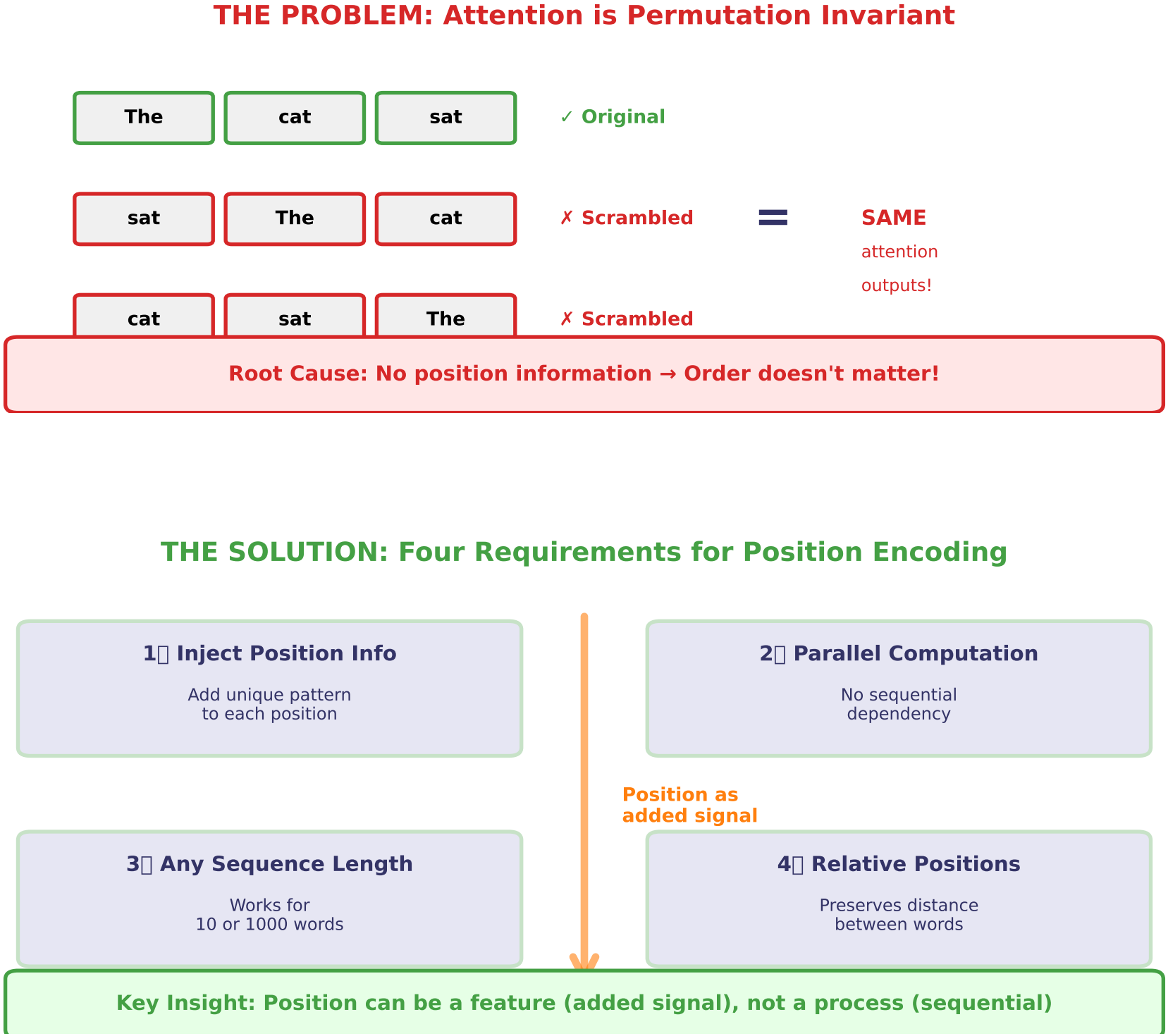
THE PROBLEM: Attention is Permutation Invariant
The
cat
sat
✓ Original
sat
The
cat
✗ Scrambled
=
SAME
attention
outputs!
cat
sat
The
✗ Scrambled
Root Cause: No position information → Order doesn't matter!
THE SOLUTION: Four Requirements for Position Encoding
1 Inject Position Info
Add unique pattern
to each position
2 Parallel Computation
No sequential
dependency
Position as
added signal
3 Any Sequence Length
Works for
10 or 1000 words
4 Relative Positions
Preserves distance
between words
Key Insight: Position can be a feature (added signal), not a process (sequential)