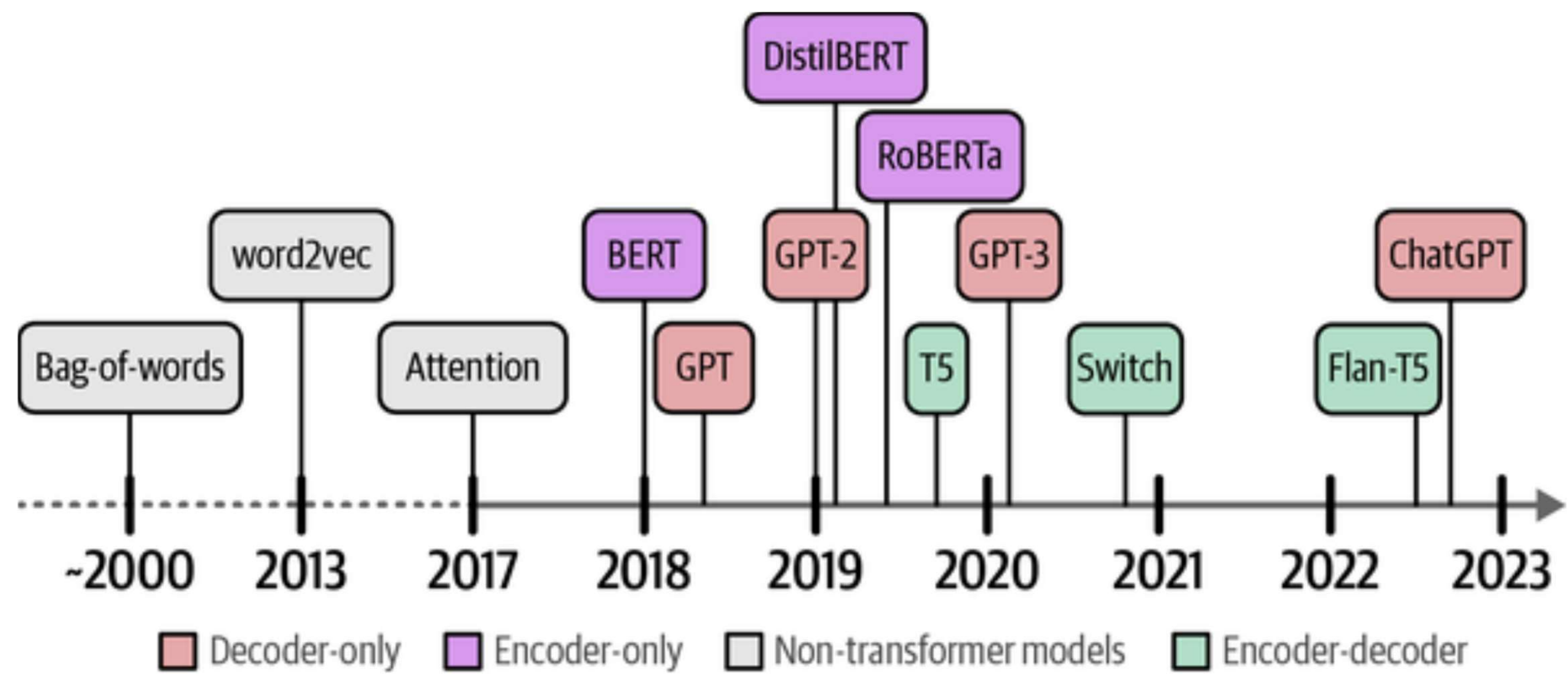


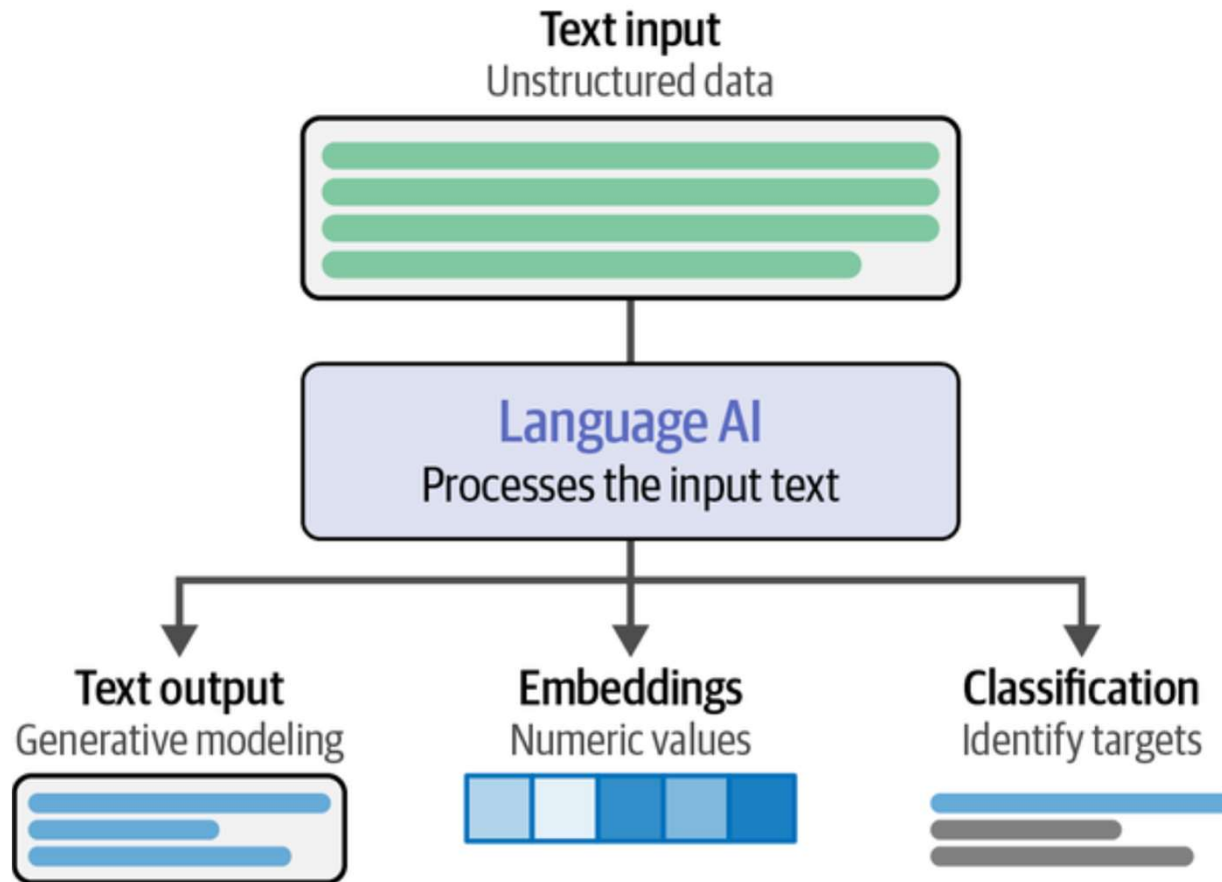
Generative AI and Agentic AI

LLM Fundamentals

A Recent History of Language AI



Language AI capabilities



