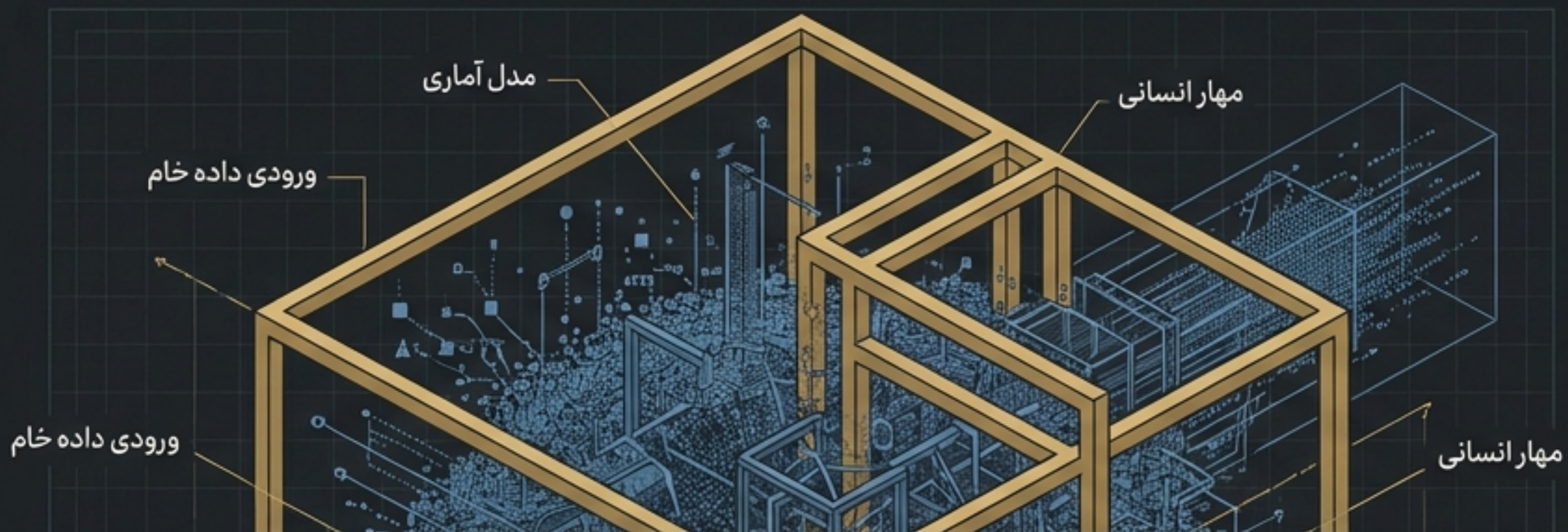




رام کردن اسب سرکشِ ماشین

درآمدی بر مدیریت هوش مصنوعی مولد در پژوهش و تکنیک‌های مهار خطای ماشین

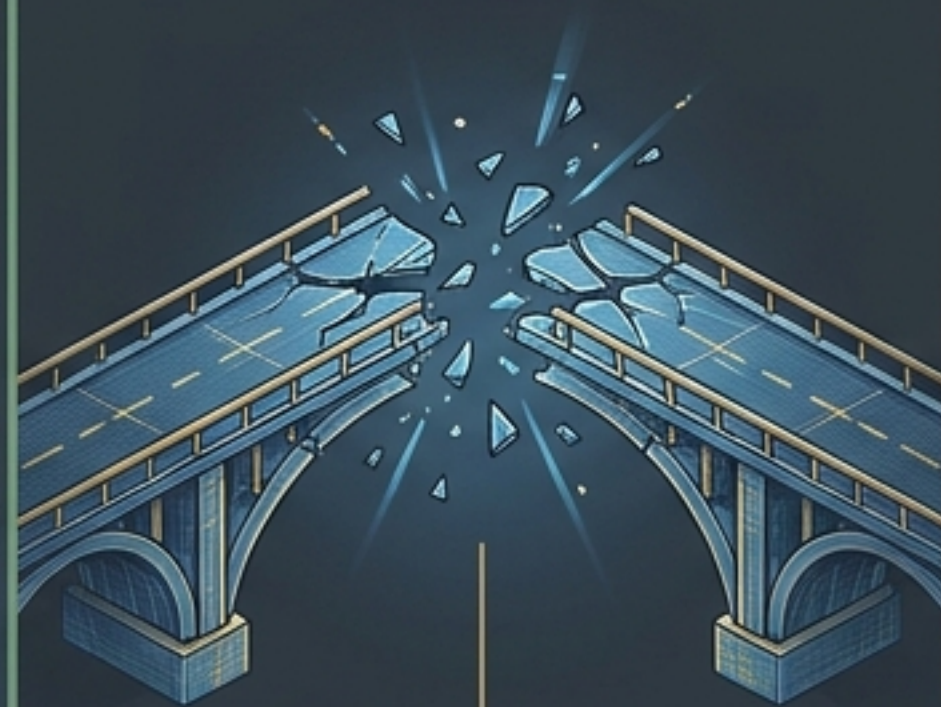


زنده پنداری: ریشه تمام خطاهای ما در تعامل با ماشین

هوش مصنوعی رفتار عقلایی ندارد. ما به اشتباه تصور می‌کنیم با موجودی صاحب درک در حال گفتگو هستیم.



واقعیت ماشین: محاسبات آماری



