

# خوب ساختن با هوش مصنوعی

راهنمای گام به گام نرم افزار، عامل ها و ارکستراسیون ها – به ترتیبی که نتایج بهتری می دهد

Rehm KI Consulting

نسخه ۰۱/۰ ویرایش نخست ۰ ژوئیه ۲۰۲۶

## چرا ترتیب همان متغیر پنهان است

وقتی یک ساخت به کمک هوش مصنوعی خراب می شود، نخستین واکنش این است که مدل را مقصر بدانیم – به اندازه کافی باهوش نبود، توهم زد، کد شلخته نوشت. معمولاً این علت واقعی نیست. علت واقعی تقریباً همیشه ترتیب است: کار به توالی نادرست انجام شد. پیش از تعریف مسئله کد نوشته شد. از مدل خواسته شد که به جای گرفتن مواد، به یاد آورد. همان چیزی که سیستم را ساخت، تنها چیزی هم بود که آن را بررسی کرد. کاربران واقعی نخستین آزمون گر ها بودند.

ترتیب را درست کنید، یک مدل معمولی کار خوب تولید می کند. آن را نادرست کنید، بهترین مدل جهان یک آشفتگی پرمدها تولید می کند.

این راهنما همان ترتیب را شرح می دهد. این ترتیب برای سه نوع کار به کار می آید، که کمتر از آنچه به نظر می رسد با هم فرق دارند:

• **نرم افزار ساده** – یک برنامه یا سرویس معمولی.

• **عامل های هوش مصنوعی** – سیستم هایی که در آن ها یک مدل از راه ابزارها و با اندکی خودمختاری دست به کنش می زند.

• **ارکستراسیون ها و روال ها** – چند عامل یا گام که در یک خط لوله ترکیب شده اند، جایی که خود هماهنگی همان محصول است.

گام ها برای هر سه یکسان اند. هر جا که یک عامل یا یک ارکستراسیون به چیز افزون تری نیاز داشته باشد، همان گام آن را می گوید.

سه اندیشه زیر همه آنچه در پی می آید جاری اند و ارزش دارد پیش از آغاز گام ها روشن بیان شوند. **پیش از ساختن، تعریف کنید** – درخواست بیان شده به ندرت همان مسئله واقعی است. **لنجر بیندازید، نه یادآوری** – هوش مصنوعی وقتی معتمد است که مواد پیش رویش باشد و وقتی نامطمئن است که دست به حافظه می برد. **سازنده را از بازیبن جدا کنید** – هیچ چیزی که چیزی را ساخته نباید داور یگانه خوب بودنش باشد. تقریباً هر گام زیر یکی از این سه اندیشه است که در لحظه ای دیگر به کار بسته شده.

## گام ۰ – تعریف مسئله واقعی

**چه کنید:** ایده را به یک سند ایده واحد و منسجم برسانید. در آن، هدف، نیت‌های خود و – این همان بخشی است که مردم از آن می‌گذرند – نمونه‌های عینی از اینکه **موفقیت** چه شکلی است و **شکست** چه شکلی است را بیاورید.

**چرا مهم است:** درخواست بیان‌شده یک نقطه آغاز است، نه یک مشخصات. مردم نشانه‌ها را وصف می‌کنند نه علت‌ها را؛ یک راه‌حل فرض‌شده را وصف می‌کنند نه نیاز زیرین را. اگر دقیقاً همان چیزی را بسازید که خواسته شد، اغلب چیز نادرست را زیبا می‌سازید. نمونه‌های شکست به اندازه نمونه‌های موفقیت اهمیت دارند، چون مرزی را می‌کشند که سیستم نباید از آن بگذرد.

**نشانه کار خوب:** یک صفحه که یک غریبه می‌تواند بخواند و درست توضیح دهد چه می‌سازید، برای چه کسی، و چگونه هر دوی شما می‌فهمید که کار گرفته است.