Input

That is a cute dog



Split input by a **whitespace**

that

is

a

cute

dog

Input

My cat is cute



my

cat

is

cute

Tokenized sentences



Create a **vocabulary**



Input

My cat is cute

my cat is cute

Tokenization

Split input by a whitespace

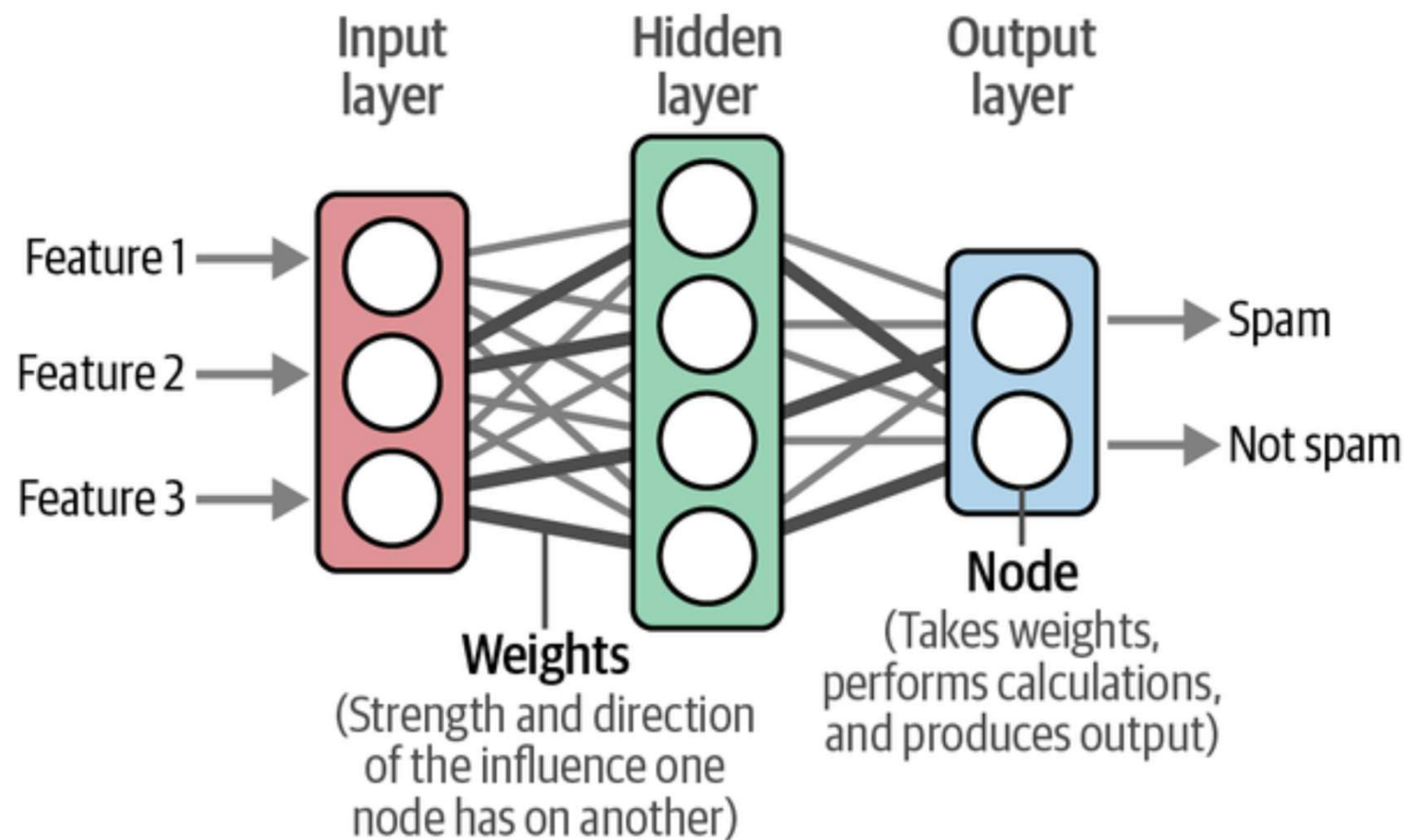
that	is	a	cute	dog	my	cat
0	1	0	1	0	1	1

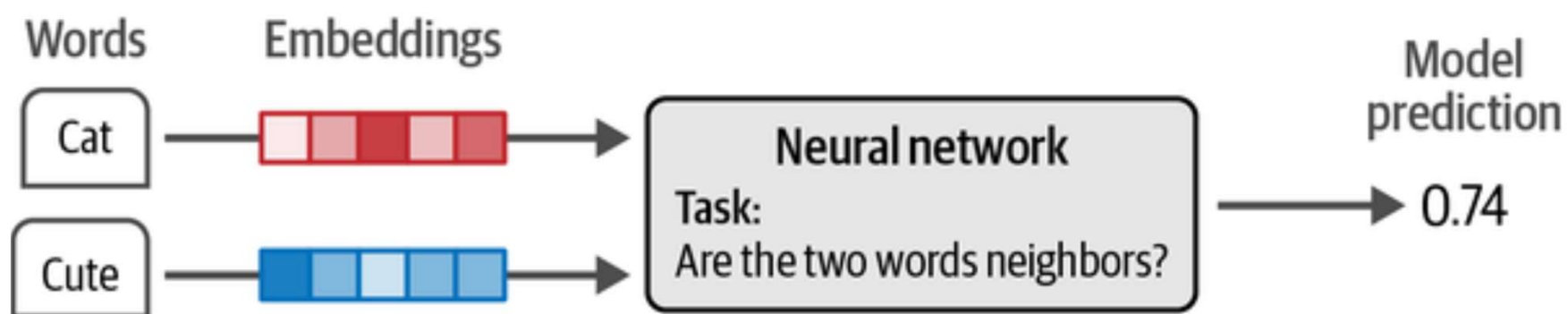
Bag-of-words