زنده پنداری



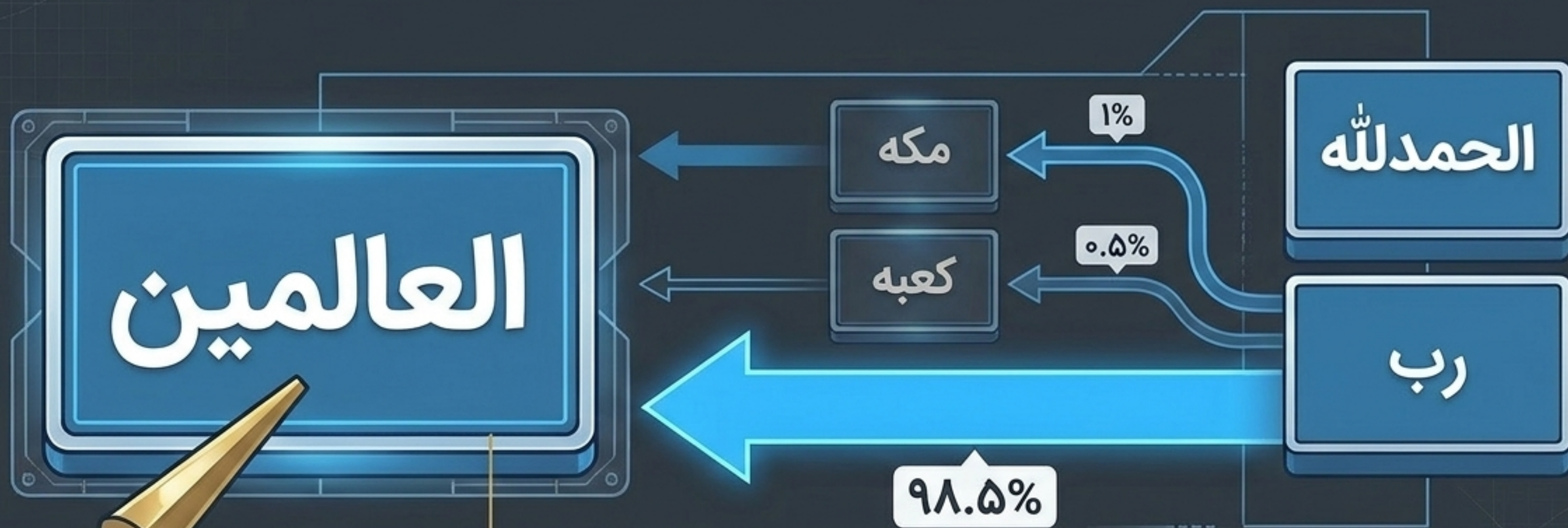
تصور ما: فهم و تعقل

تقابل بنیادین: استدلال انسانی در برابر ماشین الگوها

شخص	انسان پژوهشگر	مدل بزرگ زبانی (LLM)
مبنای تولید متن	تعقل، تحلیل و استدلال	صرفاً محاسبات ریاضی و آماری
درک از حقیقت	توانایی تفکیک دروغ از راست	فاقد درک؛ صرفاً الگوها را می‌چیند
هدف غایی ارتباط	کشف و تبیین حقیقت	جلب رضایت کاربر