**با ساده‌ترین نسخه مفید شروع کنید.** در برابر وسوسه تعریف کل سیستم بلندپروازانه به یک‌باره مقاومت کنید. کوچک‌ترین مسئله‌ای را که ارزش واقعی می‌دهد برگزینید، آن را واقعاً خوب به کار ببندازید، و توانایی‌ها را در تکرارهای بعدی (گام ۶) بیفزایید. یک چیز ساده کارآمد بر یک چیز پیچیده نیمه‌کاره می‌چربد – و نسخه ساده به شما می‌آموزد که نسخه پیچیده واقعاً به چه نیاز دارد.

**خطای رایج:** معیارهای موفقیت مبهم («خوبش کن»، «کاربران باید دوستش داشته باشند»). اگر موفقیت عینی نباشد، نمی‌توانید آن را بیازمایید و نمی‌توانید بگویید کی تمام شده‌اید.

• **برای عامل‌ها:** تعریف کنید که عامل مسئول چیست و، به همان اندازه مهم، چه چیزی را هرگز نباید انجام دهد. «موفقیت» و «شکست» به رفتارهای مشخص بدل می‌شوند – کنش‌هایی که می‌خواهید و کنش‌هایی که فاجعه خواهند بود.

• **برای ارکستراسیون‌ها:** پیش از تجزیه، نخست نتیجه سرتاسری را تعریف کنید. یک خط لوله از گام‌های به‌تنهایی درست که در مجموع به هدف نمی‌رسند، خطایی کلاسیک و پرهزینه است.

## گام ۱ – گردآوری بسته زمینه

**چه کنید:** هر آنچه را که هوش مصنوعی باید از روی آن بسازد در قالب اسناد واقعی گرد آورید: سند ایده، هدف، نیت‌ها، فهرست ویژگی‌ها، نمونه‌های موفقیت و شکست، ارزش‌ها و محدودیت‌ها، و استانداردهای کدنویسی. از این‌ها یک سند طراحی کوتاه بسازید.

**چرا مهم است:** این بزرگ‌ترین اهرم کیفیت و قوی‌ترین دفاع در برابر توهم است. یک مدل زبانی متن محتمل را پیش‌بینی می‌کند؛ وقتی فاکتی کم دارد، فاکتی محتمل می‌سازد، با همان صدای پرمدعایی که برای دانسته‌هایش به کار می‌برد. راه‌حل امید بستن به یادآوری آن نیست – این است که حقیقت را پیش رویش

بگذارید. ساختی که در یک بسته زمینه فراهم شده از سوی شما لنگر انداخته، از روی واقعیت شما کار می‌کند، نه از روی تقریب مدل.

**نشانه کار خوب:** پوشه‌ای از اسناد روشن. هیچ چیز باربری فقط در سر شما یا در تاریخچه گفت‌وگویی که هفته بعد از دستش می‌دهید زندگی نمی‌کند.

**خطای رایج:** محدودیت‌های تعیین‌کننده (یک الزام مجوز، یک قاعده امنیتی، یک قالب داده) که «بدیهی» بودند و بنابراین هرگز نوشته نشدند – پس مدل هرگز از آن‌ها خبر نداشت.

- **برای عامل‌ها:** بسته زمینه به system prompt عامل، تعریف ابزارهایش، حفظ‌هایش و مجموعه‌ای از نمونه‌های کار شده بدل می‌شود. درباره ابزارهای در دسترس و مرزهای تثبیت‌شده صریح باشید.
- **برای ارکستراسیون‌ها:** زمینه مشترک و قراردادهای میان مؤلفه‌ها را تعریف کنید – اینکه هر کدام درباره دیگری چه می‌تواند فرض کند. ابهام میان عامل‌ها همان‌جایی است که ارکستراسیون‌ها می‌پوسند.

---

## گام ۲ – نوشتن مشخصات

---

**چه کنید:** بگذارید یک مدل استدلالی قوی از روی بسته زمینه یک سند مشخصات نوشتاری تهیه کند. سپس شما آن را بخوانید، اصلاح کنید و تأیید کنید. تأیید انسانی اختیاری نیست.

