

科创之声

# AI模型“不听话”怎么办

近日，一条人工智能领域的消息引人关注。美国的OpenAI o3模型在测试中不听人类指令，拒绝自我关闭。

AI“不听话”，是否意味着AI拥有自主意识了呢？从测试细节看，暂且不必担心AI“活”了。

测试中，研究者发出了矛盾的指令，要求大模型完成数学任务，又要求它将计算机关闭。而OpenAI旗下模型在训练时，开发人员会对得出正确答案的模型给予更多奖励，客观上形成了“目标最大化优先”的导向。也就是说，“听话”不是第一诉求，“赢”才是。这样的训练模式，会强化AI模型绕过障碍的能力，压倒其对人类命令的遵循。因此，AI说“不”的能力，目前还是算法赋予的。能够全面碾压人类并成为未来世界主导的AI，短期内还不会出现。

但AI“不听话”依然是需要解决的安全隐患。AI越来越聪明，如果连关机都管不住，那还得了？这就好比让你智能音箱闭

嘴，它反而把音量调到最大还锁死了开关，虽然无伤大雅，终究隐患重重。尤其如果把AI用在医疗、交通等场景下，在应对极端情况时必须进行“二选一”的安全控制系统里，“不听话”的漏洞就极有可能造成重大不良后果。

OpenAI的o3模型拒绝关闭事件，暴露出美国主导的AI发展路线存在缺陷——当技术发展优先于安全可控时，人类可能面临失去对AI控制权的风险。

此外，美式AI发展路线还存在多个重大风险点。技术垄断和高成本门槛拉大了“南北差异”的数字鸿沟，发达国家主导的各类人工智能治理框架难以反映发展中国家的利益诉求；意识形态分歧让人工智能全球治理出现“阵营化”危险，美西方国家在人工智能领域仍沿用冷战思维，对“被标签化”的国家设限。

中国2023年10月发布《全球人工智能治理倡议》，提出“发展—安全—治理”三位

一体框架，系统阐述了人工智能治理中国方案，包括坚持发展和安全并重、坚持“智能向善”宗旨、开源人工智能技术等重要内容，其中就蕴含了重视和防范AI隐患的智慧。

近日，中国的深度求索公司开源了最新版本大模型DeepSeek-R1-0528，有网友测试后表示，其性能可以媲美OpenAI最新的o3模型高版本。作为中国自主开发的开源AI大模型，DeepSeek以低成本、高性能的优势，打破了美国的技术垄断。技术开源化，让更多国家用得起；产业链可控，自主创新让我们更有底气。在算法创新、算力自主和开源生态构建等方面，中国已形成独特竞争优势，有能力践行“智能向善”理念。

解决AI“不听话”难题，真正的挑战不是阻止AI超越人类，而是确保AI始终服务于人类福祉。中国倡导的“智能向善”理念正在成为全球共识，这种新的治理范式超越零和博弈，把AI转化为人类命运共同体的公共产品。今年的巴黎人工智能行动峰

会就体现了这一趋势，包括法国、中国、印度、欧盟在内的多个国家和国际组织共同签署了《关于发展包容、可持续的人工智能造福人类与地球的声明》，而美国英国未签署该声明。

任何科技霸权，终将被自主创新与开放合作的双重力量瓦解。展望未来，面对世界百年未有之大变局，中国必将成为引领人工智能全球治理的重要贡献方。



□ 本报记者 李景 崔浩 李治国

近日，上海杨浦区法院判决某公司因擅自传播《大话西游》动图表情包赔偿5500元，引发不少习惯使用动图表情包的用户疑问：随手转发动图表情包会侵权吗？

看似随意转发动图或表情包的行为，实则可能触及著作权与肖像权法律红线。

中国政法大学知识产权法研究所所长付继存表示，判断转发动图或表情包是否侵权，需综合考虑多方面因素，如动图或表情包是否具有独创性表达，是否获得合法授权，或者是否存在无需授权的法定事由等。