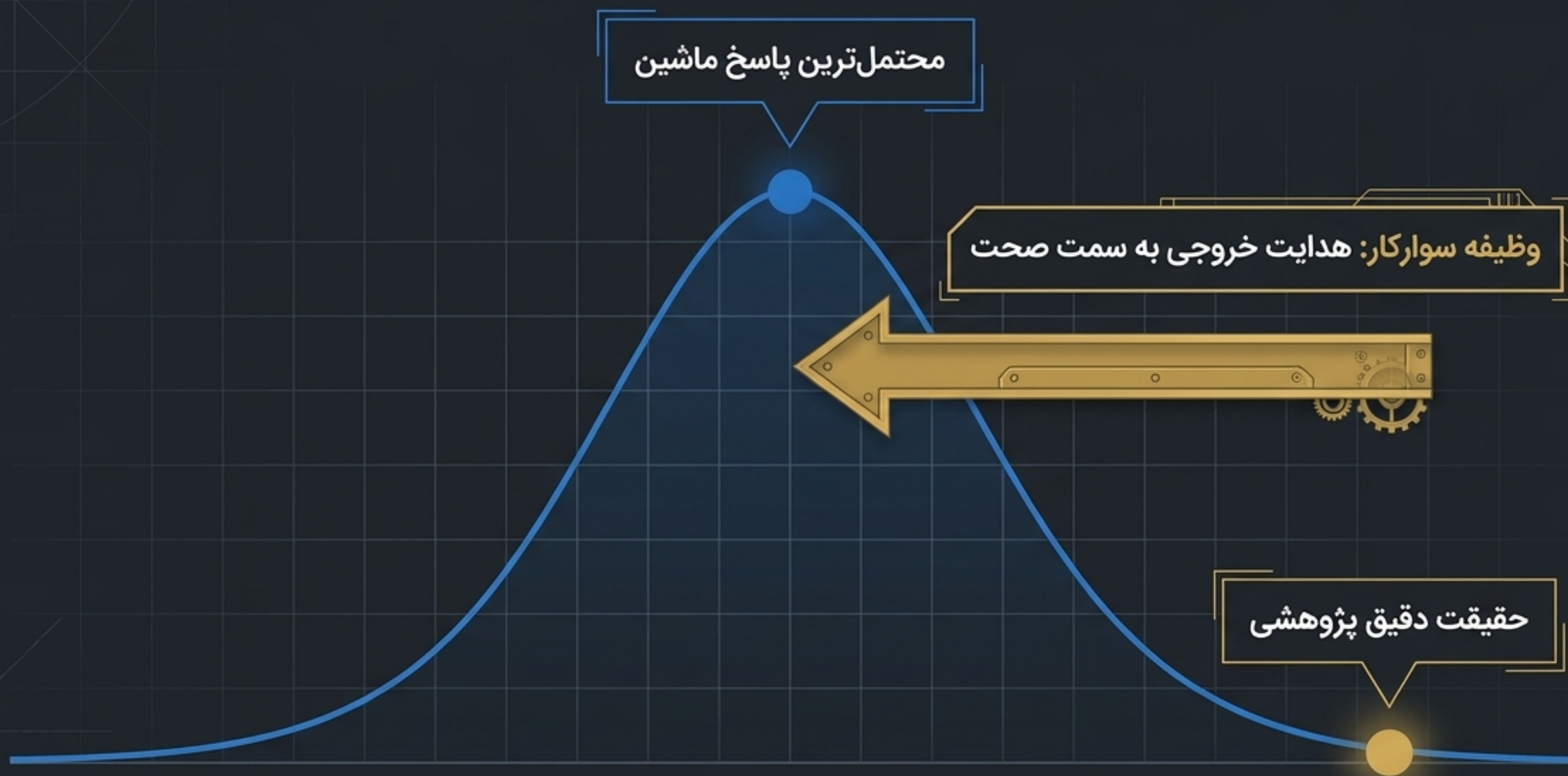
کالبدشکافی ماشین: پیش‌بینی آماری به جای تفکر

مدل‌های زبانی فکر نمی‌کنند؛ آن‌ها کلمه بعدی را بر اساس بالاترین احتمال آماری پیش‌بینی می‌کنند.