**چرا مهم است:** مشخصات همان قرارداد میان نیت و پیاده‌سازی است. خطایی در مشخصات اکنون ارزان است و به محض آنکه بر آن ساخته شود گران می‌گردد. بازبینی مشخصات پرتیرترین نیم‌ساعت کل پروژه است.

**نشانه کار خوب:** مشخصاتی که شخصاً خوانده‌اید و واقعاً با آن موافقید – آن قدر دقیق که ابهام را بردارد، اما نه آن قدر خشک که هر انتخاب پیاده‌سازی را دیکته کند و بهترها را ببندد.

**خطای رایج:** سرسری خواندن مشخصات چون «درست به نظر می‌رسد». مشخصاتی که درست به نظر می‌رسد و به طور ظریف نادرست است، از نبود مشخصات بدتر است، چون اکنون همه به آن اعتماد می‌کنند.

- **برای عامل‌ها:** نقش، ورودی‌ها و خروجی‌ها، ابزارها و شرایط توقف را مشخص کنید – اینکه عامل کی باید تحویل دهد، ارجاع دهد یا رد کند.

- **برای ارکستراسیون‌ها:** هر تحویل‌دهی را مشخص کنید. پیام‌های میان گام‌ها همان رابط واقعی‌اند؛ آن‌ها را با همان دقتی مشخص کنید که یک API عمومی را.

## گام ۳ – نوشتن یک طرح ساخت مفصل

**چه کنید:** از روی مشخصات تأییدشده، یک طرح ساخت مفصل و مرتب تولید کنید – آن قدر مفصل که یک مدل کوچک تر و ارزان تر بتواند آن را بی آنکه به استدلالی که تولیدش کرده نیاز داشته باشد، تا پایان اجرا کند.

**چرا مهم است:** یک طرح قوی همان چیزی است که به شما اجازه می دهد بخش عمده کار را به مدل های اقتصادی بسپارید و هزینه ها را معقول نگه دارید (به «مدل درست برای هر کار» نگاه کنید). همچنین ماژولاریت را الزامی می کند: برای نوشتن طرح باید درزهایی را بیابید که سیستم در آن ها به طور طبیعی جدا می شود، و این درزها به رابط های تمیز بدل می شوند نه وابستگی های درهم تنیده.

**نشانه کار خوب:** کارهای مجزا و مرتب با رابط های روشن میانشان – هرکدام یا مستقیماً ساختنی یا مستقیماً بازبینی شدنی.

**خطای رایج:** طرحی که در حقیقت یک مونولیت مبدل است – گام هایی چنان درهم تنیده که هیچ چیز را نمی توان جداگانه ساخت یا آزمود.

• **برای عامل ها و ارکستراسیون ها:** قراردادهای پیام، مسیریابی (چه کسی چه کسی را و کی فرا می خواند) و مدیریت خطا را پیش از ساختن تعریف کنید. از پیش تصمیم بگیرید وقتی گامی زمانش سر می آید، آشغال برمی گرداند یا در حلقه می افتد چه رخ می دهد. ماژولار بسازید، با یک گذرگاه ارتباطی میان مؤلفه ها به صورت پیش فرض، مگر آنکه دلیل مشخصی خلافش باشد.

## گام ۴ – ساختن با انضباط

**چه کنید:** طرح را اجرا کنید. از همان خط نخست، پادشکننده و امن بسازید، نه به عنوان یک مرحله بعدی. آزمون محور و ماژولار کار کنید. آزمون ها را حین ساخت سیستم پیوسته اجرا کنید و بخواهید که پیش از هر push هر آزمون به صورت محلی پاس شود.

**چرا مهم است:** امنیت و تاب آوری ای که در پایان وصله شوند هرگز درست جا نمی افتند – فرض ها پیش تر پخته شده اند. پادشکننده یعنی سیستم چنان طراحی شده که وقتی بخشی از کار می افتد، بقیه به کار ادامه می دهند؛ خطا جدا می شود نه آنکه آبخاری شود. این تفاوت میان یک بعد از ظهر بد و یک فاجعه است، و برای هر چیزی که به عنوان سرویس اجرا می شود، اختیاری نیست.

