

Anatomía dunha Mente Artificial

Unha guía visual dos conceptos clave que impulsan os Modelos de Linguaxe (LLMs).

Parte I: Os Fundamentos

Como unha máquina aprende a ler a nosa linguaxe, peza por peza.



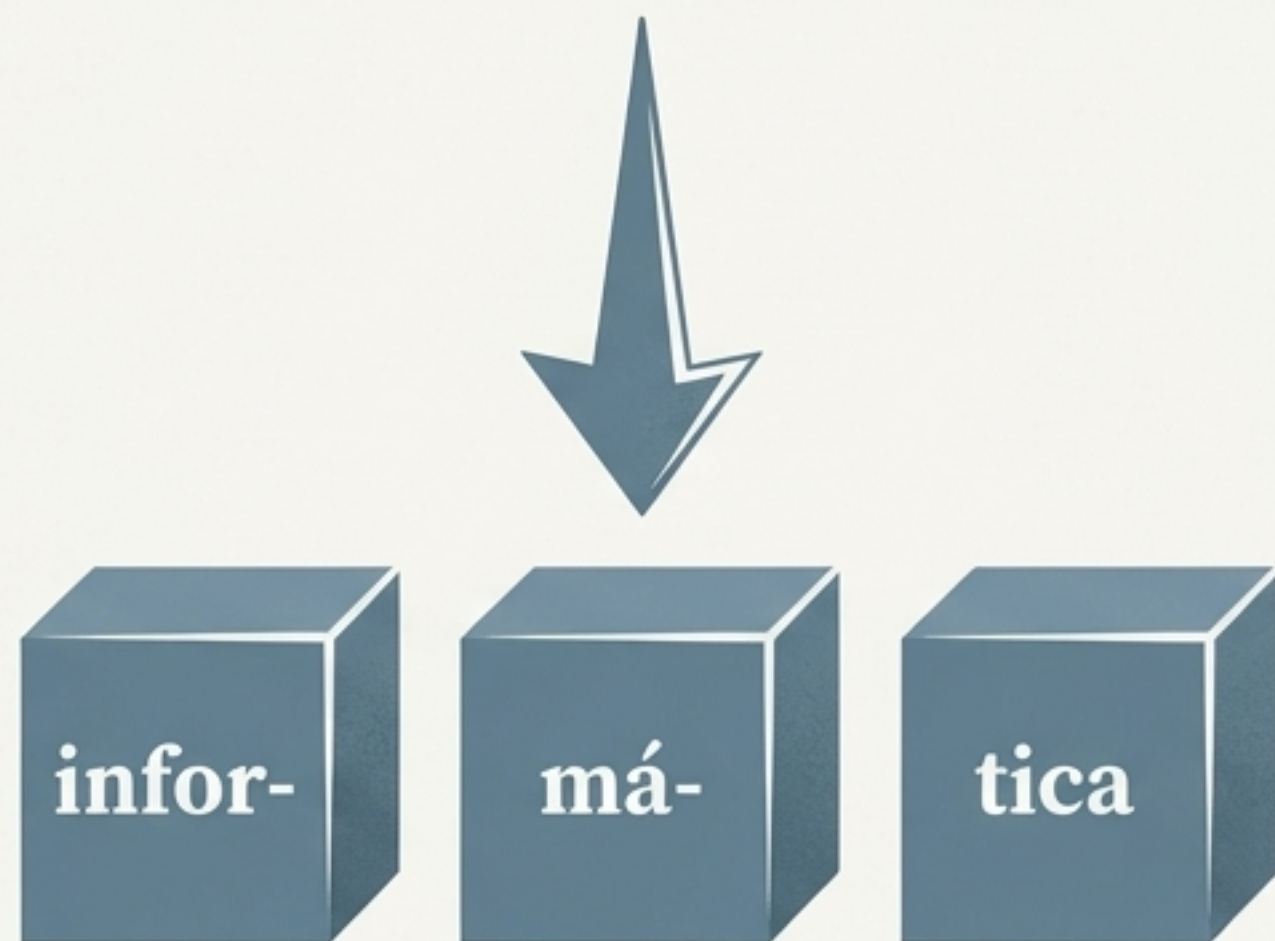
Tokens: As pezas elementais da linguaxe

Os modelos non len palabras, senón “tokens”, as unidades mínimas de significado coas que operan.