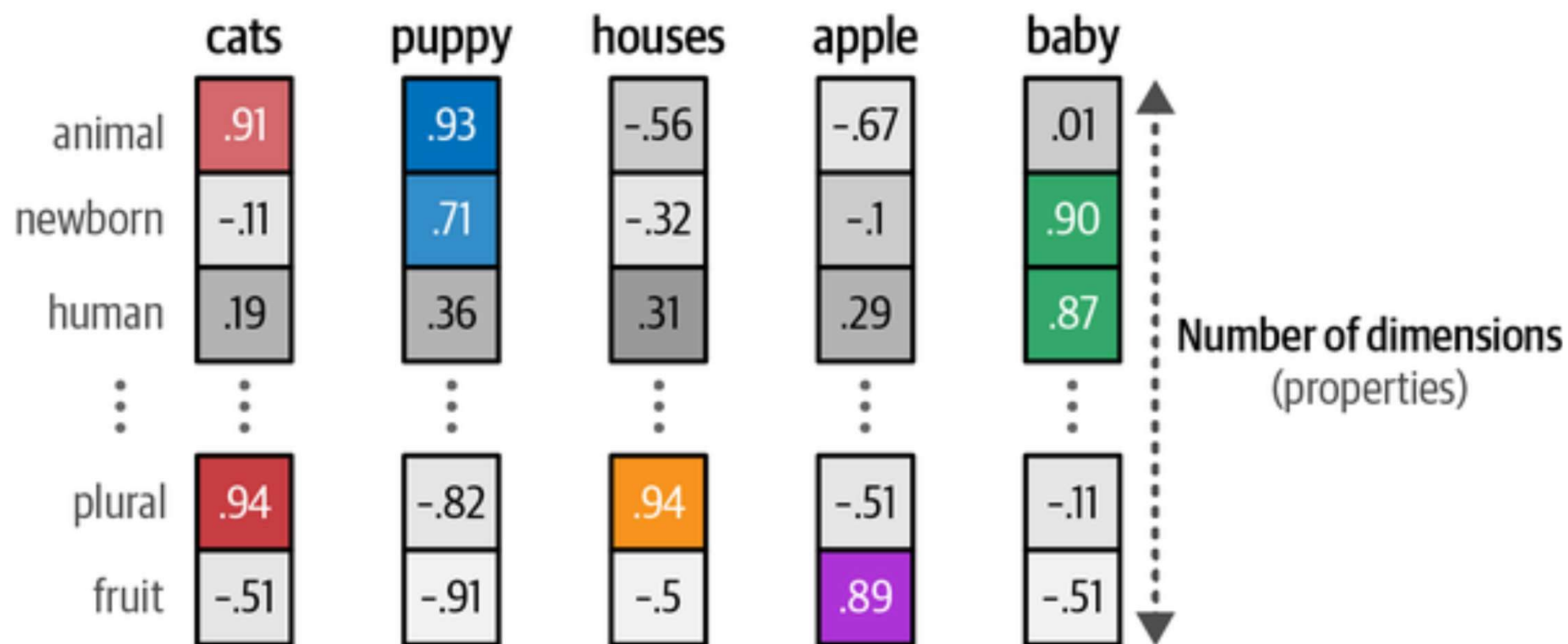
Count individual words

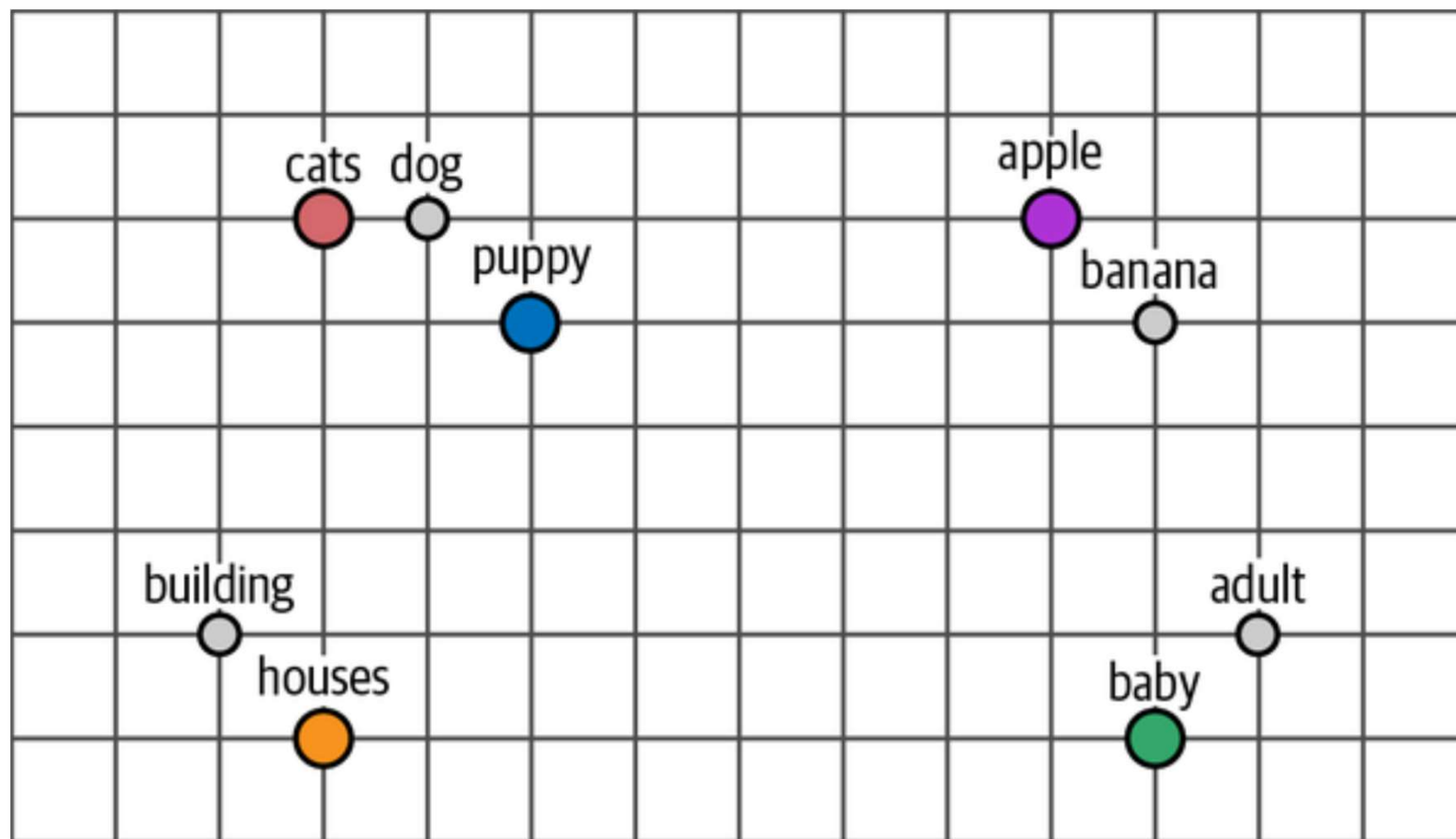


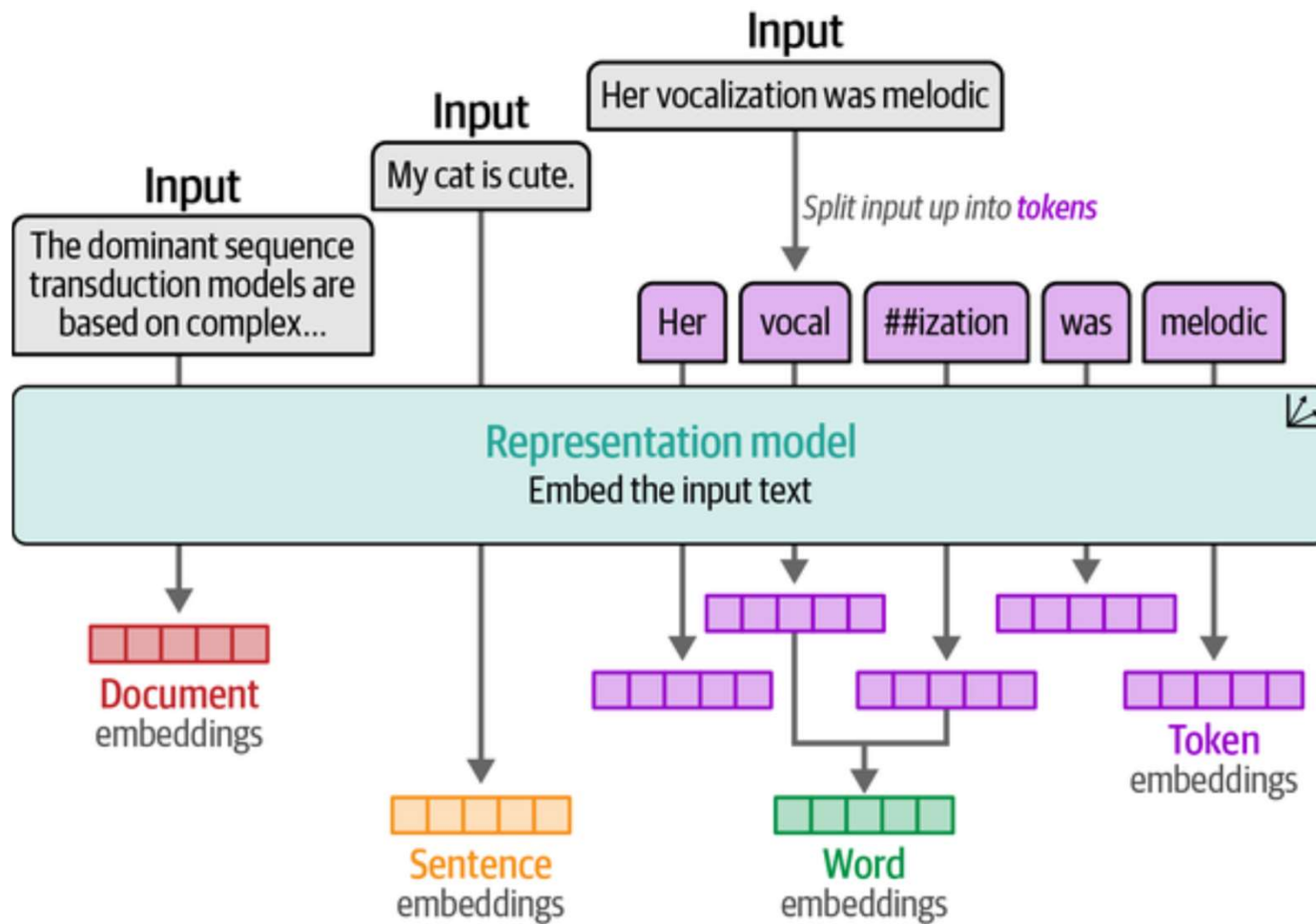
Vector representation