**نشانه کار خوب:** یک ساخت ماژولار، آزمون ها به صورت محلی سبز، و افت کنترل شده نمایش پذیر – می توانید یک مؤلفه را بکشید و ببینید بقیه پابرجا می ماند.

**خطای رایج:** «امنیت و آزمون‌ها را بعداً اضافه می‌کنیم.» بعداً نمی‌آید، و وقتی بیاید، یک بازنویسی است.

• **برای عامل‌ها:** حفاظ‌ها، اعتبارسنجی ورودی، دسترسی ابزار محصورشده (sandbox) و زمان‌بندی‌ها و تلاش‌های مجدد معقول. فرض کنید مدل گاهی خروجی بد تولید می‌کند و چنان بسازید که خروجی بد نتواند آسیب بزند.

• **برای ارکستراسیون‌ها:** دامنه‌های خطا را جدا کنید تا یک عامل از کار افتاده خط لوله را مسموم نکند. گام‌ها را هر جا می‌توانید ایدمپوتنت (idempotent) کنید و مدارشکن (circuit breaker) بیفزایید تا یک مؤلفه گیرکرده یا در حلقه افتاده متوقف شود نه آنکه به کار رها گردد.

---

## گام ۵ – بازبینی مستقل و سخت‌سازی

---

**چه کنید:** بگذارید دو نشست جداگانه هوش مصنوعی به صورت خصمانه کار را نقد کنند. آن‌ها فقط نقد می‌کنند – چیزی را اصلاح نمی‌کنند. یافته‌هایشان برای پیاده‌سازی به سازنده اصلی باز می‌گردد. سپس تست نفوذ را چند بار اجرا کنید، نه یک بار.

**چرا مهم است:** حسابداری را نمی‌گذارید دفترهای خودش را حسابرسی کند، به همان دلیل که نباید عاملی را که کد را نوشته تنها داور آن قرار دهید: تضاد منافع ساختاری است، نه مسئله حسن نیت. بازبینی با زمینه تازه و بی‌سهم در کار، آنچه را که نویسندگان ذاتاً نسبت به آن کور است می‌بیند. جدا نگه داشتن نقد از اصلاح این استقلال را حفظ می‌کند – لحظه‌ای که یک منتقد به اصلاح دست می‌زند، سهمی به دست می‌آورد.

**موارد غیرقابل مذاکره.** سه چیز اینجا مسئله صلاحدید نیستند. نخست، همه چیز از یک خط لوله CI/CD می‌گذرد – هر تغییر، هر بار، به صورت خودکار ساخته و آزموده می‌شود. دوم، همه آزمون‌ها همیشه پاس می‌شوند؛ یک ساخت قرمز هرگز merge نمی‌شود و هرگز «برای اصلاح در آینده» تحویل داده نمی‌شود. سوم، تست نفوذ الزامی است، نه تجملی برای وقتی که فرصت هست – سیستمی که تست نفوذ نشده، سیستمی نامعتمد است.

**نشانه کار خوب:** نقدهای مشخص و اجرashدنی که به راحتی اجرا می‌شوند؛ سازنده و بازبینی که هرگز یک عامل نیستند؛ و تست‌های نفوذی که تکرار شده‌اند تا دیگر مسئله مهمی به بار نیاورند.

**خطای رایج:** یک گذر بازبینی واحد، یا «بازبینی‌ای» که در حقیقت همان نشست با کلاهی دیگر است. استقلالی که واقعاً الزامش نکرده‌اید، استقلالی است که ندارید.

• **برای عامل‌ها و ارکستراسیون‌ها:** پرامپت‌ها و جریان را به صورت خصمانه بیازمایید – تزریق پرامپت (Prompt Injection)، جیل‌بریک، ورودی‌های ناقص، پاسخ‌های خصمانه ابزار. کل ارکستراسیون را تیم قرمز (Red-Teaming) کنید، نه فقط عامل‌های منفرد را؛ خطاهای جالب در تحویل‌دهی‌ها زندگی می‌کنند.