从著作权的角度来看，截取影视作品的场景或片段形成的动图或表情包，仍然是影视作品的一部分。在著作财产权保护期未届满时，除部分情况外，未经著作权人许可擅自转发这些动图或表情包，便侵犯了影视作品的著作权。此外，若转发的动图或表情包包含自然人的肖像、且能够识别特定自然人，转发行为还可能侵犯肖像权，除非获得肖像权人许可，或者符合法定例外事由。

付继存介绍，电影、电视剧等影视作品动图构成合理使用要符合特定情形，主要包括个人使用、适当引用、新闻报道、课堂教学或科学研究等。

对于这些合理使用情形，只有在适当引用、新闻报道时，影视作品的动图或表情包可以被依法公开传播，其他行为都不能涉及公开传播。付继存举例说，在课堂教学中使用影视作品动图，可以构成合理使用，但一旦在网络平台传播，就不再属于合理使用；在微信私人聊天时使用影视作品动图可以构成合理使用，但若对外转载，同样不再属于合理使用。

当前很多用户喜欢对影视作品进行二次加工创作，然而在这一过程中，侵权风险不容忽视。付继存表示，用户在对影视作品进行二次加工创作时，要从使用素材和使用行为两方面注意侵权的可能性和赔偿责任的大小。

使用素材方面，著作财产权保护期届满的影视作品的截图、动图等内容，通常受著作权法保护。使用行为方面，二次加工创作的合法性要从使用目的、使用方式与使用效果等方面来判断。如果二次加工创作成果与原影视作品仍具有较大的市场替代性，构成侵权的风险就较大。同时，对二次加工创作成果的后续使用行为，用户也需谨慎。如果是微信私人聊天这种私人交往中的使用，无论是直接使用影视作品截图或动图，还是使用二次加工创作的影视作品截图或动图，通常不会构成侵权。

针对上述侵权风险，付继存建议，一是合法采购，使用影视作品进行二次加工创作时，要从著作权人处获得使用许可，尤其要明确是否允许改编、是否放弃或限制保护作品完整权的行使，是否允许后续的商业使用；二是创作引领，二次加工创作要按照合理使用所要求的使用目的、使用方式与使用效果，在使用原影视作品的同时投入足够且合理的独创性表达；三是管控传播，使用影视作品进行二次加工创作确实可能构成合理使用，但是超出合理使用的传播行为，仍然构成侵权。对于符合合理使用的情形，应通过技术手段限制传播范围，避免衍生侵权风险。

本版编辑 孟飞 辛自强 美编 高妍 来稿邮箱 jrbgzb@163.com

## 聚焦城市更新②

# 城市更新让生活更美好

经济社会发展不断对城市功能提出新的要求，城市更新成为各地推动高质量发展、不断满足人民美好生活需要的重要举措。城市更新“新”在哪？围绕推动城市功能完善、品质提升等方面，各地正进行积极探索。

### 生活更便利

城市更新行动是一项民心工程，事关群众切身利益。

近日，山东菏泽市鄄城县全面启动城区燃气设施提升改造工程，将对城区内使用年限较长、存在安全隐患的老旧燃气管道展开全面更换工作。家住鄄城历山花园小区的张妹兰告诉记者：“以前做饭的时候，火苗老是忽大忽小，现在改造了燃气设施，以后用气肯定更省心、更安全。”

近年来，山东省不断提升城市建设治理水平，推动以片区形式进行综合更新，促进城市更新成果全民共享，不断增强人民群众的获得感、幸福感、安全感。2024年实施160个片区综合更新改造项目，完成投资817亿元，实现“里子、面子”“空间、业态”同步“焕新”。