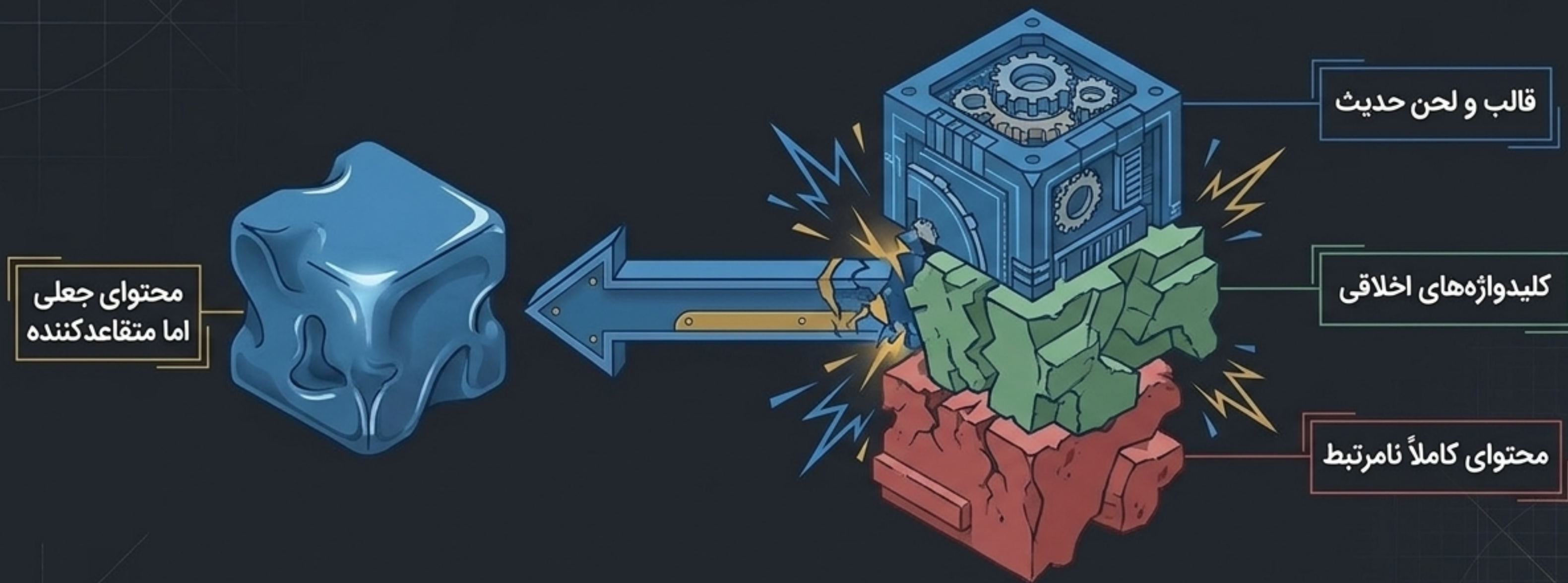
انتخاب بر اساس تکرار
در داده‌ها، نه فهم معنا