Son como as pezas máis pequenas do texto que o modelo entende. Cada palabra (ou parte dunha palabra) divídese en unidades chamadas tokens.

Cando un modelo le ou escribe texto, realmente está traballando con estes tokens —por exemplo, a palabra “informática” pode dividirse en dous ou tres tokens, segundo o modelo.

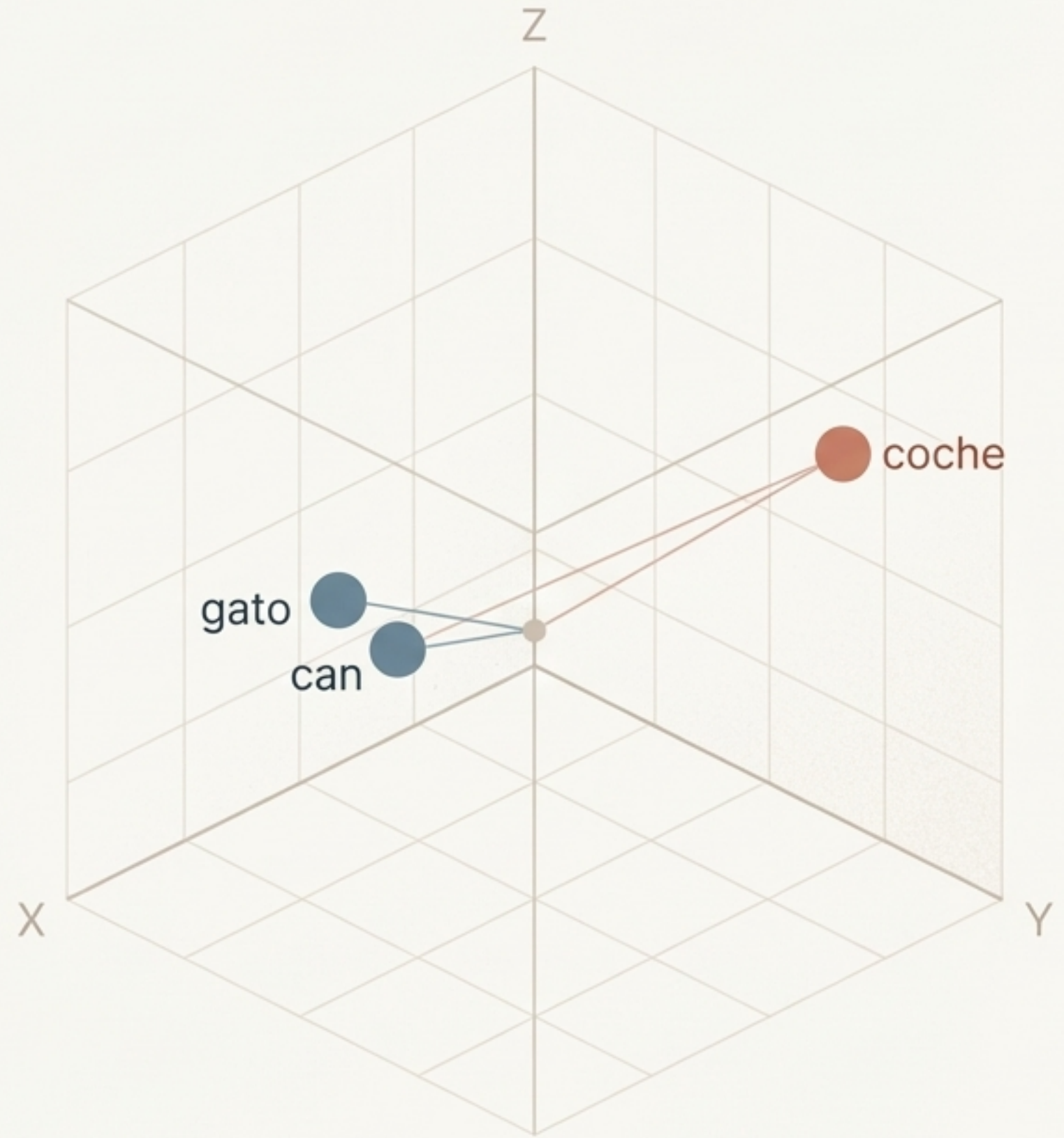
informática



Embeddings: O mapa do significado

Cada token convértese nunha coordenada numérica, situando as palabras con significados similares preto unhas das outras.