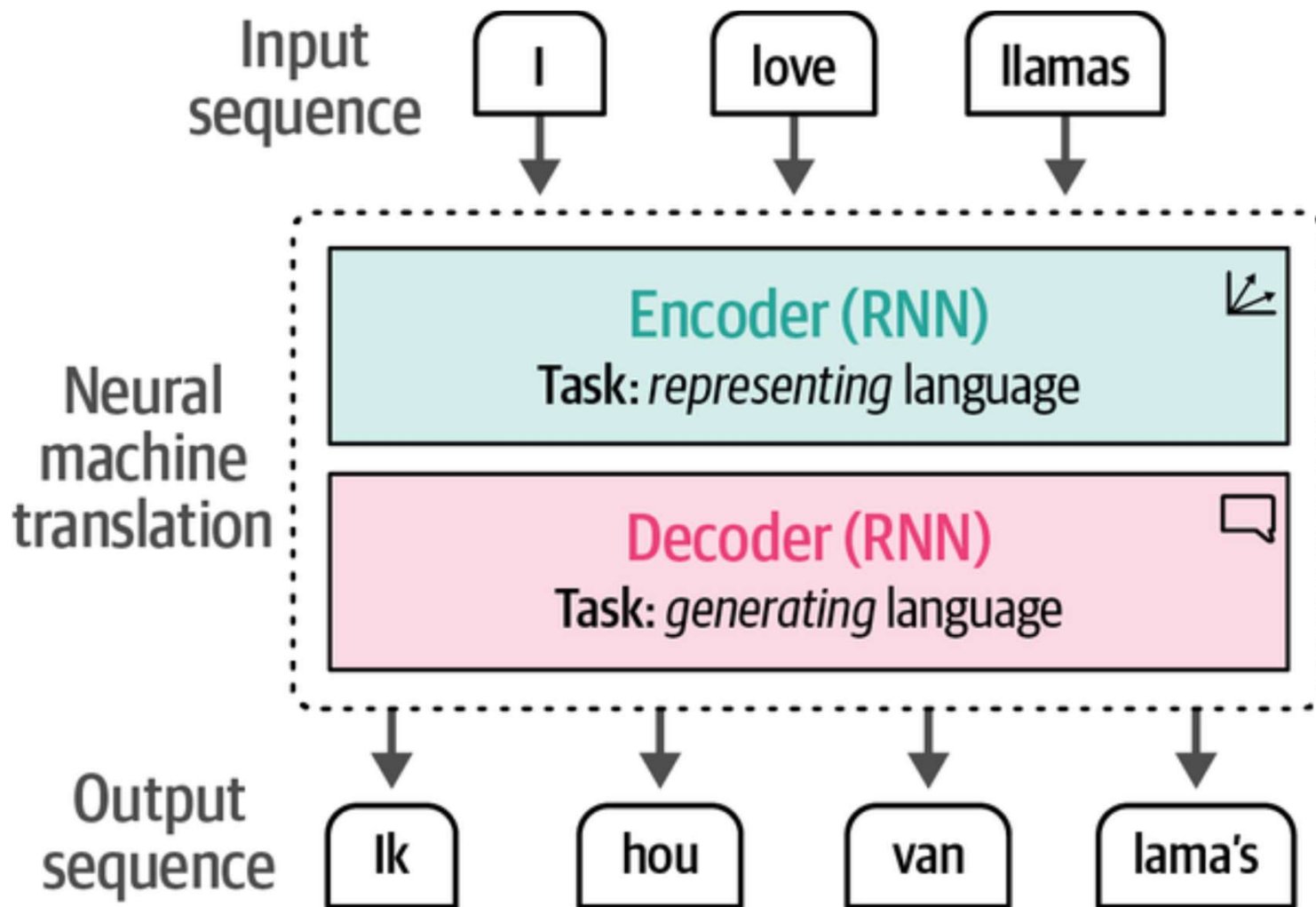


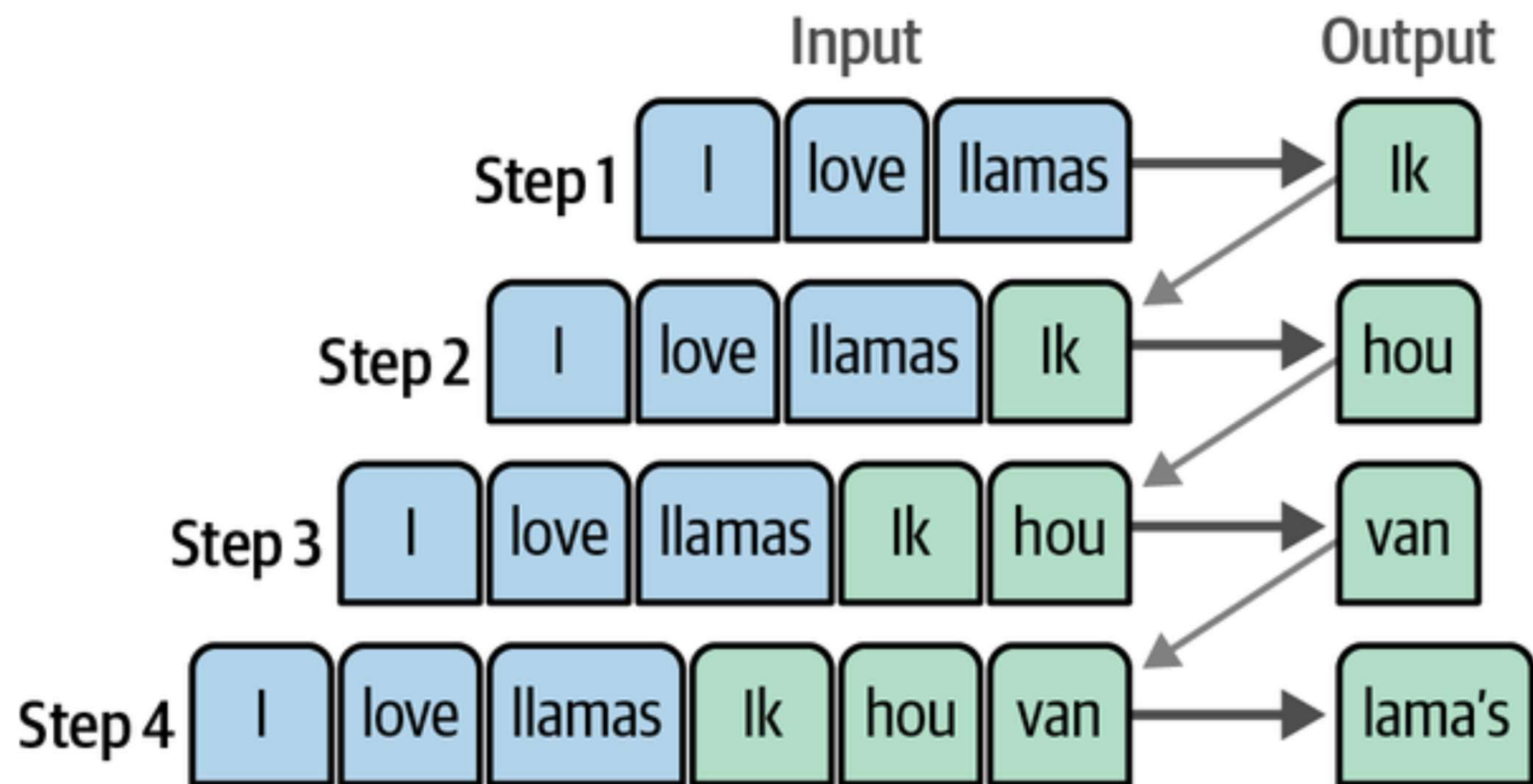


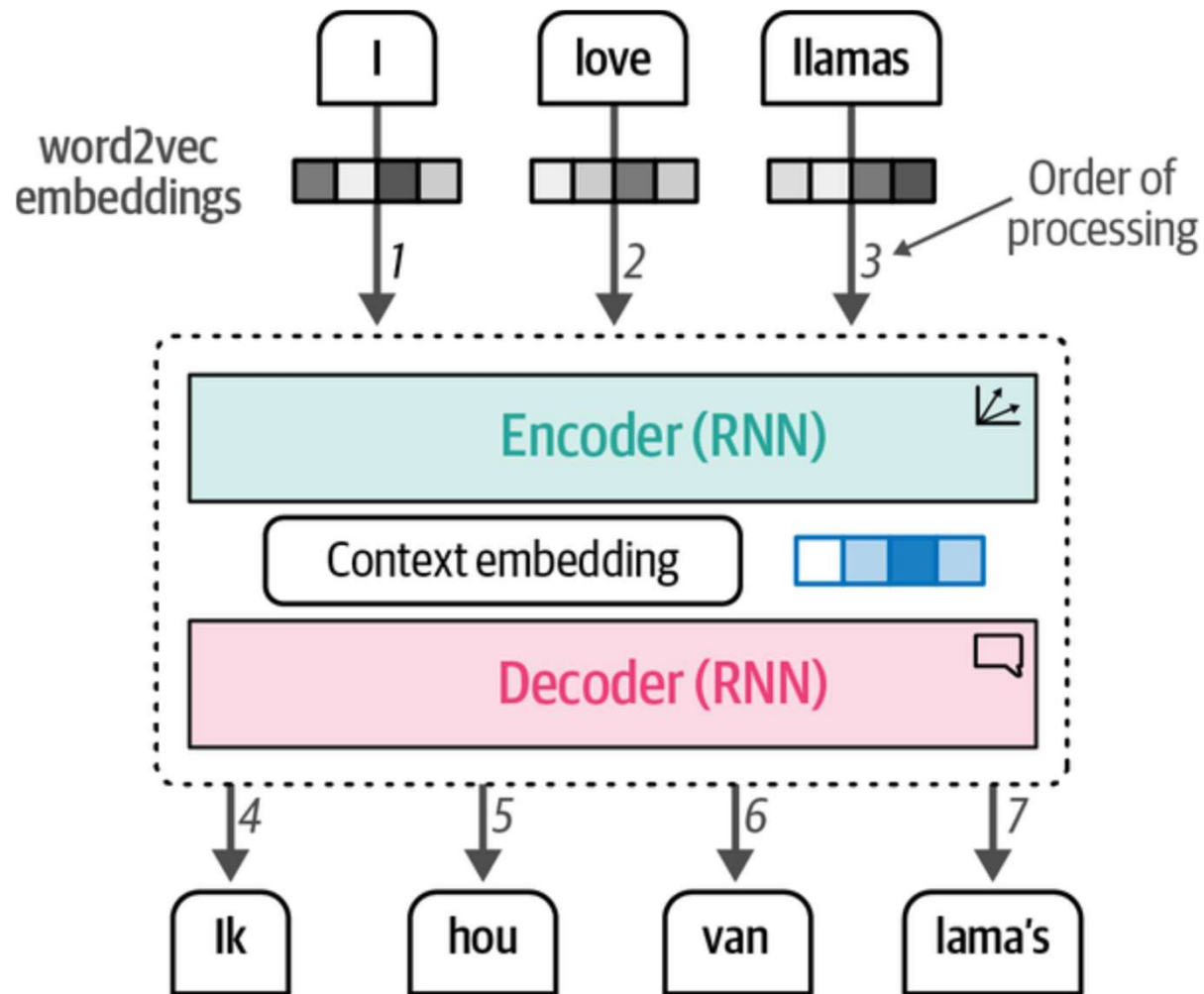




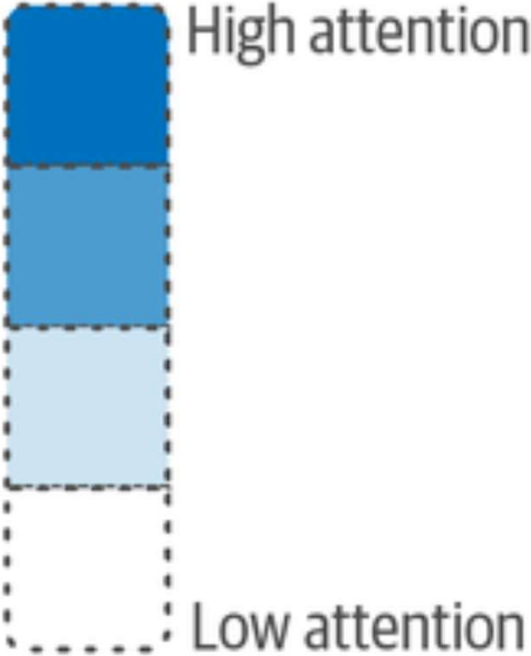
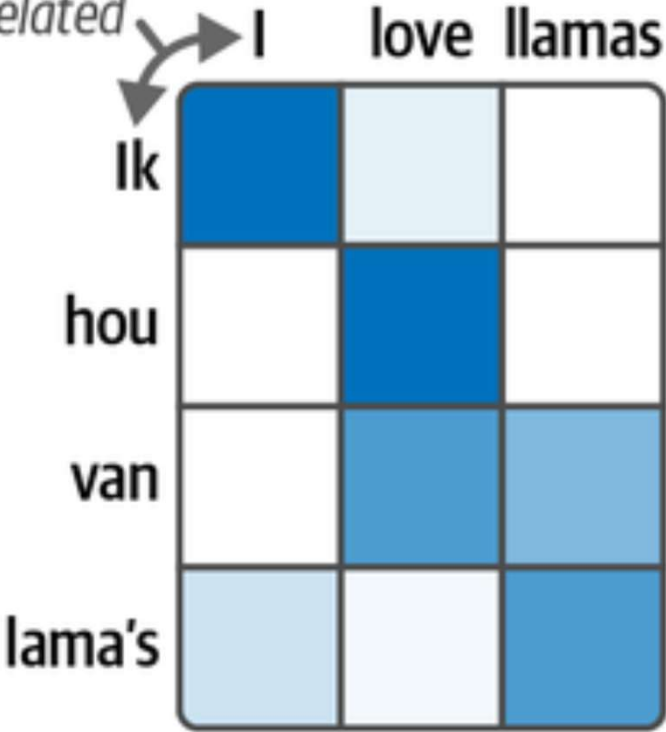


