Apuntes

IA xenerativa e protección de datos

Modelos de Linguaxe Extensa (LLM)

Os **Modelos de Linguaxe Extensa** (LLM polas súas siglas en inglés, *Large Language Models*) son sistemas de intelixencia artificial que traballan co texto. A súa función principal é **predicir cal é a palabra que vai vir a continuación** nunha secuencia de palabras. Pódese pensar neles como "adiviños de palabras" moi avanzados, capaces de continuar frases e xerar textos completos a partir de instrucións ou preguntas que reciben.

A capacidade que teñen estes modelos para predicir depende directamente dos **datos cos que foron adestrados**. Canto máis variados e numerosos sexan os textos que aprenderon, mellor poden xerar respostas coherentes e útiles.

Existen dous grandes enfoques á hora de construír sistemas de intelixencia artificial:

- Por unha banda, están os **modelos baseados en regras**, que utilizan a lóxica e a gramática para decidir cal pode ser a seguinte palabra dunha frase.

- Por outra banda, están os **modelos baseados en datos**, que aprenden a partir de enormes coleccións de exemplos reais (libros, artigos, páxinas web) e, grazas a iso, son capaces de recoñecer patróns e xerar novos textos.

Os LLM actuais pertencen a este segundo grupo, o dos modelos baseados en datos.

As orixes e os primeiros modelos

Nos inicios da intelixencia artificial, os sistemas estaban inspirados no enfoque baseado en regras. Un dos primeiros programas coñecidos foi **ELIZA**, creado entre 1964 e 1966 por Joseph Weizenbaum no MIT. Este programa era capaz de manter unha especie de conversa simple con persoas, recoñecendo palabras clave e respondendo con frases predefinidas segundo patróns. Aínda que non comprendía realmente o que lle dicían, moitos usuarios tiñan a sensación de estar interactuando cunha máquina intelixente.

Un pouco antes, en 1950, o matemático británico **Alan Turing** formulara o que se coñece como o **Test de Turing**. Esta proba propoñía unha maneira de avaliar se unha máquina podería ser considerada intelixente: unha persoa debía facer preguntas sen saber se respondía outro ser humano ou unha máquina. Se non era capaz de distinguir cal era a máquina, entón esta podería considerarse "intelixente".

Co paso do tempo, o enfoque baseado en regras demostrou ser limitado, porque non permitía xerar textos naturais nin adaptarse a contextos máis complexos. Foi necesario esperar ata dispoñer de grandes cantidades de información dixital e de ordenadores máis potentes para dar o salto ao enfoque **baseado en datos**, no que se fundamentan os LLM modernos.

O nacemento dos LLM modernos

A partir do ano 2010 producíronse dous cambios fundamentais que fixeron posible a aparición dos LLM actuais:

1. **O fenómeno do Big Data.** Grandes compañías como Google, Amazon, Microsoft ou Facebook comezaron a acumular enormes cantidades de datos xerados polos usuarios. Esa información converteuse nunha fonte esencial para adestrar modelos capaces de recoñecer patróns no texto.

2. **O avance das GPU.** As tarxetas gráficas, coñecidas como **GPU**, deixaron de usarse só para videoxogos e convertéronse en ferramentas fundamentais para a intelixencia artificial. Grazas á súa potencia, permitiron procesar grandes cantidades de datos a gran velocidade e adestrar modelos moito máis complexos.

No ano 2017 apareceu a **arquitectura Transformer**, que supuxo unha auténtica revolución. Este tipo de modelo incluía un mecanismo de **atención**, capaz de analizar todo o texto ao mesmo tempo e decidir cales son as palabras máis relevantes para comprender o seu significado. Ademais, esta arquitectura permite adestrar os modelos utilizando moitas GPU en paralelo, o que acelera o proceso e fai posible construír sistemas cada vez máis grandes e potentes.

Precisamente as siglas de Chat**GPT** significan *Generative Pre-trained Transformer*. (Transformer xenerativo pre-adestrado para conversar)

Desde entón, a cantidade de datos utilizada para adestrar estes modelos e o poder de cálculo empregado aumentou de forma extraordinaria. Isto explica por que os LLM actuais poden responder con tanta coherencia e xerar textos tan parecidos aos que escribiría unha persoa.

Da arquitectura ao modelo: por que hai tantos LLM diferentes?

Xa vimos que a arquitectura **Transformer** (2017) é a base técnica común á meirande parte dos LLM actuais: ChatGPT, Claude, Gemini, Mistral, DeepSeek e moitos outros compárteno. Se todos parten do mesmo deseño, por que se comportan de xeito tan distinto entre si?