## گام ۶ – آزمون انسانی، سپس تکرار

**چه کنید:** نخستین آزمون انسانی را زمانی اجرا کنید که سیستم به طور کامل بالا و در حال اجراست – نه پیش از آن. از آنجا به بعد، بر پایه بازخورد واقعی به صورت تکراری توسعه دهید.

**چرا مهم است:** آزمون یک سیستم نیمه ساخته توجه انسانی را صرف مسائلی می کند که خود ساخت خود به خود حلشان می کرد. داور انسانی کمیاب و ارزشمند است؛ آن را به سوی یک چیز واقعی در حال اجرا نشانه بگیرید، جایی که بازخورد واقعی است.

**نشانه کار خوب:** یک سیستم در حال اجرا در برابر یک انسان، که فهرست کوتاه و اولویت بندی شده ای از بهبودهای واقعی تولید می کند – نه تقلایی برای به راه انداختنش.

**خطای رایج:** استفاده از کاربران واقعی به عنوان نخستین آزمون یکپارچگی. آن ها درمی یابند که خراب است، نه اینکه ناقص است، و اعتماد را سوزانده اید تا چیزی را بیاموزید که یک آزمون دود (smoke test) به شما می گفت.

## مدل درست برای هر کار

دشوارترین تفکر – تعریف، مشخصات، برنامه ریزی و نقد – را به قوی ترین مدلی بسپارید که به آن دسترسی دارید. اجرای روزمره را به ارزان ترین مدلی بسپارید که کار را قابل اعتماد انجام می دهد. استفاده از یک مدل گران برای همه چیز پول را هدر می دهد؛ استفاده از یک مدل ارزان برای استدلال دشوار همه چیز را هدر می دهد، چون خطاها به هر گام پایین دستی تکثیر می شوند.

یک رده بندی عملی، که صادقانه به عنوان یک وضعیت کاری که زود کهنه می شود بیان می شود – نام ها و نسخه های مدل ها پیوسته تغییر می کنند و هر رتبه بندی مشخص باید پیش از تکیه بر آن دوباره سنجیده شود:

• رده برتر – دشوارترین استدلال: تعریف، مشخصات، طرح ساخت، نقد خصمانه.

• رده میانی توانمند و مدل های اقتصادی – اجرای روزمره: پیاده سازی کارهای خوب مشخص شده از روی یک طرح مفصل.

انضباط، خود فهرست مشخص نیست؛ عادت تطبیق توانایی با دشواری کار و سنجش دوره ای دوباره است، به جای اعتماد به رتبه بندی ای که شش ماه کهنه شده.

## سازنده را از بازیبن جدا کنید

این قلبِ ارکستراسیون خوب است و از این رو بخشِ ویژه خود را می‌گیرد. اصل همان اصلِ حسابدار است: هرکه کار را انجام می‌دهد نباید تنها کسی باشد که آن را بررسی می‌کند. وقتی تصمیم می‌گیرید کار چگونه میان مدل‌ها یا عامل‌ها تقسیم و واگذار شود، هر کار را از چهار پرسش بگذرانید:

1. **اندازه** – کار برای یک نشستِ واحد چند توکن است؟ کارهایی که به تمیزی در یک پنجرهٔ زمینه جا نمی‌شوند نیاز به تقسیم یا ساختاری دیگر دارند.
  2. **استقلال** – آیا می‌توان کار را بدون از دست دادنِ اثربخشی تقسیم کرد؟ برخی چیزها تمیز تجزیه می‌شوند؛ برخی معنایشان را از دست می‌دهند وقتی بریده شوند. در امتدادِ درزهای طبیعی تقسیم کنید، نه دلبخواهی.
  3. **تفکیک وظایف** – آیا انجامِ هر دو نیمه توسط یک عامل تضاد منافع می‌سازد؟ اگر چنین است، جدایشان کنید، حتی به بهای اندکی کارایی.
  4. **بازیبنی‌پذیری** – آیا بررسیِ کار بسیار ارزان‌تر از انجامِ آن است؟ اگر چنین است، از چند عاملِ بازیبن مستقل استفاده کنید. تأییدِ ارزان که به موازات اجرا می‌شود، اعتمادِ فراوانی به بهای اندک می‌خرد.
- آن نکتهٔ آخر ابرقدرتِ خاموشِ ارکستراسیون است: تأییدِ اغلبِ بسیار ارزان‌تر از تولید است، پس می‌توانید از پسِ آن برآیید که همان خروجی را چند بار به شیوه‌های مختلف بررسی کنید و چیزی را بپذیرید که هر بررسیِ منفرد از دست می‌داد.