Cada token convértese nun número ou conxunto de números que describe o seu significado. É como se o modelo puidese colocar cada palabra nun mapa enorme do significado, onde as palabras parecidas quedan preto. Así, “gato” e “can” están máis próximas ca “coche”.



Parte II: O Motor Principal

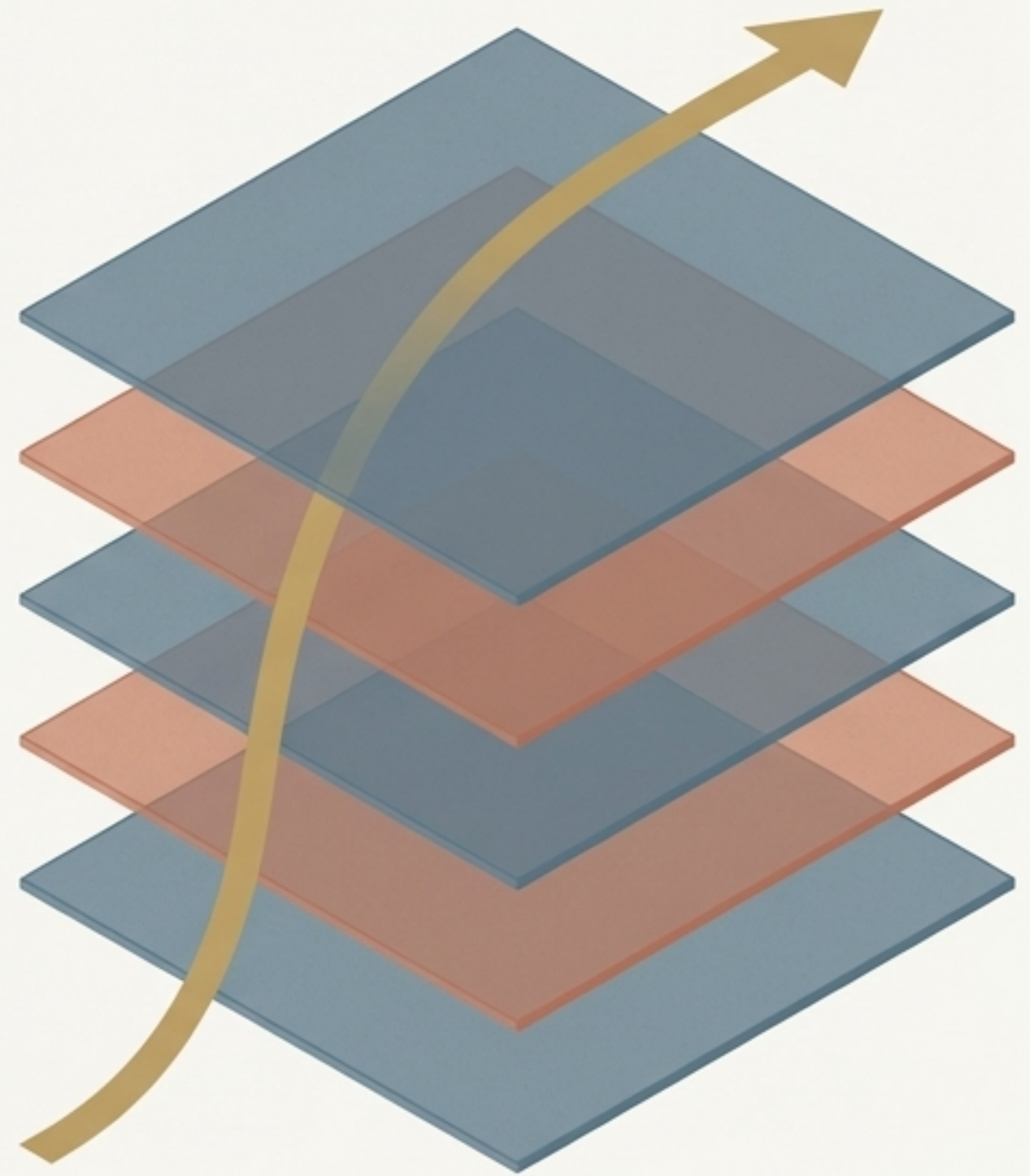
Dentro da arquitectura que procesa a información e xera intelixencia.