A arquitectura define **como** procesa o texto un modelo —o mecanismo de atención, como se organizan as capas internas—. Pero o resultado final depende doutros factores, independentes da arquitectura, que cada empresa decide á súa maneira:

- **Os datos de adestramento.** Con que textos, en que idiomas e de que calidade se adestrou o modelo. Un modelo adestrado maioritariamente con textos en inglés responderá mellor en inglés que en galego, por exemplo.
- **O tamaño do modelo.** Cantos **parámetros** ten —as conexións internas que se axustan durante o adestramento—. Un modelo maior adoita recoñecer patróns máis complexos, aínda que máis grande non sempre é mellor: tamén importa moito como se adestrou.

- **A potencia de cálculo empregada no adestramento.** Cantas GPU e canto tempo se usaron. Isto inflúe directamente no custo económico de crear o modelo.
- **O axuste fino e o aliñamento posterior.** Despois do adestramento inicial cun gran volume de texto xenérico, cada empresa "educa" ao seu modelo á súa maneira para que responda de forma útil e segura. Anthropic, por exemplo, usa un método chamado **Constitutional AI**, no que o propio modelo revisa as súas respostas comparándoas cunha lista de principios escritos. Outras empresas apóianse máis en persoas que valoran e corrixen respostas unha a unha. **É sobre todo este proceso —non a arquitectura— o que explica por que un modelo se nega a facer algo que outro sí fai, ou por que un responde de forma máis breve e outro máis conversacional.**
- **O obxectivo e a especialización.** Uns modelos buscan ser xeneralistas, outros priorizan a programación, outros a eficiencia de custo, outros a soberanía lingüística ou tecnolóxica dun país ou rexión.
- **Se os seus "pesos" son abertos ou pechados.** Un modelo de **pesos abertos** pódese descargar e executar en servidores propios; un de **pesos pechados** só se pode usar conectándose aos servidores da empresa que o creou.

Comparativa de familias de modelos (curso 2026-2027)

Modelo	Empresa / orixe	Pesos	Que o caracteriza
ChatGPT (GPT)	OpenAI, EE.UU.	Pechados	O primeiro en popularizarse a finais de 2022; moi estendido en uso xeral, con amplo ecosistema de complementos
Claude	Anthropic, EE.UU.	Pechados	Foco en seguridade e fiabilidade; adestrado con Constitutional AI; moi usado en programación e tarefas de longa duración
Gemini	Google DeepMind, EE.UU./RU	Pechados	Multimodal de fábrica (texto, imaxe, vídeo e audio á vez); moi integrado no ecosistema Google (Android, Workspace, buscador)
Mistral	Mistral AI, Francia	Algúns abertos	Empresa europea; aposta por modelos descargables que non dependen sempre dun servidor externo; alternativa europea ás grandes tecnolóxicas dos EE.UU.

DeepSeek	DeepSeek, China	Abertos	Chamou a atención mundial a comezos de 2025 por acadar un rendemento comparable ao dos grandes modelos occidentais cun custo de adestramento moito menor; pesos abertos e prezo de uso moi baixo
ALIA	Goberno de España + BSC	Abertos	Primeiro modelo público pensado especificamente para o castelán e as linguas cooficiais —entre elas **o galego**—; adestrado no supercomputador MareNostrum 5; busca soberanía tecnolóxica, non competir comercialmente coas grandes empresas

Dime se queres que axuste algo máis —por exemplo, podería engadir unha liña curta que explique por que DeepSeek é un caso especialmente relevante para falar de custo e eficiencia (foi noticia porque puxo en dúbida que faga falta gastar centos de millóns para adestrar un modelo punteiro), se cadra útil para conectar con contidos económicos ou de impacto da IA que xa tratades noutros bloques.

A tokenización

Para que un modelo de linguaxe poida traballar cun texto, este debe ser dividido en **tokens**. Un **token** é un fragmento pequeno en que se descompón o texto: pode ser unha palabra enteira, unha sílaba, unha parte dunha palabra ou mesmo un signo de puntuación.

Por exemplo, a frase “as vacas comen herba” podería dividirse nos seguintes tokens: “as”, “vac”, “as”, “com”, “en”, “herba”.

O conxunto de tokens que un modelo pode manexar ao mesmo tempo chámase **contexto**. Este contexto inclúe tanto o texto da **pregunta** que fai o usuario como a **resposta** que xera a máquina. Canto máis grande sexa o contexto, máis información pode recordar a IA e máis longos poden ser os textos cos que traballa.