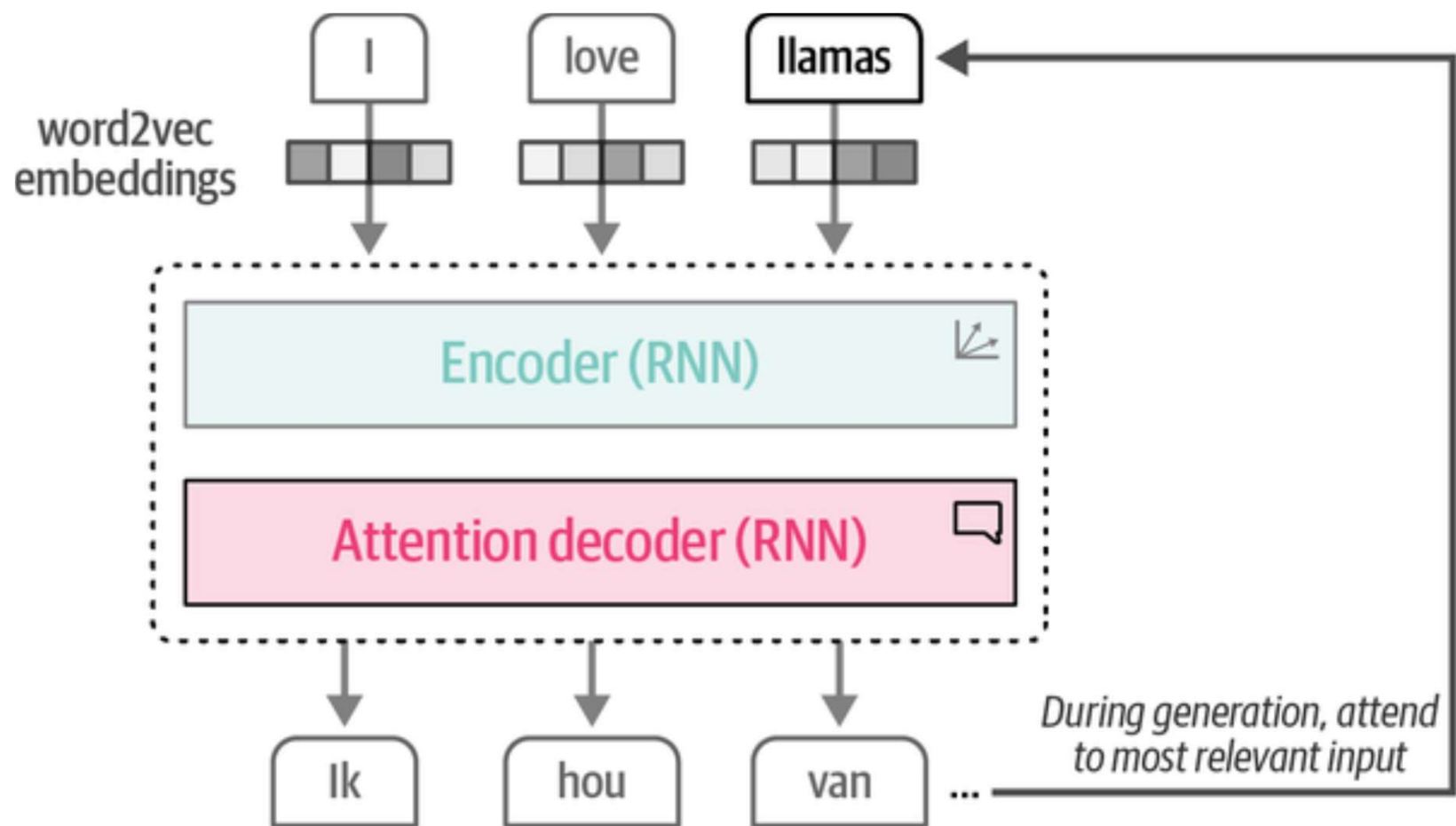




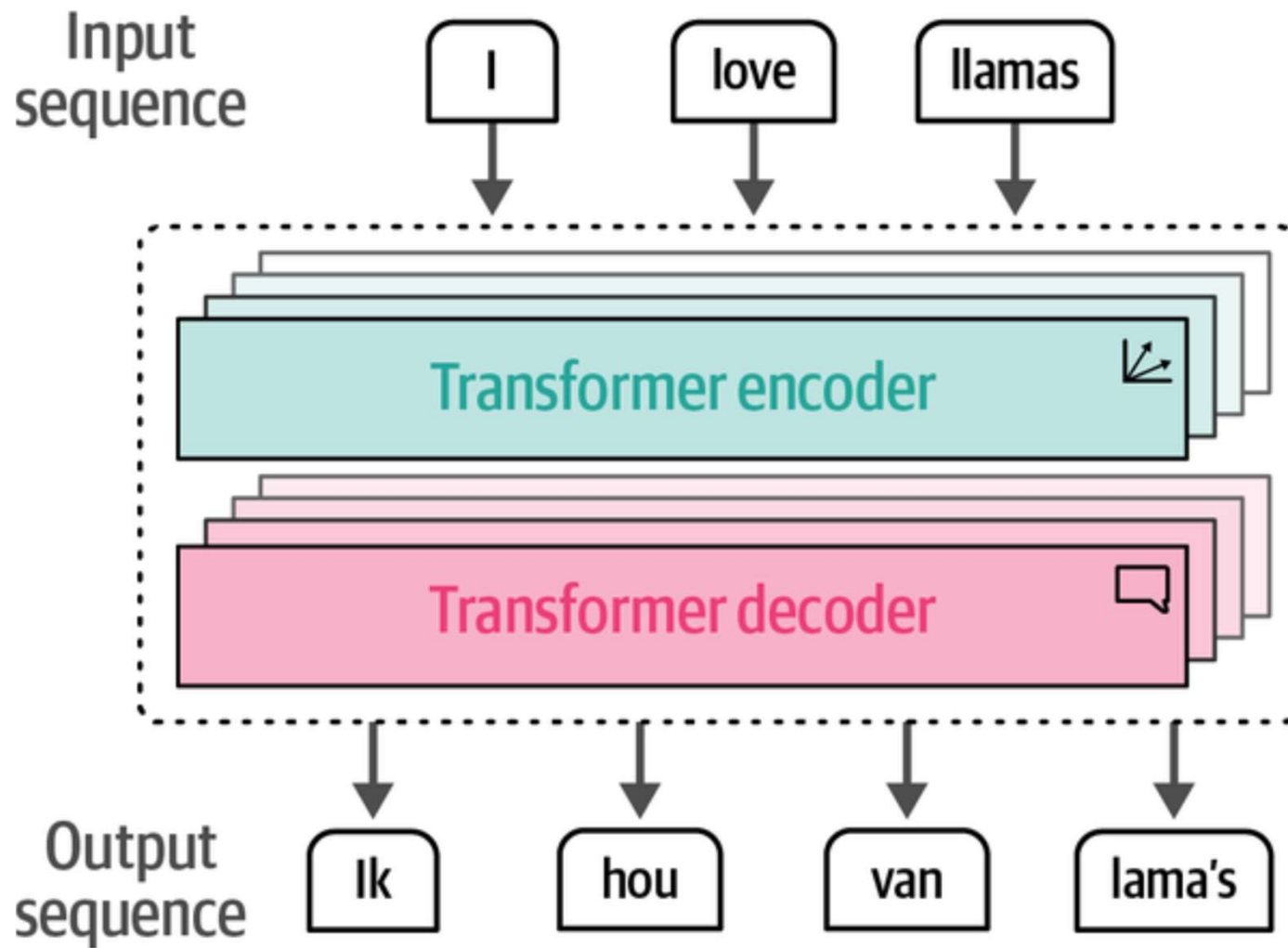


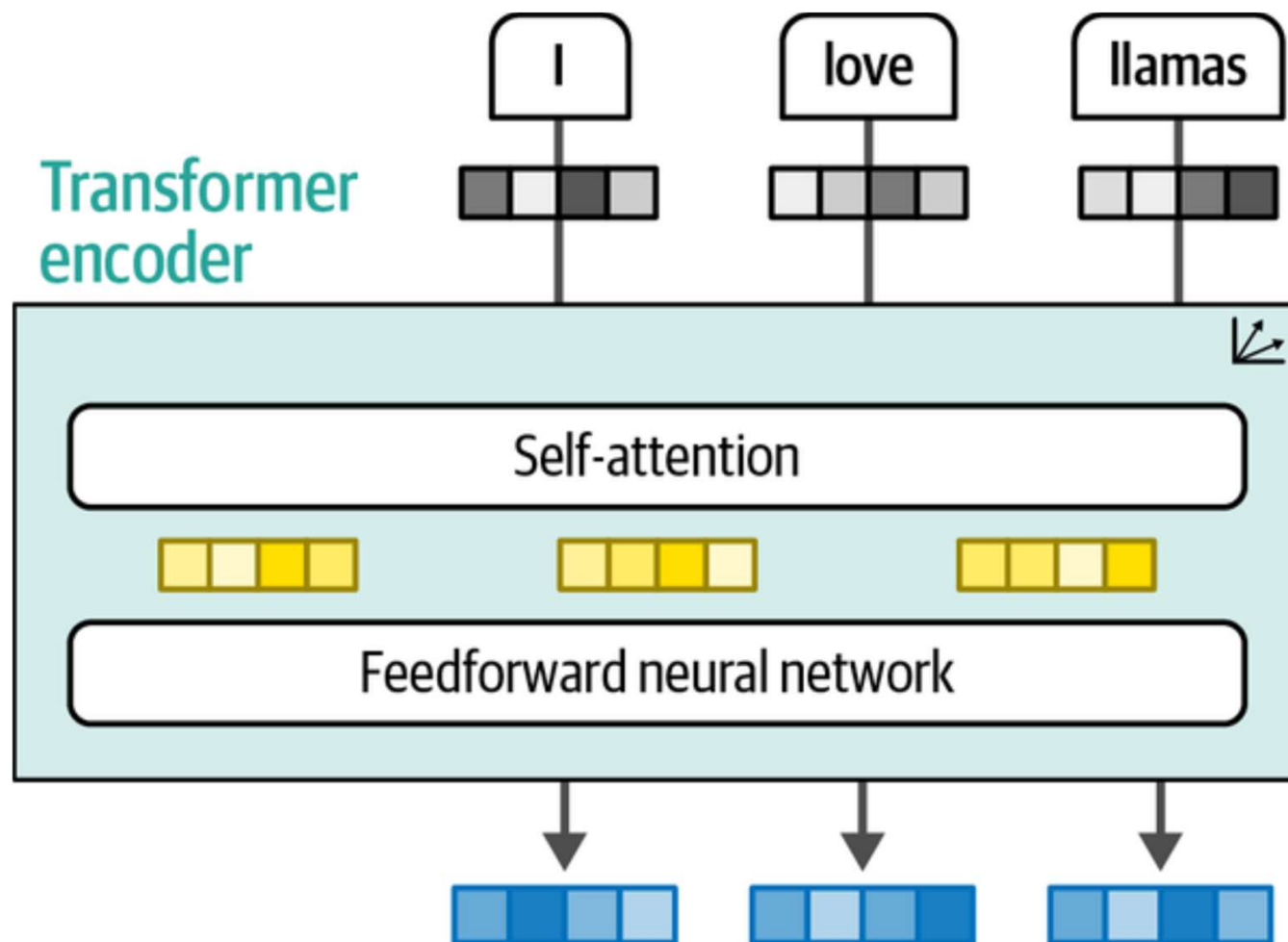
*Words with similar meaning
have higher attention weights
since they are highly related*