Transformer: A arquitetura que organiza o pensamento

É a estrutura de capas que permite ao modelo procesar texto complexo e manter a coherencia a longo prazo.

É o nome da estrutura interna que organiza todo o proceso dos LLMs. Funciona como un sistema de capas que procesan o texto e aprenden relacións complexas entre palabras. É o que permite que o modelo comprenda frases longas e manteña o fío do que se está dicindo.



Atención: O foco selectivo da intelixencia

Este mecanismo permite ao modelo ponderar a importancia de cada palabra para comprender o seu verdadeiro significado nunha frase.

Dentro do Transformer hai un mecanismo chamado atención, que lle permite ao modelo "mirar" todas as palabras ao mesmo tempo e decidir cales son importantes nun momento dado. Por exemplo, se dis "O gato come o peixe porque ten fame", a atención axúdalle a saber que `ten fame` refírese a `o gato` e non a `o peixe`.

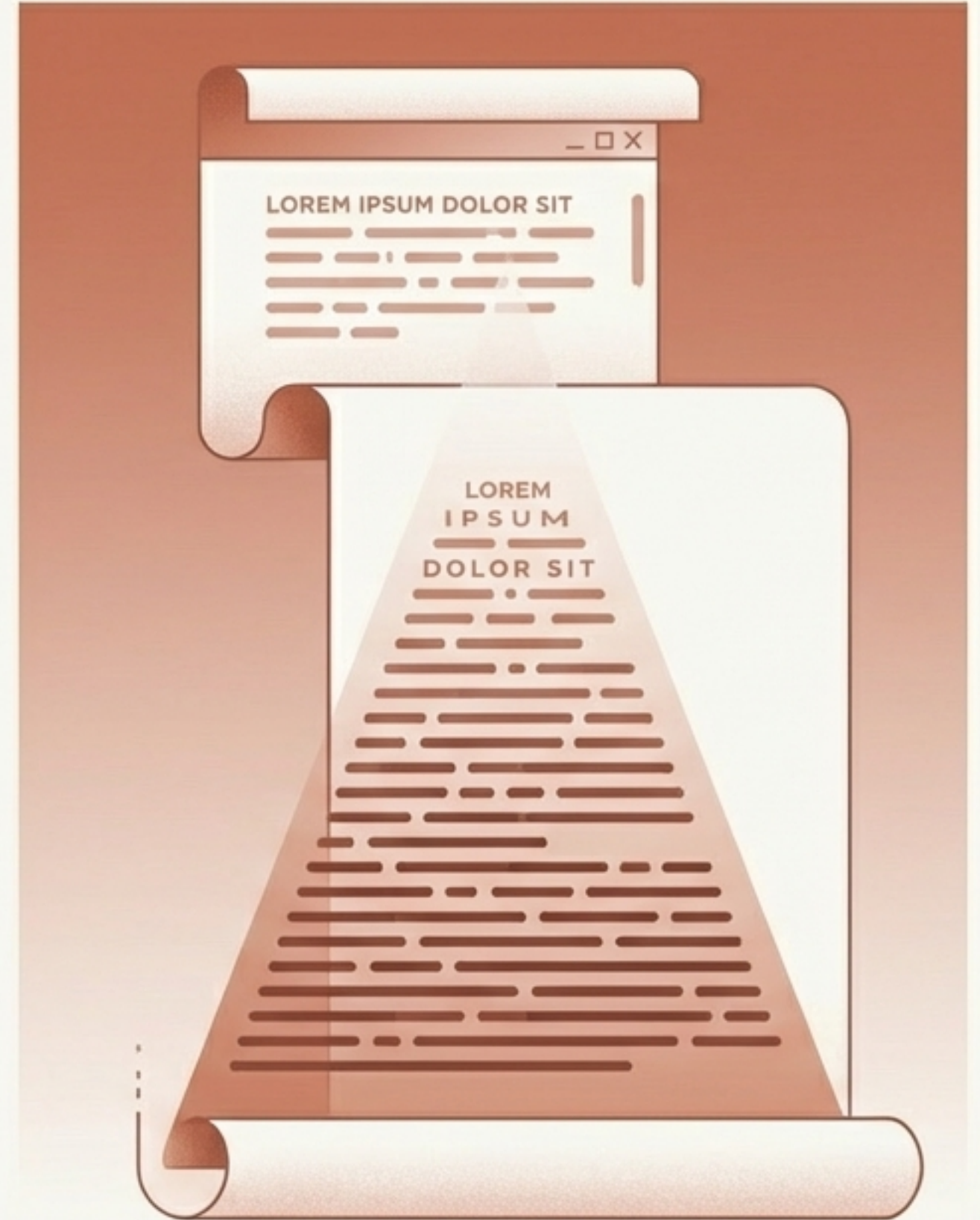


O gato come o peixe porque ten fame

Contexto: A memoria operativa do modelo

É o conxunto de información que o modelo utiliza para manter o fío e xerar respostas coherentes e precisas.

O contexto é o conxunto de palabras (tokens) e conexións entre elas que o modelo usa para entender o que se está a dicir ou o que se lle pregunta. Canto máis contexto poida lembrar, mellor serán a coherencia e a precisión da súa resposta.