Segundo o modelo, o límite de tokens pode variar moito. Por exemplo, GPT-3.5 pode xestionar arredor de 4.000 tokens, mentres que GPT-4 chega ata 32.000. Os modelos máis recentes, como GPT-5, poden alcanzar os 128.000 tokens ou mesmo máis nas versións empresariais.

Os prompts

Para falar cun LLM é necesario escribir un **prompt**. Un prompt é simplemente a instrución ou a pregunta que lle facemos ao modelo para que realice unha tarefa.

Antes de redactar un **prompt**, é importante **escoller ben o modelo de adestramento máis axeitado á tarefa**. Non todos os modelos están pensados para facer o mesmo: algúns son máis rápidos e útiles para tarefas sinxelas, como redactar un texto ou facer un resumo, mentres que outros están deseñados para investigacións máis complexas, nas que se necesita un razoamento profundo e máis capacidade de memoria. Escoller o modelo correcto axuda a que a resposta sexa máis adecuada ao que realmente precisamos.

A calidade dun prompt inflúe directamente na utilidade da resposta. Por iso é importante aprender a estruturalo ben. Un bo prompt pode incluír os seguintes elementos:

- **O contexto:** consiste en explicarlle á IA a situación na que estamos ou a información de fondo que debe ter en conta. Por exemplo: “Estou en 4º da ESO e cústame entender os sistemas de ecuacións”.

- **O rol:** aquí dicímoslle á IA que papel debe asumir. Por exemplo: “Explícame como se foses o meu profesor particular de matemáticas, paciente e claro”.

- **A acción que necesito:** definimos con claridade o que queremos que faga. Por exemplo: “Explícame paso a paso como se resolve un sistema de dúas ecuacións con dúas incógnitas”.

- **A información que debe conter:** indicamos que elementos queremos ver na resposta. Por exemplo: “Pon exemplos con números pequenos, explica cada paso e propón un exercicio para practicar”.

- **O estilo ou formato:** especificamos como debe redactarse a resposta. Por exemplo: “Usa frases curtas e claras, vai paso a paso, dáme pistas se me equivoco e resume ao final os pasos máis importantes”.

Por exemplo:

Modelo de Prompt
Modelo de razoamento *Quero que uses o modelo **Instant**, porque necesito explicacións rápidas e claras, non análises moi profundas.* ***Contexto*** *Estou en 4º da ESO e cústame lembrar os verbos irregulares en inglés, sobre todo as formas de pasado e participio.* ***Rol*** *Ti es o meu profesor particular de inglés, paciente e con exemplos prácticos.* ***Acción que necesito que realice*** *Axúdame a repasar os verbos irregulares máis importantes en inglés.* ***Información que quero que conteña*** • *Unha lista cos 15 verbos irregulares máis usados.* • *As tres formas de cada verbo (infinitivo, pasado, participio).* • *A tradución ao galego.* • *Un par de frases de exemplo con algúns verbos.* ***Como quero que fagas a tarefa*** • *Explica con frases curtas e claras.* • *Organiza a lista en táboa para que sexa doado de ler.* • *Pon exemplos sinxelos de frases.* • *Ao final, dame un pequeno exercicio para practicar.*

A elección da IA:

Protección de datos

Cando usamos Internet, deixamos moitos datos persoais: o que buscamos, as páxinas que visitamos, as fotos que compartimos ou mesmo a nosa localización. Estes datos utilízanse para diferentes cousas, como mellorar servizos, recomendar contidos ou, moitas veces, para facer publicidade adaptada a cada usuario. Pero tamén teñen perigos: toda esta información forma a nosa pegada dixital, é dicir, o rastro que deixamos na rede e que pode permanecer durante moito tempo. Ademais, se cae en mans

equivocadas, os datos poden empregarse con fins delituosos, como roubos de identidade, estafas en liña ou accesos non autorizados ás nosas contas.

É importante non compartir información privada, como enderezos, números de teléfono ou contrasinais, e lembrar que todo o que subimos pode ser analizado polos sistemas. Tamén convén **empregar contas de usuario que non estean directamente ligadas ao noso perfil persoal**, para garantir un maior grao de anonimato cando utilizamos ferramentas de intelixencia artificial, **especialmente no caso dos menores de idade**, que deben ser máis coidadosos coa súa privacidade.

Enlaces

https://youtu.be/I9MLwYPpJAY?feature=shared

https://www.aprendemachinelearning.com/como-funcionan-los-transformers-espanol-nlp-gpt-bert/

Pestana 3