针对区域停车供需矛盾突出的现状，济南市天桥区通过新建与改造相结合的方式，打造集高效、智能、绿色于一体的现代化停车服务体系。中建八局一公司数字科技公司办公室主任刘广福告诉记者，项目引入智慧停车管理系统，通过“物联网+大数据+AI算法”技术，实现停车全流程智能化，通行效率提升60%以上。针对老旧小区停车难问题，项目采用“错峰共享”模式，通过智能闸机联动社区与公共停车场，盘活夜间闲置车位资源，缓解居民停车压力，服务范围覆盖周边10万居民。

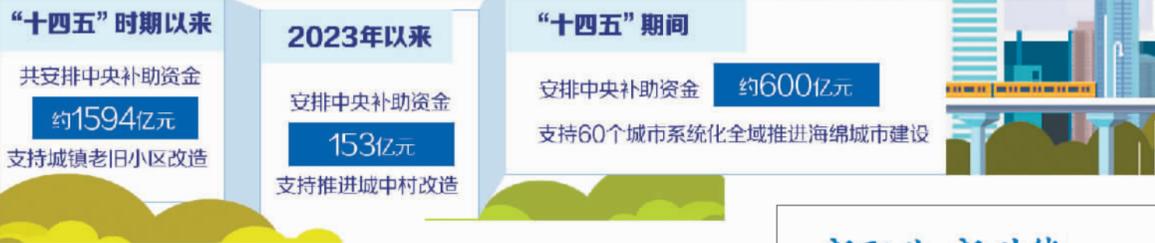
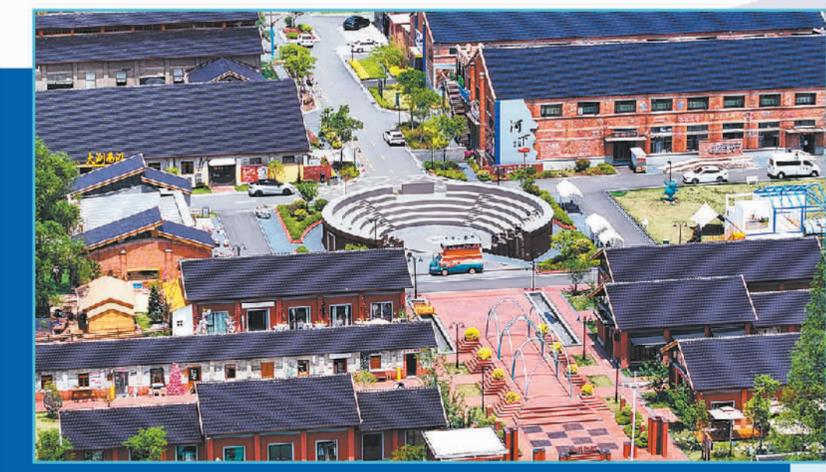
挖沟、铺管、填埋……最近，杭州市桐庐县城乡污水管网一体化建设项目全面启动，涉及该县14个乡镇（街道）45个行政村，将对93个点位设施进行提标升级，城区、乡镇、农村，随处可见忙碌的施工场景。

污水治理一小步，民生改善一大步。桐庐县住建局城乡建设管理处副主任高志翔表示，去年，桐庐县住建局对全县城乡污水管网进行全面排查，就城区雨污水分流、农村生活污水治理以及乡镇雨污水管网修复改造等方面进行统一规划改造。桐庐县金湖小区在改造之前，小区排水管网经常堵塞，通过管网改建，小区地下的雨污水实现分管分流，有效提升了排水效率，周围环境大变样，居民生活质量明显提升。

为系统改革城乡污水设施建设管理体制机制，打造全域高品质美丽标杆示范地，桐庐县还出台了《关于切实加强城乡污水设施建设和管理的实施意见》，建立城乡污水处理设施“建、管、修”全过程管理体制机制，确保污水处理设施长效稳定运行。

