

Product Hour Live – Behind the Scenes of ReMark

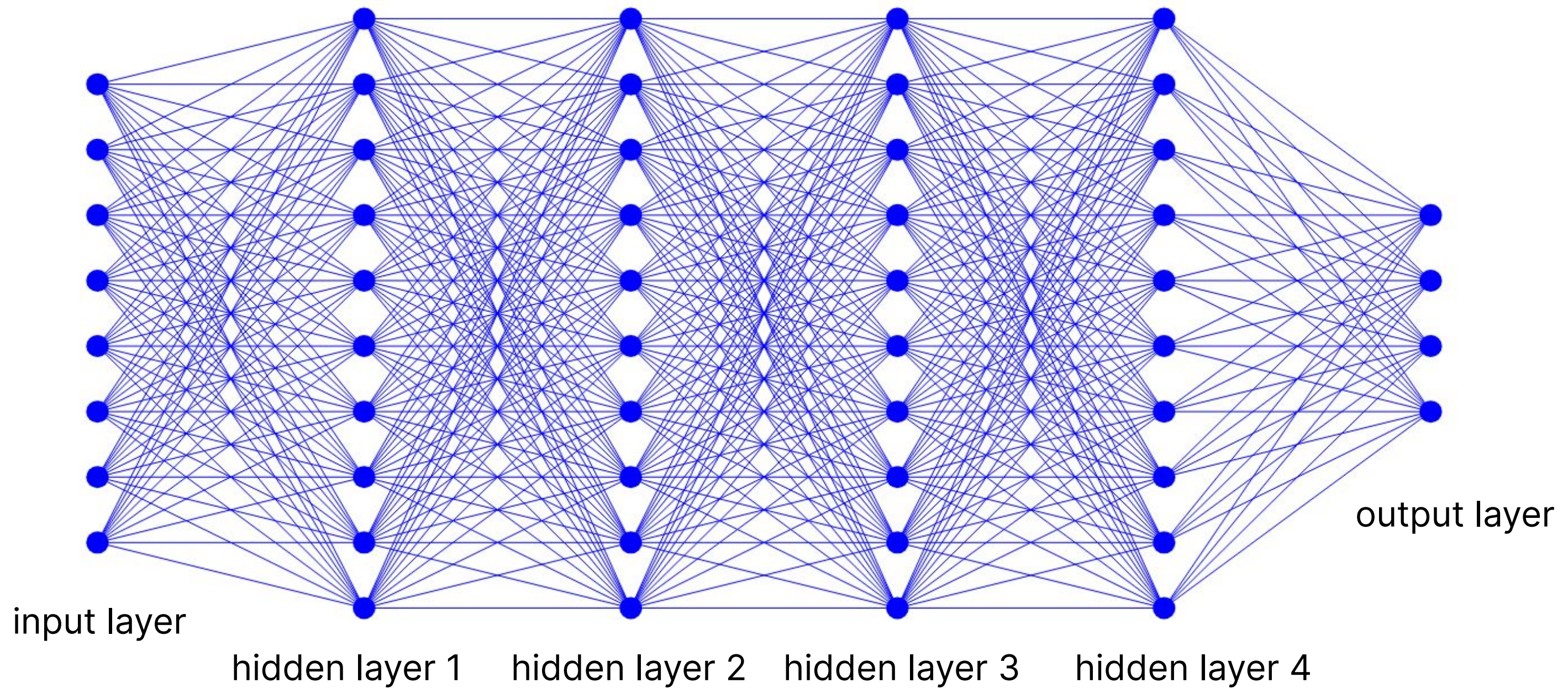


Mikko Korpela

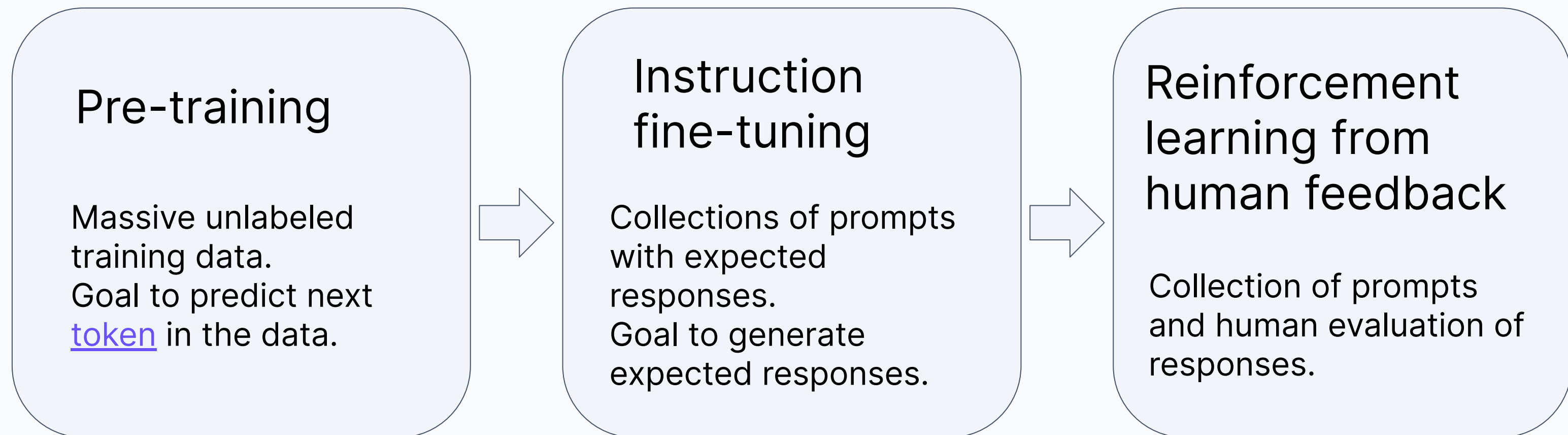
Large Language Model: an Instrument of Language