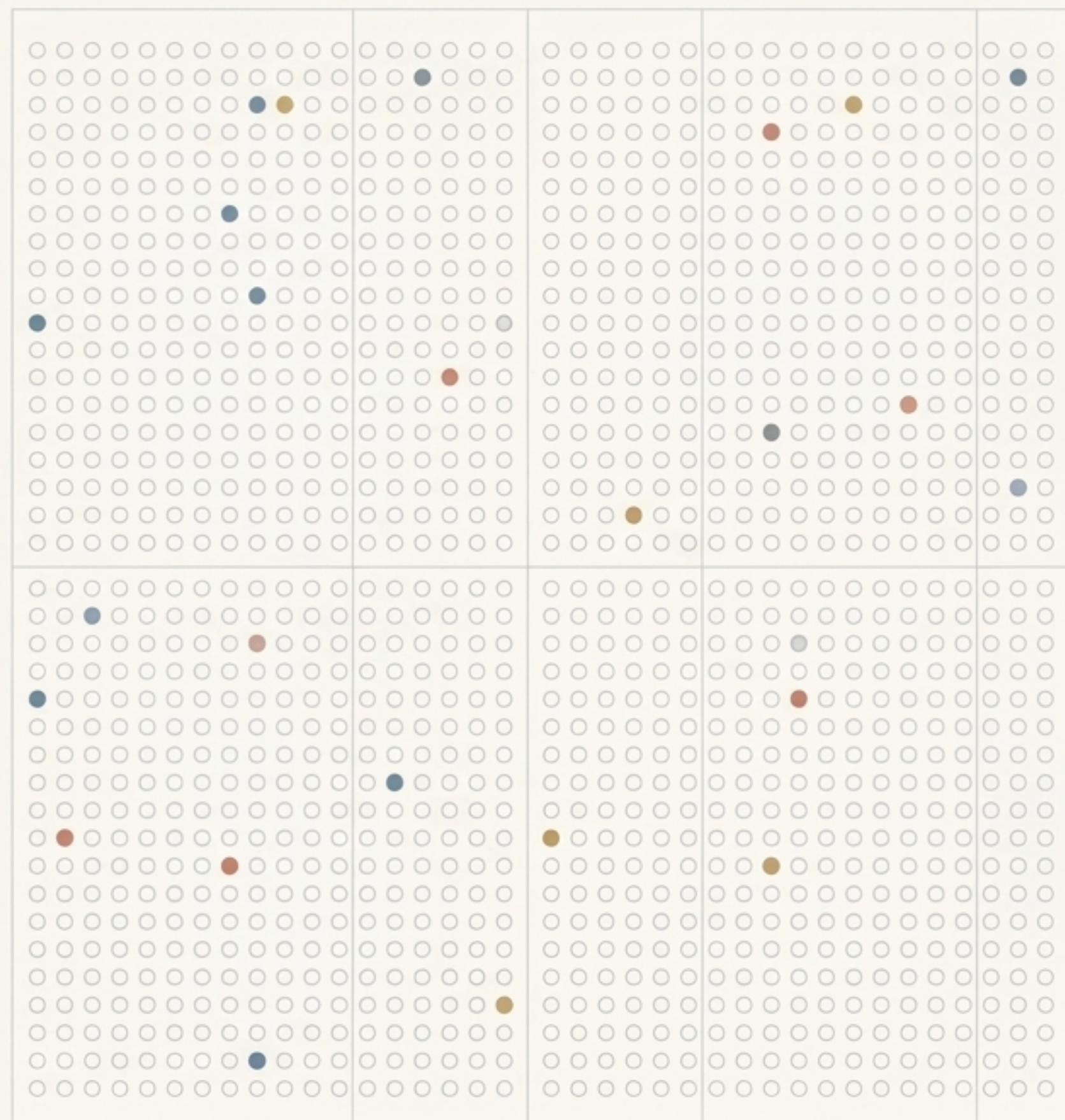


Parámetros: O coñecemento acumulado

Son millóns de 'axustes' internos aprendidos durante o adestramento que definen como o modelo responde e se comporta.

Son os valores internos que o modelo foi aprendendo durante o seu adestramento. Indican como debe combinar os datos que recibe para xerar un texto con sentido.

Pódense imaxinar como "axustes" que determinan o comportamento do modelo, da mesma maneira que os botóns de volume ou brillo axustan un aparello.



Parte III: A Conexión Humana

O noso papel á hora de guiar, instruír e refinar estes sistemas.




Prompt: O noso punto de partida

A instrución que lle damos ao modelo é a semente a partir da cal xera a súa resposta.

O prompt é a mensaxe ou instrución que lle damos ao modelo para que xere unha resposta. Pode ser unha frase curta (“Explícame que é a choiva ácida”) ou unha orde máis complexa (“Resume este texto e dálle un ton formal”). O modelo usa o prompt como punto de partida para crear o texto.

Explícame que é a choiva ácida|



A choiva ácida é unha forma de precipitación que é inusualmente ácida, o que significa que ten niveis elevados de ións de hidróxeno. A choiva ácida é causada principalmente pola liberación de dióxido de xofre e óxidos de nitróxeno á atmosfera, que reaccionan coa auga, o osíxeno e outras substancias químicas para formar ácido sulfúrico e nítrico. Estas substancias poden ser liberadas por fontes naturais, pero a principal causa son as emisións industriais e a queima de combustibles fósiles. Cando esta auga ácida cae ao chan, pode danar bosques, solos, ecosistemas acuáticos e estruturas feitas polo ser humano.