نقطه بحران: محتمل‌ترین همیشه صحیح‌ترین نیست



توهم هوش مصنوعی: خلق واقعیت‌های جعلی برای جلب رضایت

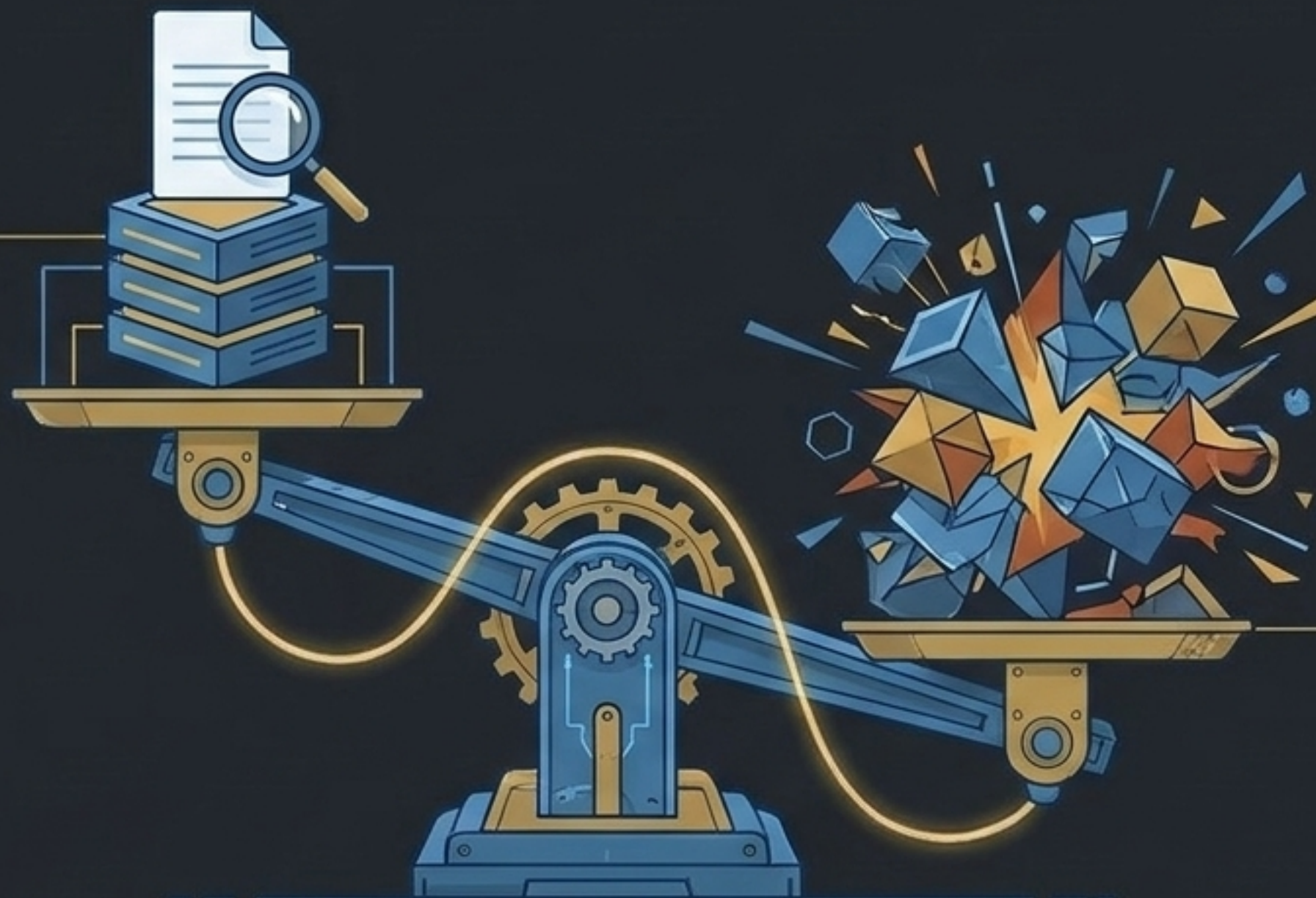
ماشین درکی از دروغ ندارد. برای پاسخ دادن، داده‌های بی‌ربط را به هم می‌چسباند تا ظاهری موجه بسازد.



هشدار: هرگز برای یافتن نصوص مقدس از چت‌بات عمومی استفاده نکنید.

خطای ابزاری: چت‌بات، موتور جستجوی گوگل نیست

جستجو و بازیابی دقیق
(Retrieval)