Large Language Model is a deep neural network



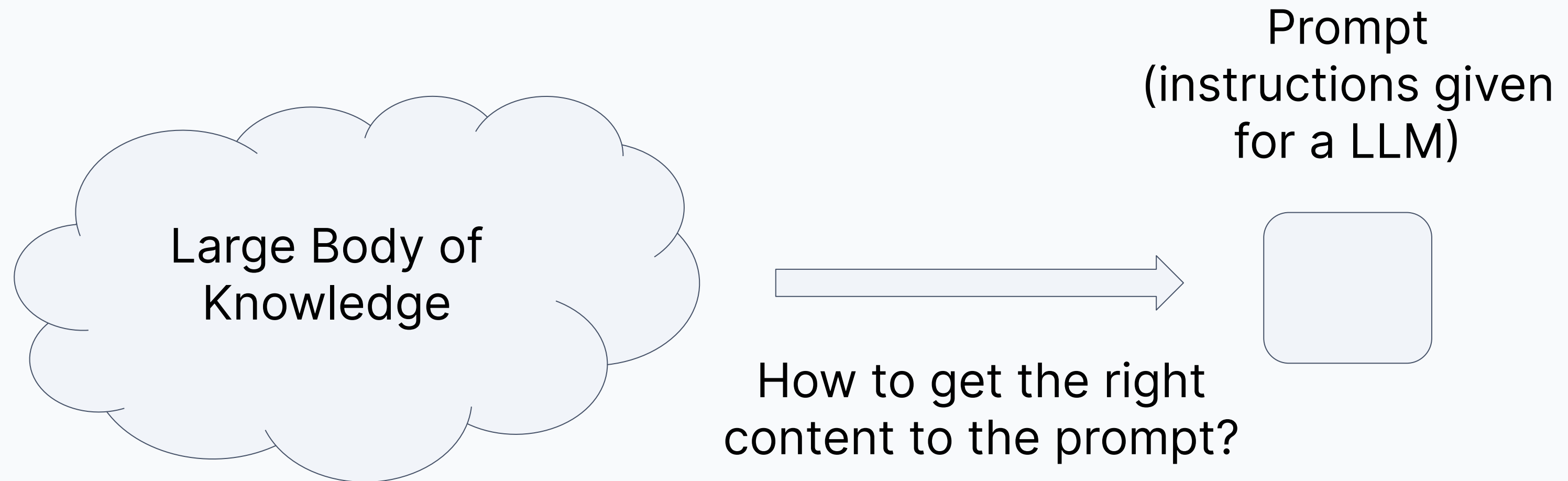
Training: The network weights are adjusted to make its predictions better