Transformer

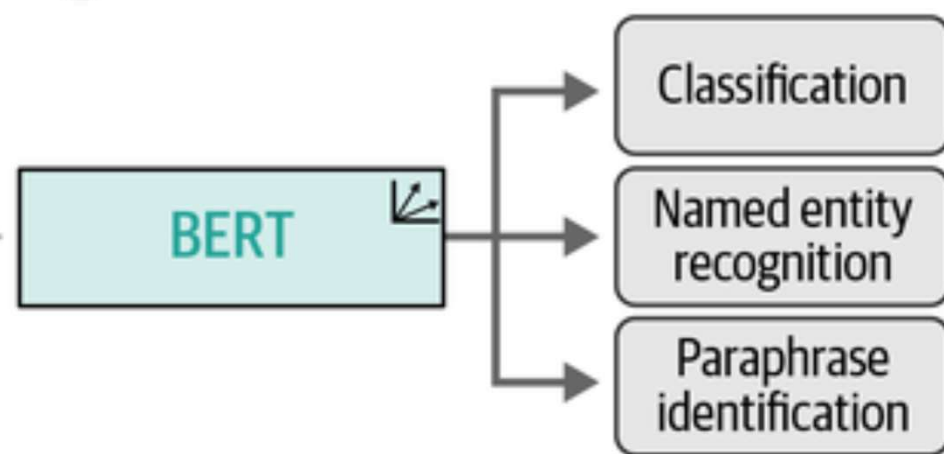




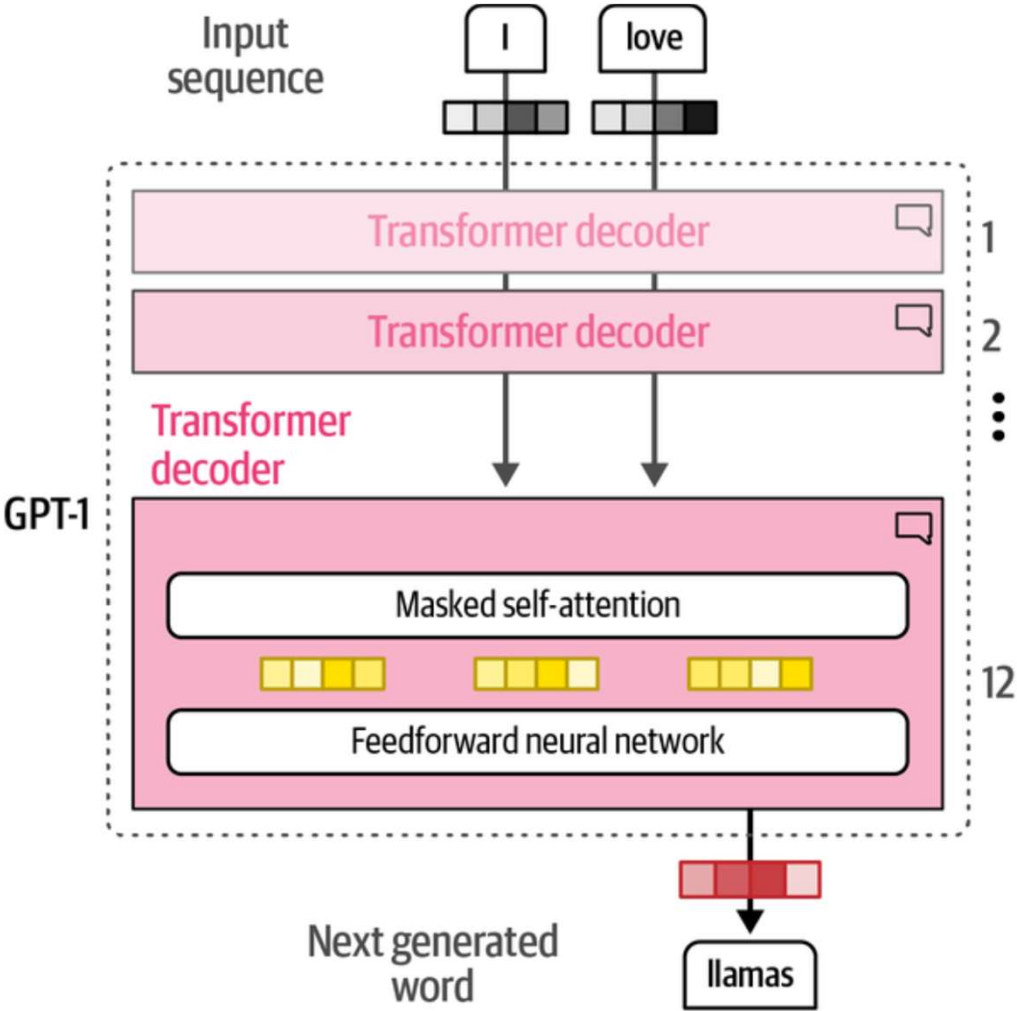
1 Pretrain on large dataset

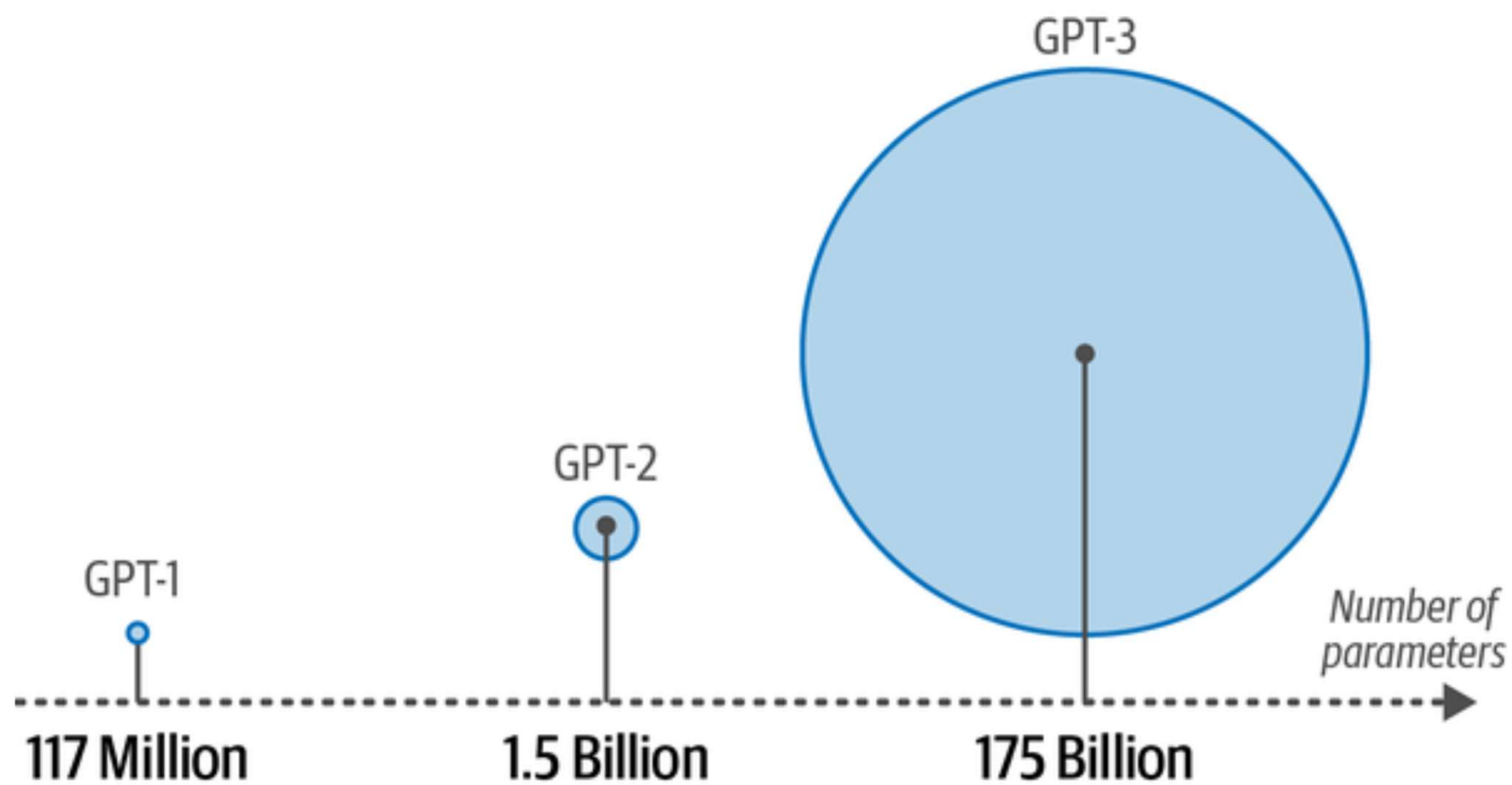


2 Fine-tune for downstream task



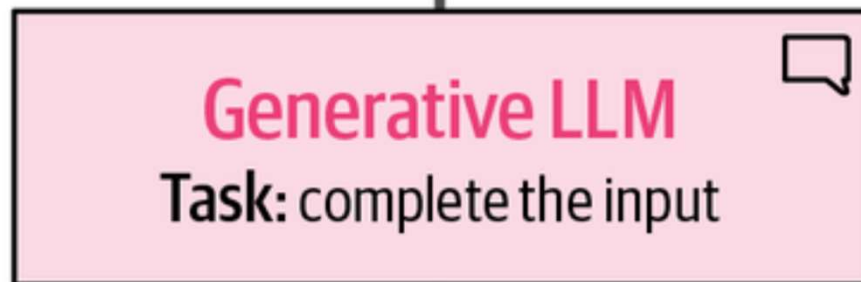
decoder-only architecture





User query
(*prompt*)

