Common Mistakes Students Make

WRONG

CORRECT

WRONG	CORRECT
Vectors are arrows	Vectors are LISTS of numbers (arrows are just one visualization)
Attention copies words	Attention creates WEIGHTED AVERAGES (blends information, not copies)
Softmax changes total value	Softmax creates PROBABILITIES (sum always = 100%)
More heads = more accuracy	More heads = more PERSPECTIVES (not automatically better)
Transformers understand meaning	Transformers find PATTERNS (statistical correlations, not understanding)
Position encoding adds dimensions	Position encoding MODIFIES existing dims (same 8 dimensions before and after)