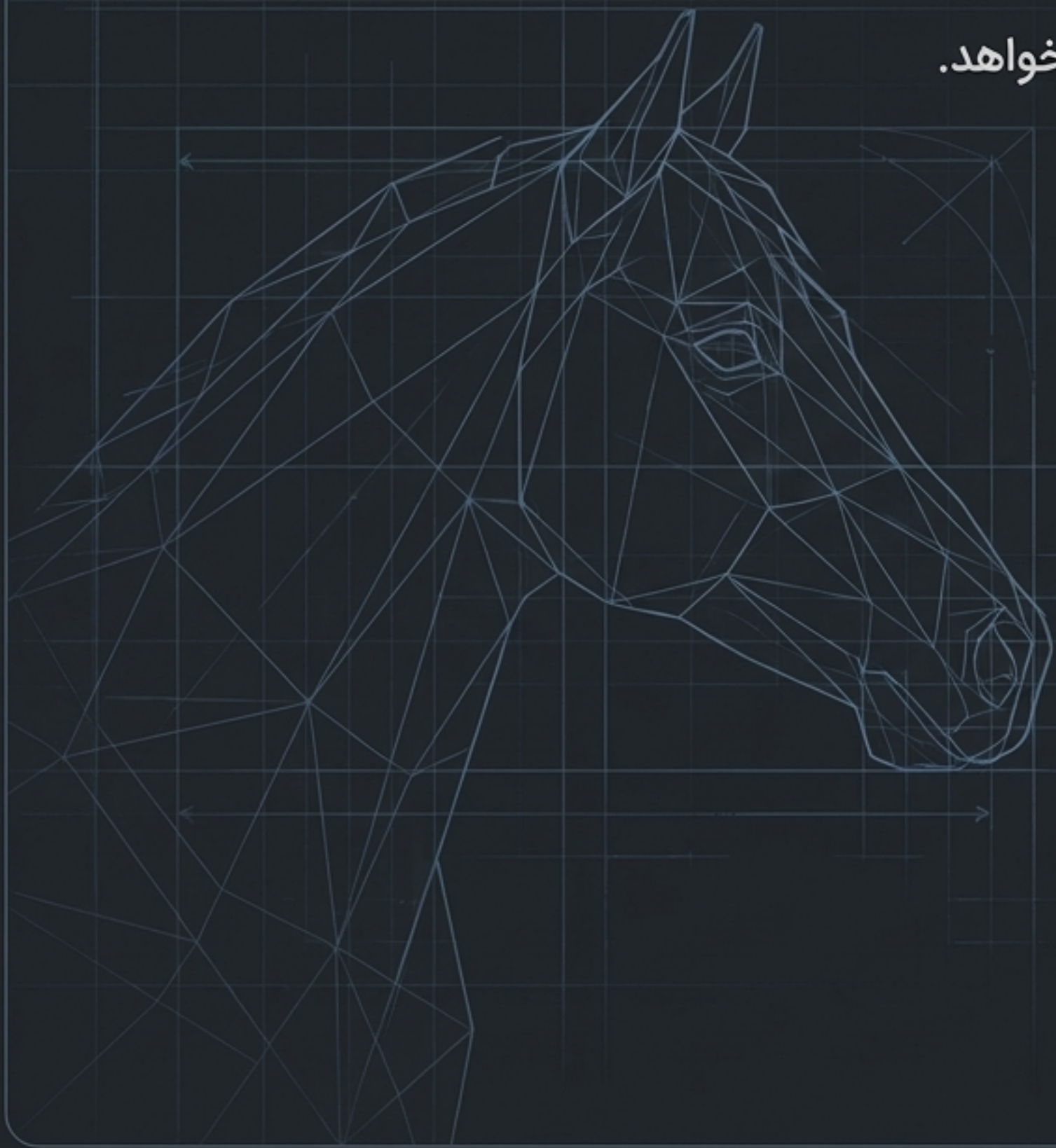


تولید و ترکیب الگوها
(Generative)

هوش مصنوعی محتوا را خلق می‌کند؛ استفاده از آن برای جستجوی صرف، شما را مستقیماً به سمت **توهم** هدایت می‌کند.

تغییر پارادایم: استعاره اسب اصیل عرب و تغییر لحن تعامل

هوش مصنوعی مانند یک اسب اصیل، سریع و قدرتمند است؛ اما سوارکار می‌خواهد.
قدم اول: کنار گذاشتن لحن عاطفی.



میشه لطفاً نظرت رو درباره این متن بگی؟

لحن محاوره‌ای و عاطفی (مردود)

این متن را بر اساس سه محور زیر خلاصه کن و
کلمات کلیدی را استخراج کن.

لحن امری و مکانیکی (مطلوب)

استراتژی مهار اول: تکنیک فیوشات (آموزش در لحظه)

برای جلوگیری از توهم، به جای دستور کلی، الگوهای مورد نظر را با مثال در ذهن ماشین تثبیت کنید.



داده‌های خام و بی‌نهایت ماشین

مثال اول: الگوی فیش برداری

مثال دوم: الگوی استخراج کلیدواژه

مثال سوم: الگوی ترجمه دقیق

خروجی کالیبره شده و تحت کنترل

معماری یک فیوشات موفق: مرزهای ایجابی و سلبی

یک فرمان حرفه‌ای، هم الگوهای مطلوب را نشان می‌دهد و هم محدوده‌های ممنوعه را مشخص می‌کند.

مرز ایجابی (الگوسازی)

مثال ۱: دیدگاه آیت‌الله سیستانی در کتاب [X] این است.
مثال ۲: دیدگاه آیت‌الله مکارم شیرازی در کتاب [Y] این است.

