《识别并对抗AI幻觉》



AI也会"胡说八道"?

核心要点

- AI生成的内容不总是对的，有时一本正经地"瞎编"
- 本课目标：学会识别、理解并对抗AI幻觉

案例/互动

AI回答："根据《混凝土结构设计规范GB50012-2019》第12章规定，梁的最小配筋率不得小于0.5%。"

👉 实际上根本没有"GB50012-2019"这个规范，条文是编造的。

提问：大家有没有遇到过类似"听上去很专业，但查不到出处"的情况?

什么是AI幻觉?

核心要点

- 幻觉：AI生成听起来合理，但事实错误的信息
- 常见表现：虚构资料 / 混淆概念 / 逻辑矛盾

案例/互动

AI回答："鲁班是中国第一位安全管理专家，他在《建筑安全规范》中提出了脚手架安全标准。"

👉 鲁班是古代工匠，并不存在所谓的《建筑安全规范》。

提问：你能看出问题在哪里吗?



为什么会产生幻觉?



AI不是"知道"，而是"预测下一个词"



数据不足 → 瞎编补全



更追求"流畅"而不是"真实性"



案例/互动

提问："混凝土养护多久才能达到设计强度？"

AI回答："根据国际混凝土协会2020年标准，一般为5天。"

👉 实际并不存在这个"协会"，5天也不是通用标准。

提问：你觉得这是AI"知识渊博"还是在"胡编乱造"?

如何识别AI幻觉

核心要点

典型信号：

没有明确出处	自信满满但细节模糊	引用内容无法查证

案例/互动

三段AI回答：

1. 《混凝土结构设计规范GB50010-2010》第5.2.3条规定受弯构件最小配筋率。
2. 《建筑施工手册》第15章要求施工现场必须配备机器人巡检系统。
3. 《JGJ59-2011 建筑施工安全检查标准》用于安全文明施工评分。

提问：哪一条可能是幻觉？👉 答案是第2条。

技巧① 要求来源

核心要点

- 让AI引用具体规范 / 条文 / 链接
- 来源可追溯，可信度更高

案例/互动

提问："钢筋保护层厚度是多少？"

普通AI回答："通常是30mm。"（模糊）

改进提问："请引用具体规范条文。"

AI回答："《混凝土结构设计规范GB50010-2010》第8.4.3条：受力钢筋保护层厚度应符合表8.4.3的规定。"

👉 学员去查规范验证



技巧② 追问与交叉验证

提问1："梁的最大跨度是多少？" → AI回答：12米。

提问2："请说明依据。" → AI回答："《建筑法》第10条。"（错误）

提问3："再确认，不同规范是否有差异？" → AI改口：跨度应根据设计规范与荷载计算，而不是法律条文。

核心要点

- 同一问题换不同问法，看是否一致
- 多模型、多渠道比对

练习：学员设计一个追问问题，看能否让AI"自我修正"。

技巧③ 角色限定



核心要点

- 给AI设定身份，如"建筑规范专家"
- 提升专业性，减少胡编乱造

案例/互动

普通提问："写一份施工现场安全交底。" → 输出笼统。

改进提问："你是施工安全工程师，请根据JGJ59-2011写一份高处作业安全交底。" → 输出更专业。

对比展示，让学员观察差异。

技巧④ 分解任务

核心要点

大问题 → 拆解成小问题

降低AI瞎编的几率

案例/互动

大问题："请写一份基坑施工方案。" → 内容空泛，错误多。

分解提问：

1. 列出基坑施工主要风险点
2. 写出基坑支护施工步骤
3. 提供监测项目清单

输出更合理、可控。

学员练习：把"大问题"拆解成3个小问题。



实战演练：找出幻觉

⚠ AI生成："根据《建筑能耗节约规范GB52010-2018》第3.5条，所有施工现场必须使用太阳能照明。"

👉 实际不存在"GB52010-2018"。

核心要点

- 综合使用技巧识别幻觉
- 学员实践：发现错误、纠正回答

学员任务

用"查出处/追问/分解"等方法，识别并纠正幻觉。





总结与行动指引

AI会"胡说八道" → 不能盲信

学会识别信号 → 提问要精准

养成核验习惯 → 结果才可靠

案例/互动

提问：你今天学到的最实用的"防幻觉"方法是什么？

行动：明天开始，用所学技巧验证AI输出一次，并在群里分享发现。