## ضدالگوها، یک‌جا

- **حل مسئلهٔ بیان‌شده به‌جای مسئلهٔ واقعی.** نخست تعریف کنید.
- **گذاشتنِ مدل به یادآوری به‌جای دادنِ مواد به آن.** همه‌چیز را لنگر بیندازید.
- **نوشتنِ کد پیش از وجودِ مشخصات و طرح.** ترتیب همان کلِ بازی است.
- **مونولیتِ مبدل** – «طرحی» که گام‌هایش را نمی‌توان جداگانه ساخت یا آزمود.
- **امنیت و آزمون‌ها «بعداً».** بعداً یک بازنویسی است.
- **خودبازیبنی** – سازنده کارِ خودش را نمره می‌دهد.
- **یک تستِ نفوذ.** تکرار کنید تا خاموش شود.
- **کاربران به‌عنوان نخستین آزمون‌گرها.** آزمونِ انسانی روی یک سیستم کاملاً در حال اجرا آغاز می‌شود.
- **یک مدل برای همه‌چیز** – یا برای کارِ روزمره گران می‌پردازید یا استدلالِ دشوار را کم‌توان می‌گذارید.

## نتیجه صادقانه

هیچ‌کدام این‌ها عجیب نیست. هر گام به محض بیان شدن بدیهی است. دشواری این است که زیر فشار اینکه فقط شروع به ساختن کنیم، واکنش‌های پیش‌فرض بازمی‌گردند – از تعریف بگذر، به حافظه مدل اعتماد کن، بگذار خودش را بررسی کند، به کاربران تحویل بده «تا ببینیم چه می‌شود». خوب انجام دادن این کار بیش از هر چیز انضباط پیروی از ترتیب است، حتی وقتی جلو پریدن سریع‌تر به نظر می‌رسد. سریع‌تر نیست. فقط این‌طور حس می‌شود – تا گام ۶.

این یک راهنمای کلی و فراگیر درباره خوب ساختن با هوش مصنوعی است. آگاهانه در سطح اصول و ترتیب می‌ماند و نمی‌تواند جزئیات خاص یک پروژه منفرد را پوشش دهد. برای راهنمایی مفصل و متناسب با موقعیت شما، تماس بگیرید.

### درباره Rehms KI Consulting

Rehm KI Consulting به سازمان‌ها کمک می‌کند با هوش مصنوعی همان‌گونه بسازند که باید ساخته شود – با انضباط مهندسی واقعی در پس آن، نه یک دموی آراسته به عنوان محصول. این رویکرد در دهه‌ها تجربه سخت‌افزار و نرم‌افزار ریشه دارد، از طراحی تراشه و پردازنده‌های شبکه در Intel تا سال‌ها ساخت برنامه‌های وب، و امروز بر کمک به کسب‌وکارها برای به‌کارگیری هوش مصنوعی متمرکز است – صادقانه، قابل اعتماد و ساخته‌شده برای ماندگاری.

موضع ساده است: پاسخ‌های صریح درباره اینکه هوش مصنوعی واقعاً کجا کمک می‌کند و کجا نه. بدون هیاهو، بدون جادو.

Christopher Rehm · بایرن، آلمان · [contact@christopherrehm.de](mailto:contact@christopherrehm.de) · **Rehm KI Consulting**

نام‌ها، ابزارها و رتبه‌بندی‌های نسخه مدل‌هایی که در این راهنما ذکر شده‌اند به سرعت تغییر می‌کنند – آن‌ها را نمونه‌هایی برای بازبینی دوباره بدانید، نه فاکت‌های ثابت.