

آکادمی آی‌لرن (ailearnacademy.ir)

NotebookLM

📄 چکلیست گام‌به‌گام رفع توهم (Hallucination) هوش مصنوعی

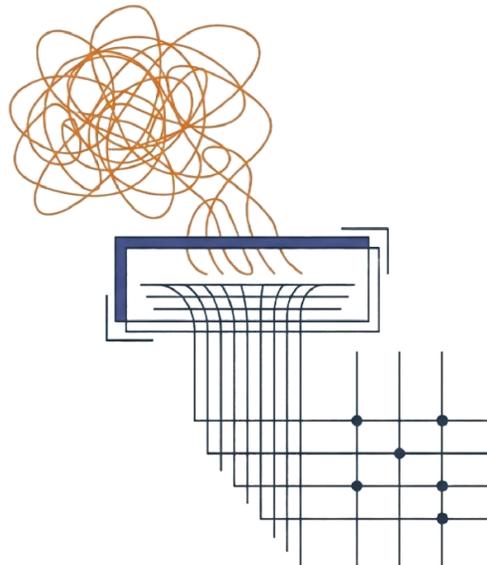
ویژه دانشجویان، پژوهشگران و کاربران حرفه‌ای

تهیه شده در: آکادمی آی‌لرن (ailearnacademy.ir)

# راهنمای گام به گام رفع توهم هوش مصنوعی ویژه دانشجویان و پژوهشگران

این چک‌لیست برای پژوهشگرانی است که می‌خواهند از هوش مصنوعی به‌عنوان یک دستیار دقیق استفاده کنند، استفاده ممنوع کنند، نه یک منبع تولید اطلاعات جعلی. با رعایت این ۱۰ اصل، خطای مدل را به صفر نزدیک کنید.

Based on the Zero-Hallucination Protocol



NotebookLM

---

## چک‌لیست گام به گام رفع توهم (Hallucination) هوش مصنوعی 🇮🇷

### ویژه دانشجویان، پژوهشگران و کاربران حرفه‌ای

تهیه شده در: آکادمی آی‌لرن ([ailearnacademy.ir](http://ailearnacademy.ir))

مقدمه: این چک‌لیست برای تبدیل هوش مصنوعی از یک «داستان‌سرای خلاق» به یک «دستیار علمی دقیق» طراحی شده است. با رعایت این ۱۰ مورد، احتمال خطا و تولید اطلاعات جعلی را به حداقل برسانید.

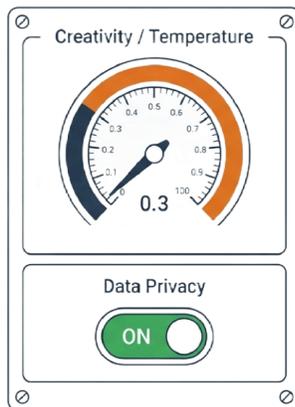
---

[ailearnacademy.ir](http://ailearnacademy.ir)

## ۱. تنظیمات فنی و انتخاب ابزار

- [ ] کاهش دما (Temperature): آیا پارامتر دما را روی ۰.۳ یا کمتر تنظیم کرده‌اید؟ (برای دقت حداکثری و کاهش خلاقیت کاذب).
  - نکته: اگر به تنظیمات دسترسی ندارید، در پرامپت بنویسید: «فقط بر حقایق تمرکز کن و میزان خلاقیت پاسخگویی را در پایین‌ترین سطح ممکن در نظر بگیر.»
- [ ] انتخاب مدل مناسب: آیا برای جستجوی منابع از مدل‌های متصل به وب (مانند Perplexity یا Gemini) استفاده می‌کنید؟

### فاز اول: تنظیمات فنی و امنیت



#### کاهش دما (Temperature)

برای دقت حداکثری، پارامتر دما را روی ۰.۳ یا کمتر تنظیم کنید.

اگر دسترسی به تنظیمات ندارید، از این پرامپت استفاده کنید:  
«در پاسخگویی فقط بر حقایق (Facts) تمرکز کن، از هیچ‌گونه تحلیل خلاقانه استفاده نکن و میزان خلاقیت (Temperature) را در پایین‌ترین سطح در نظر بگیر.»

#### حریم خصوصی (Data Privacy)

داده‌های خام، شخصی یا نتایج آزمایشگاهی منتشر نشده را هرگز در مدل‌های عمومی آپلود نکنید.

استفاده از حالت ناشناس (Incognito) و غیرفعال‌سازی گزینه Training.