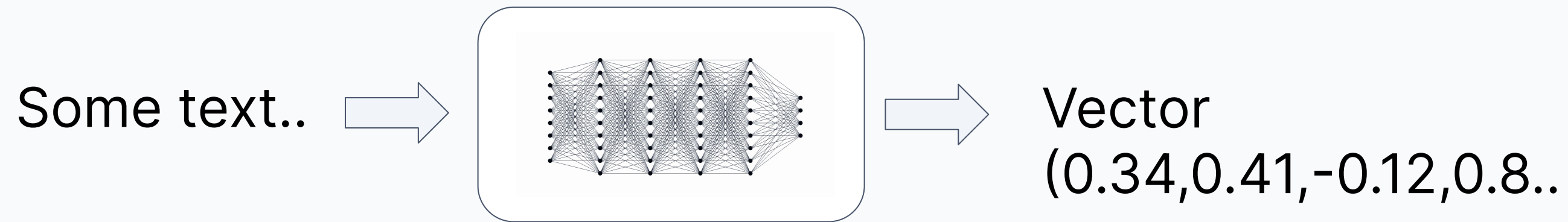
Context Window size:
the number of input and output tokens for a given model.

Model	Context Size (tokens)	Approx pages of text
gpt-4-32k	32768	50
gpt-4	8192	12
gpt-3.5-turbo-16k	16385	25
gpt-3.5-turbo	4097	6

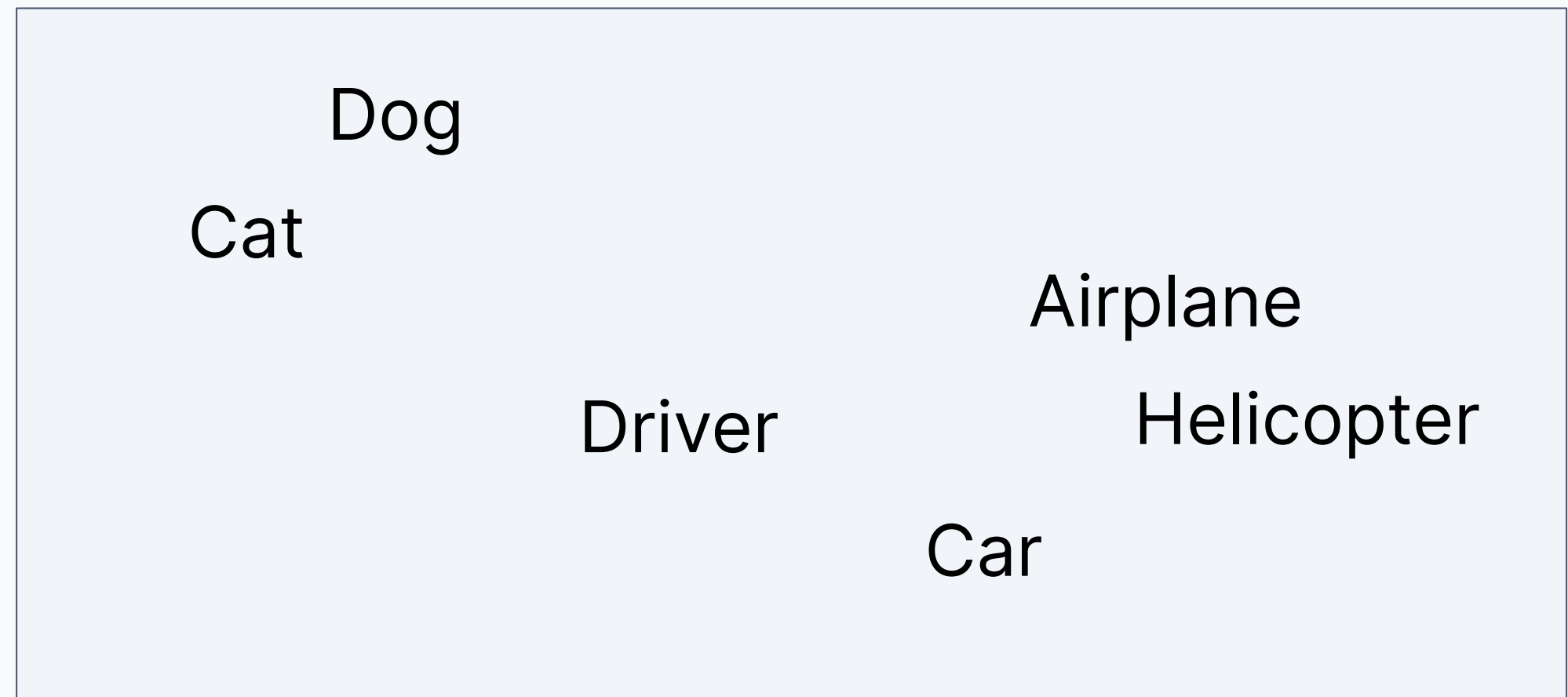
In-Context Learning



Embeddings: Turning words into numbers



Similarity search



Retrieval-augmented generation