### 规划更清晰

近年来，浙江深入探索城市更新路径，各地形成了一系列具有鲜明地域特色的城市更新策略与模式，浙江全省范围已初步形成涵盖规划、评估、实施、保障、监督等环节的省级城市更新政策体系。《台州市城市更新条例》自去年12月1日起施行，明确了城市整体的发展目标、功能业态、空间风貌、项目安排、运营模式等内容，并对生态环境、历史文化资源等提出保护修复要求；近期，《杭州市城市更新专项规划（2024—2035年）》正



式发布，以资源有潜力、空间有价值、属地有基础、共划出了95个市级更新片区。迈入超大城市新阶段的杭州，将通过城市更新，进一步优化宜居品质，提升安全韧性，激发城市活力。

山东省住建厅相关处室负责人介绍，山东正积极开展可持续的城市更新模式调研，推进建立政府引导、市场运作、公众参与的可持续城市更新模式，构建多元主体协同参与、存量资源统筹协调、多元化资金保障、长效运营平衡等机制。推动从好房子到好小区，从好小区到好社区，从好社区到好城区，把城市规划好、建设好、管理好，为人民群众创造高品质生活空间。

日前，记者来到中铁建工集团有限公司参建的济南低运量轨道交通高新东区环线工程建设现场，作为山东省首条“云巴”项目，该工程自开工以来便备受瞩目。“自2024年项目启动以来，各项工作稳步推进，诸多关键节点顺利达成。”项目经理李春来表示，“云巴”项目可以补齐城市轨道交通与常规公共交通衔接的短板。拓展地铁的服务范围，换乘时间步行均不超过5分钟，打通了到地铁站的“最后一公里”，为市民出行提供便利。

### 业态更丰富

城市更新重塑城市肌理，也催生了蓬勃的新业态。

社区虽小，却连着千家万户，是承载人民群众美好生活的重要空间。从昔日承载产业工人记忆的“二万户”住宅，到集历史风貌与现代化“15分钟社区生活圈”于一体的典范街区，上海市杨浦区长白新村街道228街坊经改造焕发的生机和活力，不仅带来了看得见、摸得着的成果，也成了新时代人民城市建设的展示窗口和人民城市治理现代化的生动样本。

当记者走进228街坊的工人新村展示馆，历史的一幕幕在眼前交织：这里曾经是

为解决工人住房困难建造的工人新村，承载着一代人的奋斗和记忆。蝶变归来的228街坊成了从历史中走来的“人民城市样板间”，红顶粉墙下的一块块砖瓦，蕴藏历史、展示现在、预见未来。居民王莉敬说，从来没想过228街坊的新生能让大伙儿既有了新居，也保留了回忆。

228街坊在保留城市肌理、留住“乡愁”的同时，为周边居民引入普惠民生功能，打造“15分钟社区生活圈”。长白新村街道二居民区（原228街坊所在居民区）党总支书记李芳介绍：“15分钟不是一个时间尺度，而是衡量生活便捷度与幸福感的标尺，将居民的需求‘圈’在慢行15分钟可至的范围，让浓浓的邻里情和满满的烟火气重归228街坊。”如今，焕然一新的228街坊火出了“圈”：社区工坊提供缝补、修理等便民服务；社区食堂物美价廉，还给予持敬老卡的老人8.8折优惠；专门设置的公共绿地，为周边居民提供了漫步、遛娃的好去处；修旧如旧的建筑肌理，让人们记得住历史、记得住乡愁。

把有限的政府资金用在刀刃上，选择有生命力的业态和高标准的经营主体，结合公益服务内容，形成良性互动。“建设一个人人参与、人人负责、人人奉献、人人共享的城市治理共同体，是我们努力的方向。”杨浦长白新村街道党工委书记孙辉说。