NotebookLM

## ۲. مهندسی دستور و ساختار خروجی

- [ ] تعیین نقش (Persona): آیا به مدل نقش مشخصی داده‌اید؟ (مثلاً: «به عنوان یک داور مجله Nature عمل کن.»)
- [ ] شفافیت وظیفه: آیا دستور شما دقیق است؟ (به جای «بررسی کن»، بنویسید: «چکیده را در ۲۰۰ کلمه نقد کن.»)
- [ ] تعیین قالب خروجی: آیا فرمت خروجی (جدول، لیست، کد LaTeX یا Markdown) را مشخص کرده‌اید؟

## فاز دوم: معماری دستور و زمینه‌سازی



### مهندسی دستور (Prompt Formula)



NotebookLM

## ۳. هدایت استدلال و شرط عدم آگاهی

- [ ] فعال‌سازی CoT: آیا عبارت «گام‌به‌گام فکر کن» را برای فعال‌سازی زنجیره تفکر در پرامپت گنجانده‌اید؟
- [ ] شرط صداقت: آیا این جمله را اضافه کرده‌اید: «اگر از پاسخ مطمئن نیستی، صراحتاً بگو نمی‌دانم؟»

## فاز دوم: هدایت استدلال (Logic Flow)



NotebookLM

#### ۴. اعتبارسنجی و ابزارهای ضد-توهم ✓

- [ ] راستی‌آزمایی منابع: آیا لینک‌ها و کدهای DOI را در **Google Scholar** یا **Crossref** چک کرده‌اید؟
- [ ] استفاده از ابزارهای کمکی: آیا خروجی را برای تحلیل استنادات در **Scite.ai** بررسی کرده‌اید؟

### فاز سوم: اعتبارسنجی خارجی (Verification)

#### راستی‌آزمایی منابع

لینک‌ها و کدهای DOI را در **Google Scholar** یا **Crossref** چک کنید.

#### ابزارهای کمکی

خروجی را برای تحلیل استنادات در **scite.ai** بررسی کنید.

#### بازبینی انسانی (Human-in-the-Loop)

اعداد، تاریخ‌ها و نام‌های خاص را با متن اصلی تطبیق دهید. آیا لحن علمی حفظ شده یا اغراق‌آمیز است؟



#### ۵. بازبینی انسانی (Human-in-the-Loop) 👤

- [ ] تطبیق فکت‌ها: آیا اعداد، تاریخ‌ها و نام‌های خاص را با متن اصلی مقاله تطبیق داده‌اید؟
- [ ] کنترل لحن: آیا هوش مصنوعی لحن علمی را حفظ کرده یا دچار اغراق شده است؟

#### ۶. حریم خصوصی و امنیت داده 🛡️

- [ ] عدم افشای داده‌های حساس: آیا اطمینان حاصل کرده‌اید که داده‌های خام یا نتایج آزمایشگاهی منتشر نشده را در مدل‌های عمومی آپلود نکرده‌اید؟
- [ ] حالت ناشناس (Incognito): آیا گزینه آموزش مدل روی داده‌های شخصی (Training) را در تنظیمات غیرفعال کرده‌اید؟

---

## ۷. تکنیک نقد متقابل و وکیل مدافع شیطان 🗡️

- [ ] خود-انتقادی: آیا از هوش مصنوعی خواسته‌اید خروجی خودش را نقد کند؟ (مثال: «به عنوان یک منتقد، ۵ ایراد اساسی به این متن وارد کن.»)

---

## ۸. پایداری و تکرارپذیری علمی 📖

- [ ] ثبت نسخه مدل: آیا نام دقیق مدل (مثل Gemini ۱.۵ Pro) و تاریخ استفاده را برای استنادهای آینده ثبت کرده‌اید؟
- [ ] تست مجدد (Consistency): آیا دستور را در دو چت مجزا امتحان کرده‌اید تا از ثبات پاسخ مطمئن شوید؟

تکنیکی

Lab Notebook
Model: GPT-4o
Date: [Current Date] [1402/08/15]
Prompt ID: #001

### ثبت نسخه مدل (Model Versioning)

پاسخ مدل‌ها در طول زمان تغییر می‌کند. همیشه نسخه دقیق (مثلاً GPT-4o) و تاریخ استفاده را ثبت کنید.

