory Efficiency

Arrays store data contiguously in memory, allowing for efficient access and manipulation.

Optimized Libraries

NumPy is built upon highly optimized C and Fortran libraries, providing a performance boost for numerical computations.

Conclusion

NumPy is a fundamental library for data science, offering a powerful and efficient platform for numerical computing. Its versatile features empower developers and researchers to tackle complex data analysis and scientific computation tasks effectively.