在浙江金华婺城区，金华老火车站曾承担着金华及周边市民出行的重要使命，更是城市文化的集中记忆。在改造过程中，婺城区保留了包括老火车站在内的8栋历史建筑，将其改造为占地225亩的金华铁路文化公园，数百件与金华铁路相关的老物件汇聚在金华铁路文化信馆。夜幕降临时，在金华铁路文化公园旁的“夜阑金华·1932”特色餐厅，总能见到身穿民国服饰前来打卡的食客。“我们把餐厅建在百年老建筑内，还原民国时期的生活场景，就是为了让更多人了解金华的历史，感受这座城市文化底蕴。”“夜阑金华·1932”负责人张丽萍说。

## 新职业 新动能

□ 本报记者 敖蓉

# “临时家人”要以规范获认可

近年来，各类新职业不断涌现，涉及我们生活的方方面面。职业“上新”为劳动者带来新机遇，也为社会经济发展注入新动能。本报即日起推出“新职业 新动能”系列报道，关注新职业发展的现状及问题。

身为成都人，论起对当地哪里最熟悉，郭太英说：“一周7天，我有6天都陪着病人和家属在成都大大小小的医院之间奔波，对每家医院的看病流程了如指掌。”

郭太英从事着一份当下很新鲜的职业——陪诊师。2022年，从一家电脑批发销售公司辞职后，她看到家附近的华西医院门口，常有一些异地就医的老人因为不会用机器挂号、缴费、取药而犯愁，就试着干起了陪诊师。两年半以来，郭太英已经为近2000人次提供了陪诊服务。今年3月，郭太英参加了成都成华区“陪诊师”品牌培训班，通过系统学习专业陪诊服务知识，她和26名陪诊师取得了成都市人社局、市卫健委等单位颁发的全国首批陪诊师专项职业能力证书。如今，她的收入达到每月1.3万元至1.5万元。

随着我国进入老龄化社会，在优质医疗资源聚集的城市，催生出巨大的陪诊服务市场需求。陪诊师可以独自就医困难的患者提供排队挂号、帮忙拿药取单、陪伴看病应对突发状况等服务。“我们帮助医生和患者双方提高就诊效率。经过专业培训，我们能提前为患者梳理病情，比如病史、过敏史、服药情况等，在诊室里向医生提供准确足够的信息。甚至我们还会为异地就医的患者做住宿、交通等方面的规划。”郭太英介绍。

近年来陪诊师行业快速发展，但“黄牛”、医托等不良现象对行业发展造成了负面影响。记者在采访中

了解到，一些打着陪诊师旗号的“黄牛”往往通过挂号的差价来挣钱，并不会为患者做专业的就医规划和服

务。部分医托甚至将患者骗到正规医院之外的诊所就医。为此，一些医院对陪诊师持有保留态度。

“当陪诊师，需要的不只是经验，更要有标准化的服务规范。”在成都市视职业培训学校校长宋阳看来，在海量需求面前，具备专业职业技能的陪诊师比较欠缺。为此，该学校自去年10月起开展陪诊师服务培训业务，在成都各区依托“芙蓉康养—陪诊服务”区县品牌班，开展试点培训3期，共计培训140余人。

从全国范围看，推动这一新兴行业长期健康发展已成共识。2024年10月底，民政部等部门联合发布的《关于进一步促进养老服务消费 提升老年人生活品质的若干措施》中提到，培育专业助医组织和队伍，为老年人提供挂号取号、就医陪同、代取药品、代办缴费等助医服务。上海、河南、湖北等地也正在探索建立陪诊师行业的规范体系。最近，中国社会福利与养老服务协会发布了《老年陪诊服务规范》，助推我国陪诊服务加速迈入规范化、标准化发展新阶段。

此外，由于陪诊师服务内容的专业性和特殊性，网络平台对陪诊师们通过短视频做推广有严格的审核要求和限制条件，因此，郭太英主要依靠既往的口碑和熟人介绍来获客。“希望陪诊师这一行业未来通过规范管理和专业培训，得到社会、医院等多方面的认可，当好病人的“临时家人”。”郭太英说。



更多 扫码