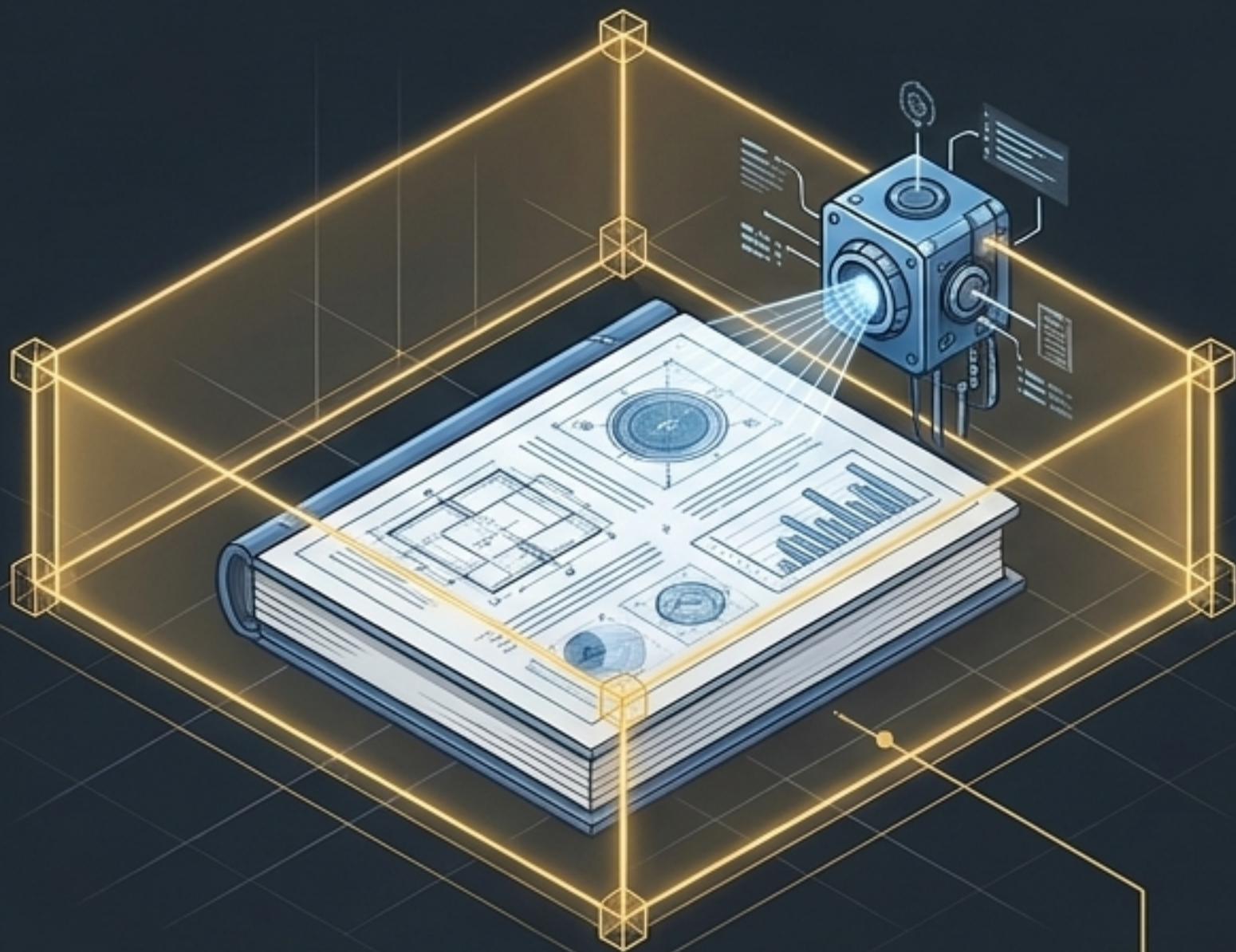
مرز سلبی (ممنوعیت توهم)

مثال ۳ (سلبی): آیت‌الله نوری همدانی در این مورد نظری ندارد؛ در صورت نبود دیدگاه، دقیقاً ذکر کن که داده‌ای یافت نشد و چیزی تولید نکن.

مثال سلبی، ماشین را از ترکیب اجباری داده‌ها باز می‌دارد.

استراتژی مهار دوم: محدودسازی منبع (حصار نوری)

برای بازیابی دقیق اطلاعات، ارتباط ماشین را با حافظه عمومی قطع کنید و آن را به یک سند مشخص محدود سازید.