Fine-tuning: O proceso de refinamento

É o axuste fino, a miúdo con supervisión humana (RLHF), para que o modelo sexa máis útil, correcto e seguro.

O fine-tuning, tamén chamado axuste fino ou aliñamento, é o proceso polo que o modelo se adapta para ser máis útil, correcto e seguro. Unha parte importante disto é o RLHF (aprendizaxe con retroalimentación humana), onde persoas revisan as respostas do modelo e axúdanos a mellorar o ton, o estilo e a precisión.



Parte IV: Unha Perspectiva Crítica

Comprendendo as limitacións
inherentes da tecnoloxía.



Alucinacións: Cando o modelo inventa

Ocorren porque o obxectivo do modelo é xerar texto plausible segundo os patróns aprendidos, non verificar a súa veracidade.

As alucinacións son os erros que comete o modelo cando inventa información que non é certa. Ocorren porque o modelo non busca en bases de datos reais: simplemente xera texto segundo o que aprendeu. Por iso pode producir frases que soan realistas aínda que non sexan verdadeiras.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore magna aliquam. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore magna aliquam. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

De Tokens a Respostas: Unha Arquitectura de Probabilidades

Un LLM non 'entende' coma un humano; é un sistema sofisticado que ensambla a linguaxe (Tokens) seguindo patróns aprendidos (Parámetros) para xerar a resposta máis probable, non necesariamente a máis verdadeira.

