

Iverson's New Notation for AI

(Draft Reference Card)

Symbol	Name	Meaning / Example
X	Data	Input array (samples $\times$ features)
Y	Labels	Target outputs (samples $\times$ classes)
θ	Parameters	Model weights (array, same algebra as data)
∇	Gradient	$\nabla \text{ loss } f = \text{gradient of loss wrt } f$
η	Learning rate	Scalar step size
$\uparrow n$	Stack	$\text{relu } \uparrow 3 \circ \text{linear} \rightarrow \text{3-layer MLP}$
$\otimes$	Linear transform	$X \otimes \theta = \text{matrix multiply}$
$\circ.$	Composition	$f \circ. g = \text{apply } g \text{ then } f$
$\circledast$	Attention	$f \circledast g = f \text{ applied with attention guided by } g$
learn	Learning operator	learn $f(X, Y) \eta$ updates f on data
$\oplus$	Merge / Concat	$F \oplus G = \text{join models (ensembling)}$
$\bar{;}$	Stack vertically	Join model blocks
$\subset$	Submodel	$F \subset 2 = \text{second layer of model } F$
$\equiv$	Loss function	$\equiv(Y, f X) = \text{loss between prediction and target}$

Model Expressions

$f \leftarrow X \otimes \theta \circ. \text{linear model}$
 $g \leftarrow f \circ. \text{softmax} \circ. \text{add softmax}$
 $h \leftarrow \text{learn } g(X, Y) \eta \circ. \text{train a model}$
 $F \leftarrow \text{relu } \uparrow 5 \circ. \text{linear} \circ. \text{deep net}$
 $A \leftarrow f \circledast g \circ. \text{attention block}$
 $F \subset 2 \circ. \text{inspect second layer}$
 $F \subset 2 \leftarrow \text{norm} \circ. \text{replace with normalization}$

Principles

1. Everything is an array: data, weights, models.
2. Operators lift verbs: stacking, learning, attention are adverbs/conjunctions.
3. Notation = Semantics: each symbol corresponds to algebra, not magic.
4. Interactive by design: every expression can be tried and observed instantly.
5. Transparent AI: models are readable, auditable, swappable.