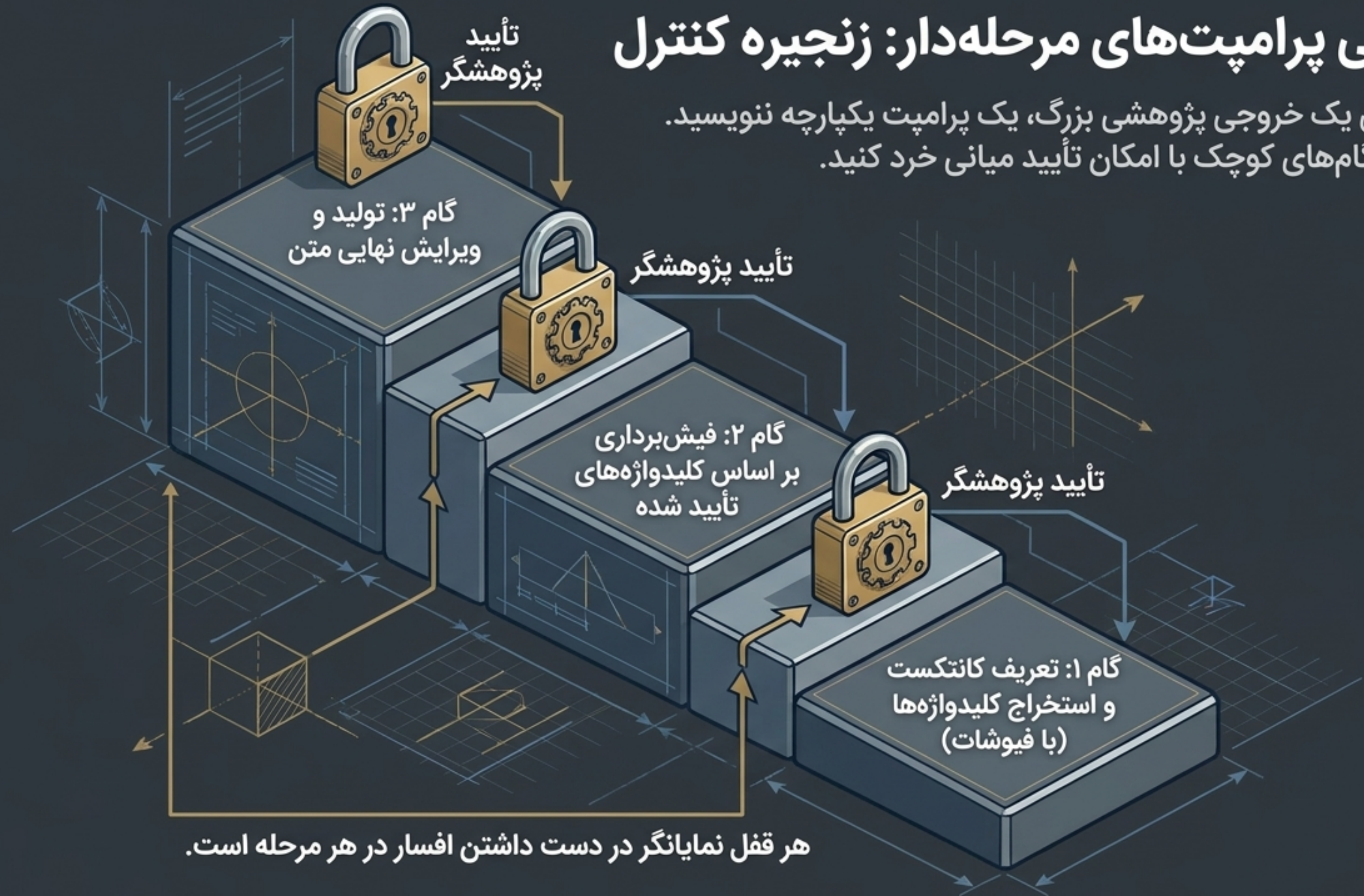
پردازش منحصرأ درون مرزهای
تعیین شده کاربر (RAG)



حافظه عمومی LLM
(مستعد توهم)

طراحی پرامپت‌های مرحله‌دار: زنجیره کنترل

هرگز برای یک خروجی پژوهشی بزرگ، یک پرامپت یکپارچه ننویسید.
کار را به گام‌های کوچک با امکان تأیید میانی خرد کنید.



مثلث مهار: معماری جامع کنترل خطای ماشین

ترکیب همزمان این سه استراتژی، توهم را به حداقل رسانده و ماشین را به یک دستیار دقیق تبدیل می‌کند.



جایگاه نهایی: هوش مصنوعی پژوهشگر نیست

“

«شما سرمربی تیم هستید.
ماشین فقط ابزار است.
مسئولیت مطلق هر ادعا، ارجاع و
ارجاع و تحلیل مستقیماً
بر عهده شماست.»



از کاربر منفعل به سوارکار مسلط

اکنون زمان آن است که دانش نظری را به مهارت عملی تبدیل کنیم.

فرمان عملیاتی کارگاه:

یک موضوع پژوهشی انتخاب کنید.
پرامپتی طراحی کنید که شامل حداقل
۲ مثال ایجابی، ۱ مثال سلبی و یک
دامنه محدود باشد.



زنده پنداری را کنار بگذارید.



افسار ماشین را با فیوشات به
دست بگیرید.



مسئولیت خروجی را تماماً
بپذیرید.