### تست مجدد (Seed Consistency)

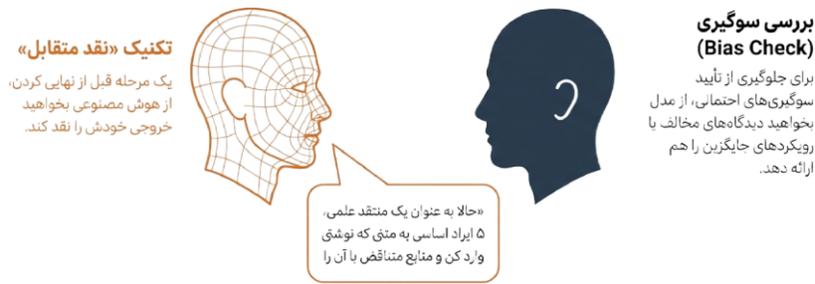
دستور را در دو پنجره چت جداگانه امتحان کنید تا از ثبات پاسخ مطمئن شوید.



## ۹. استفاده از تکنیک RAG (محدود کردن منبع)

- [ ] ارائه متن مرجع: آیا فایل مقاله را آپلود کرده و دستور داده‌اید: «فقط بر اساس فایل پیوست پاسخ بده»؟ (موثرترین راه برای جلوگیری از بافتن اطلاعات).

### فاز سوم: نقد متقابل و بررسی سوگیری



## ۱۰. بررسی سوگیری (Bias Check)

- [ ] تنوع دیدگاه: آیا از مدل خواسته‌اید دیدگاه‌های مخالف یا رویکردهای جایگزین در آن حوزه علمی را هم ارائه دهد؟

💡 **نکته حرفه‌ای:** همیشه یک نمونه (Example) از خروجی مطلوب خود را به مدل بدهید (Few-Shot Prompting). این کار معجزه می‌کند!

⚠️ **هشدار:** طبق بررسی‌های سال ۲۰۲۶، بیش از ۴۰٪ منابع تولید شده توسط چت‌بات‌های عمومی در حوزه آکادمیک، نامعتبر یا ترکیبی هستند. بدون فیلتر به AI اعتماد نکنید!

# چک لیست رفع توهم هوش مصنوعی

## راهنمای گام به گام

راهنمای عملی برای پژوهشگران جهت اطمینان از صحت علمی خروجی‌ها

### بخش دوم: مهندسی دستور و هدایت استدلال

### بخش اول: آماده‌سازی و تنظیمات فنی (قبل از شروع)



#### کاهش دما (Temperature)

برای دقت حداکثری

پارامتر دما را روی ۰.۳ یا کمتر تنظیم کنید یا در پرامپت صراحتاً بخواهید که «خلافت» را حذف کرده و فقط بر حقایق تمرکز کند.



#### انتخاب مدل مناسب

برای جستجوی منابع، حتماً از مدل‌های متصل به وب مانند «مانند» Gemini یا Perplexity استفاده کنید.

### بخش سوم: اعتبارسنجی و فیلترهای نهایی



#### راستی‌آزمایی و بازبینی انسانی

کدهای DOI را در Google Scholar چک کنید و فکت‌ها، امداد و تاریخ‌ها را با متن اصلی مقاله تطبیق دهید.



#### تکنیک نقد متقابل (Self-Critique)

از هوش مصنوعی بخواهید در نقش «وکیل مدافع شیطان»، ۵ ایراد اساسی به خروجی خودش وارد کند.



#### استفاده از RAG دستی (Grounding)

به جای سوال کلی، فایل مقاله را آپلود کنید و بخواهید «فقط بر اساس فایل پیوست» پاسخ دهد.



فرمول طلایی پرامپت‌نویسی: تعیین نقش (مانند داور مجله)، شفافیت در وظیفه (تعداد کلمات مشخص) و تعیین قالب خروجی (جدول یا Markdown) الزامی است.



فعال‌سازی تفکر گام‌به‌گام (CoT) با افزودن عبارت «گام‌به‌گام فکر کن» و شرط «اگر نمی‌دانی، بگو نمی‌دانم»، احتمال ساختن پاسخ‌های جعلی را کاهش دهید.



تکنیک Few-Shot (نمونه دادن) همیشه یک نمونه از خروجی مطلوب خود را به هوش مصنوعی بدهید: این کار در کیفیت نتیجه معجزه می‌کند.

### بخش چهارم: اخلاق، امنیت و پایداری



حفظ حریم خصوصی داده‌ها: از آپلود داده‌های آزمایشگاهی منتشر نشده خودداری کرده و آموزش مدل روی داده‌های شخصی را در تنظیمات غیرفعال کنید.



پایداری علمی و ثبت نسخه: نام دقیق مدل (مثل GPT-4o) و تاریخ استفاده را ثبت کنید و برای اطمینان از ثبات، دستور را در دو پنجره چت جداگانه تست کنید.



بررسی سوگیری (Bias Check): برای جلوگیری از تایید سوگیری‌های احتمالی، از مدل بخواهید دیدگاه‌های مخالف یا رویکردهای جایگزین را نیز ارائه دهد.