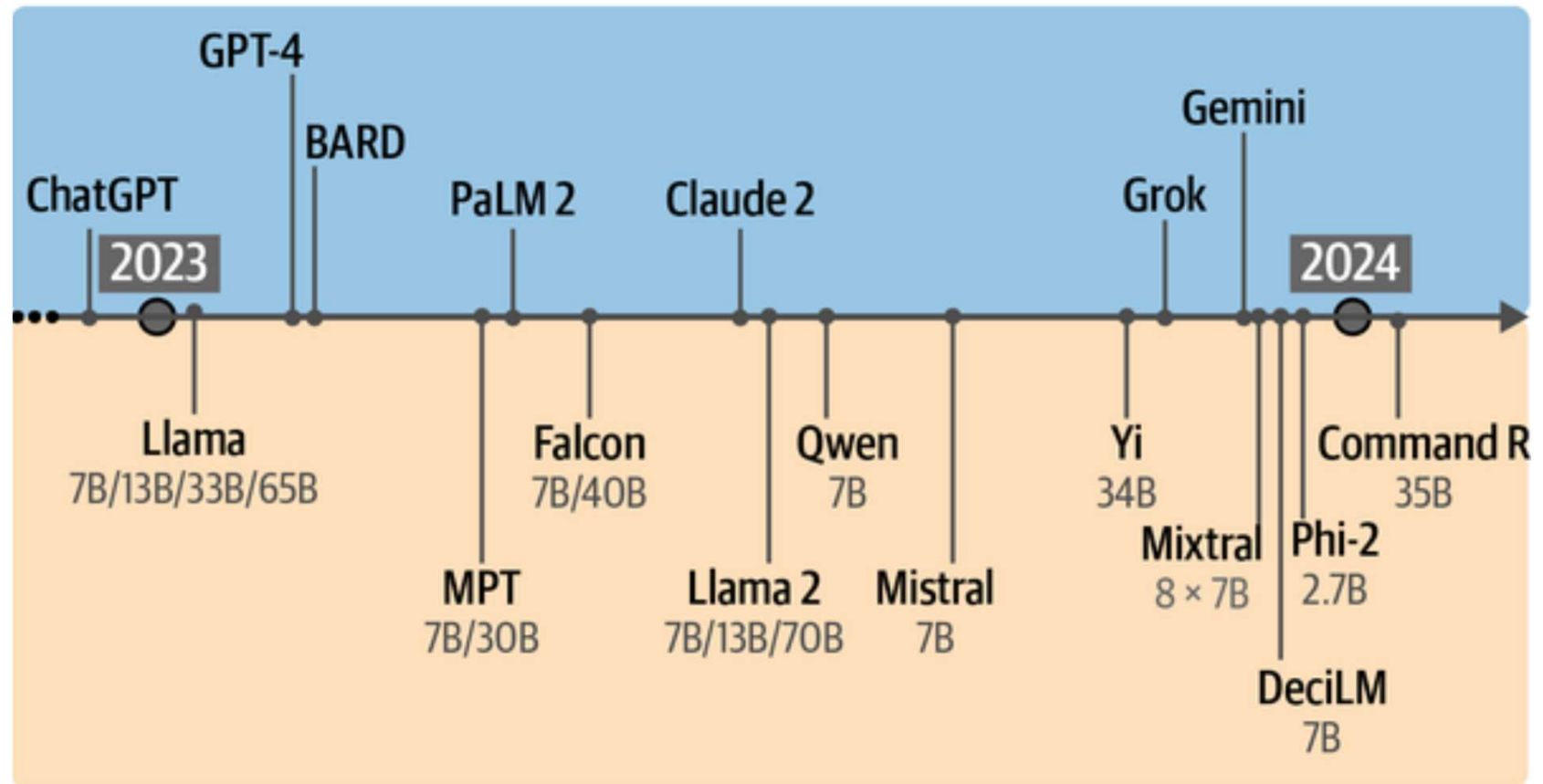
Tell me something about llamas



Output
(*completion*)

Llamas are domesticated South American camelids, widely used as pack animals by Andean cultures since pre-Hispanic times. With their fluffy coat, long neck, and distinctive facial features...

Proprietary models

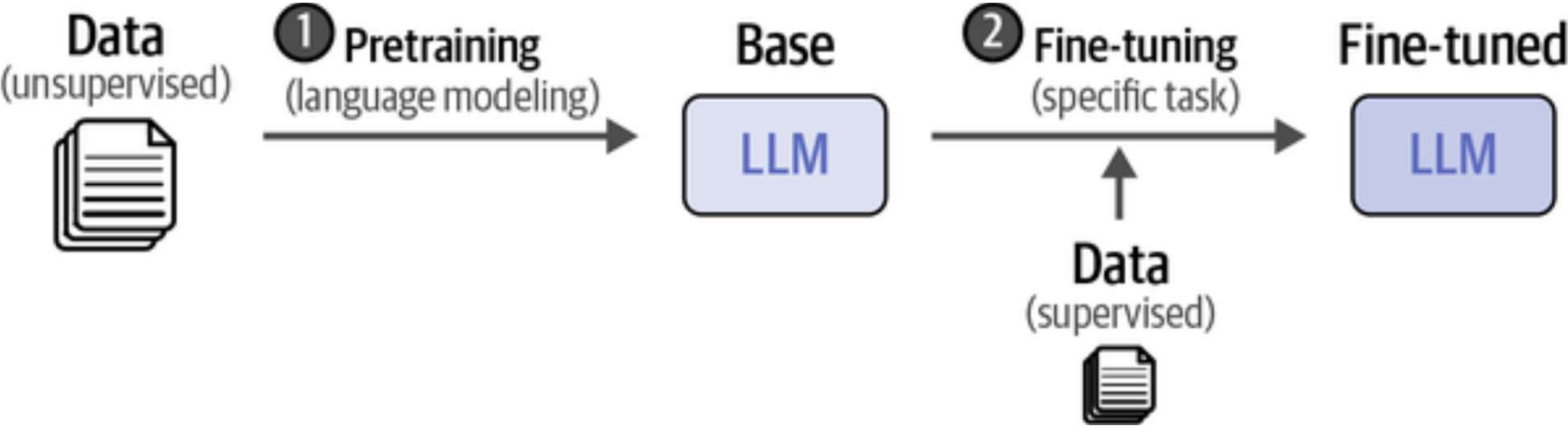


Open models

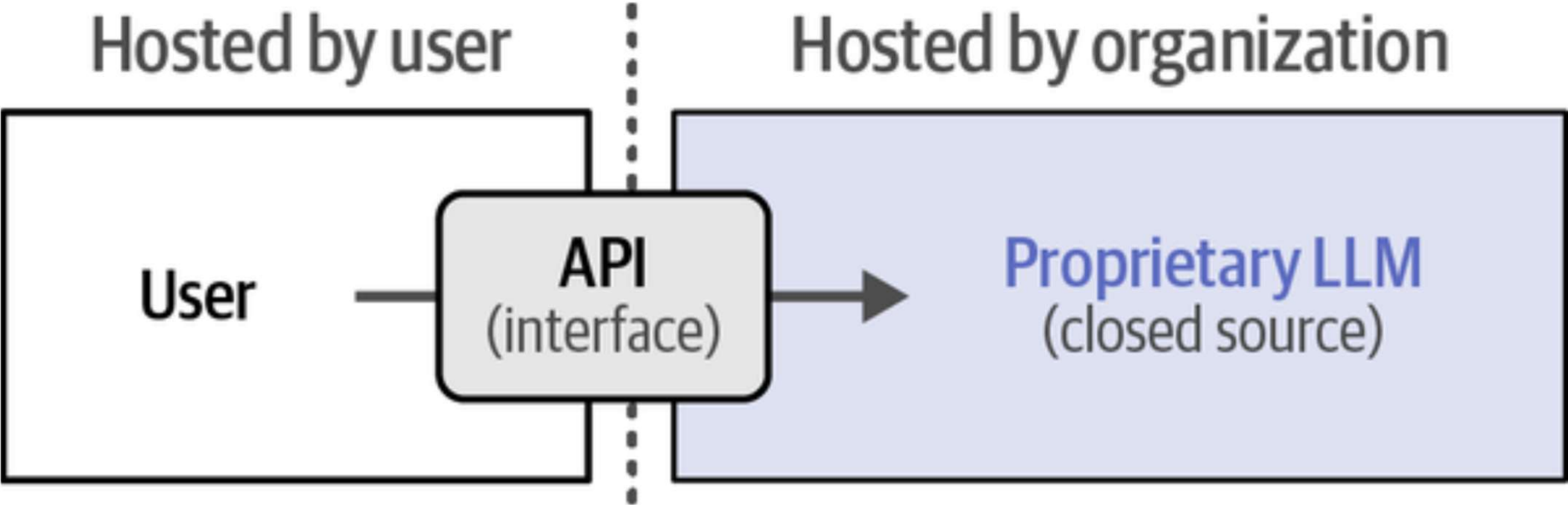
Traditional machine learning



LLM training



Closed source LLMs



Open source LLMs

