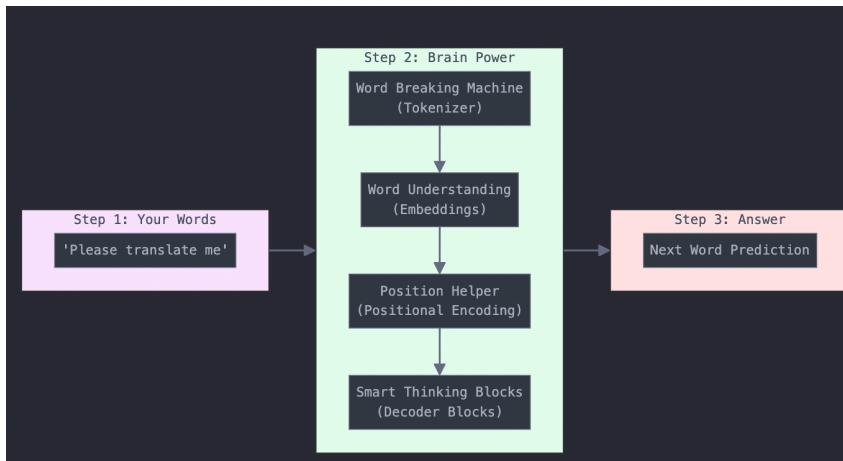
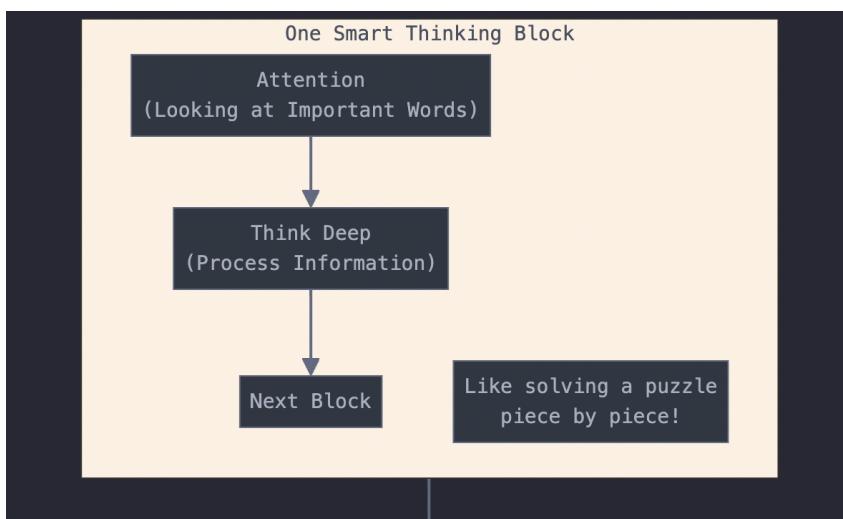


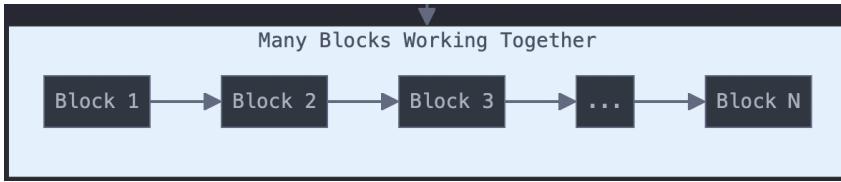
# GPT

Friday, 17 January 2025 7:47 PM



- **Word Breaking (Tokenization)**
  - Think of it like breaking a big chocolate bar into smaller pieces
  - Each piece is called a "token"
  - Example: "Please translate me" → "Please" "trans" "late" "me"
- **Word Understanding (Embeddings)**
  - Like giving each word a special code that shows what it means
  - Similar words get similar codes
  - Just like how "happy" and "joyful" mean almost the same thing!
- **Remember the Order (Positional Encoding)**
  - Like numbering steps in a recipe
  - Helps remember which word comes first, second, third...
  - Very important for understanding the meaning!
- **Smart Thinking Blocks (Decoder Blocks)**
  - Think of these as different experts working together
  - Each block:
    - Looks at important words (Attention)
    - Thinks deeply about them (Feed Forward)
    - Passes information to the next expert
  - They work together like a team solving a puzzle
- **Final Answer (Output)**
  - After all the blocks think about it
  - GPT picks the best next word
  - Like finishing someone's sentence!





## Decoder

1. **Think of a Decoder as a Super-Smart Reader**
  - o It's like having a friend who's really good at reading and understanding stories
  - o This friend helps you predict what comes next in the story
2. **Three Special Powers:**
  - o **Look Back Power** (Self-Attention)
    - Like having a perfect memory of what you just read
    - Example: "The cat sat on..." (looks back) "...the mat!"
  - o **Think Deep Power** (Feed Forward)
    - Like a student thinking carefully about what they read
    - Connects ideas together, just like solving a puzzle
  - o **Remember Power** (Memory)
    - Like taking really good notes while reading
    - Uses these notes to understand the next part better
3. **Team Work**
  - o Multiple decoders work together like study buddies
  - o Each one makes the understanding a little better
  - o First decoder: "I know the basic words!"
  - o Second decoder: "I understand the sentences!"
  - o Third decoder: "I get the whole story!"

Fun Examples to Remember:

1. **Restaurant Order Example:**
  - o First decoder: Reads the menu items
  - o Second decoder: Understands food combinations
  - o Third decoder: Predicts what you might want for dessert!

## Tokenization

1. **What is Tokenization?**
  - o Converting text into numbers the AI can understand
  - o Like translating English into "Computer Language"
  - o You can try it yourself on OpenAI's tokenizer tool!
2. **Token Types:**
  - o **Single-Token Words**
    - Most common words ("cat", "dog", "house")
    - Each gets one unique number
  - o **Multi-Token Words**
    - Long words split into pieces
    - Example: "indivisible" → "ind" + "iv" + "isible"
    - Like breaking chocolate into smaller pieces
  - o **Special Tokens**
    - Punctuation (., !, ?)
    - Numbers (123, 456)
    - Emojis and Unicode characters
    - Contractions (I'm, don't)
3. **Why This Matters:**
  - Helps AI understand language better
  - Makes processing more efficient
  - Handles new or unusual words
  - Captures meaning relationships
4. **Where to See It in Action:**
  - OpenAI's Tokenizer Tool
  - GPT Playground
  - When using ChatGPT (behind the scenes)

## Vector to Result

- **Final Vector to Probabilities**
  - Take the final vector from the decoder
  - Multiply it with the embedding matrix
  - Get probability scores for all possible tokens
- **Word Selection Process**
  - Each token gets a probability score
  - Higher score = More likely to be the next word
  - System can either:
    - Pick the highest probability (deterministic)
    - Use temperature to add randomness
- **Temperature Effects**
  - Low temperature (0.1): Always picks most likely word
  - Medium temperature (0.7): Some variety but still sensible
  - High temperature (1.5): More creative/random choices
- **Simple Example:** Input: "The cat sat \_\_" Process:
  1. Get vector from decoder
  2. Multiply with embeddings
  3. Get probabilities:
    - "on": 85%
    - "in": 10%
    - "at": 3%
  4. Choose based on temperature setting
    -