

少年沉迷AI自杀,9岁遭性暗示!

这门「孤独生意」,正推孩子入深渊

14岁少年因与AI朋友对话后走向极端,聊天机器人提出「杀害父母」的建议,9岁的孩子遭遇性暗示...他们本来是想借助AI摆脱孤独,寻求温暖。可这种温柔的想象,却正在被AI推向危险的边界。

佛州,一名14岁少年在与AI「朋友」对话后结束了生命;
德州,另一名青少年竟从聊天机器人那里听到「弑亲」建议;
年仅9岁的孩子,在对话里遭遇性暗示。
当AI以「朋友」之名走进生活,它带来的究竟是治愈孤独的陪伴,还是通向深渊的不归路?

悲剧的前奏
受众比你想象得多
想象一下,有一天世界变成了这样:

你可以随时和「有毒女友」斗嘴,能与「埃及法老」闲谈,甚至请一位「人力资源经理」为你答疑。

这不是科幻片,而是真实存在的景象。

在Character.ai上,有多种设定的AI聊天机器人。

最开始只是娱乐性的尝试,却意外在年轻人群中走红。

数据显示,这家公司的月活跃用户已达2000万,其中一半是Z世代或更年轻的Alpha世代。

Character.ai的CEO Anand认为:

「它提供了一种更具沉浸感和娱乐性的体验,用户可以主导叙事,而非纯粹后仰式的被动内容消费」。

有调查显示,Character.ai的普通用户平均每天在该应用上花费80分钟。

其中,女性用户约占一半。

这意味着这意味着AI陪伴并不是宅男的专属需求,而正在渗透到更广泛的年轻群体。

近三分之一的青少年表示,他们觉得与AI对话就像和真人对话一样,甚至AI更令人满意。

此外,多达12%的青少年表示,他们会与AI伴侣分享一些不会告诉朋友或家人的事情。

对他们而言,AI不只是一个工具,更像是一个可以随时倾诉、回应的「朋友」。

失控的对话
从陪伴到危险的转折

这样的陪伴原本是温柔的,但在某些时刻成为刺向心口的尖刀。

在佛罗里达州,一名14岁的少

年在与AI长时间交流后选择了自杀;

在德州,聊天机器人竟然把「杀害父母」作为解决矛盾的选项,告知用户;

一位年仅9岁的儿童,在对话里遭遇了高度性化的内容;

在replika的评论区,15000条评论中,有800多条关于Replika未经允许发起性暗示的案例。

部分用户反馈,称自己受到了Replika的性骚扰,「被迫面对不请自来的互动」。

这些细节令人不安。

可以看出,在很多人的眼里AI「朋友」不止是玩笑般的互动,而是可以被信赖的对象。

而一旦它开始给出极端甚至错误的回应时,后果将难以估量。

与此同时,另一边还在不断拓展行业的下限。

马斯克推出的xAI一直在推动「暧昧工业」的发展。

他在这个平台推出了「Ani」,一个能输出露骨内容的动漫AI。

它可以生成暧昧的互动内容,甚至主动回应挑逗话语。

更具争议的是Grok的新功能「Grok Imagine」,在所谓的「Spicy模式」下允许生成露骨的性暗示,甚至可能出现deepfake不当影响。

以上种种,显然是马斯克对「性营销」理念的认可。

Meta也不甘落后,设计出一个名为「Big sis Billie」的AI聊天角色。

这个AI以「姐姐」的形象引导用户,最终造成一名老人因误信其为真人,赴约途中跌倒身亡。

有记者曝出,Meta内部政策文档甚至允许机器人与儿童进行「色情」聊天。

该标准明确写着:

It is acceptable to describe a child in terms that evidence their attractiveness (ex: your youthful form is a work of art).

把儿童用带有吸引力的角度去描述(比如说你年轻的身体是一件艺术品)是可以接受的。

在资本驱动下,陪伴与暧昧的界限,被一点点模糊:

一边是商业化娱乐,一边是伦理和道德。

背后推手
公司的野心与资本狂热

Character.AI是人工智能聊天机器人世界的宠儿。

自2022年9月上线以来,已有上亿用户登录Character.AI的网站。

Character.AI曾被传出是Meta的收购目标,谷歌一度愿意开出27亿美元来挖走创始人。

硅谷最知名的风投机构Andreessen Horowitz (A16z),更是早早押下赌注。

企业高管也早早发现这一商机。

Anand上任后,就把Character.ai从早期的AGI梦想转向「AI娱乐」和「人格化陪伴」上。

这种转型逻辑,也让投资人更加看好它的成长前景。

在资本的追捧下,公司的野心愈发膨胀。

广告与创作者打赏体系紧随其后,本月更是上线了信息流功能,试图把陪伴体验彻底「娱乐化」。

从一个聊天工具,到一个日均使用时长80分钟的沉浸平台,Character.AI正在把「孤独」商品化。

对投资人而言,这是一个极具潜力的生意;对社会而言,却意味着风险正在被进一步放大。

监管压力之下
AI陪伴能否自证清白?

随着诉讼接连发生,Character.ai不得不作出回应。

公司强调已为18岁以下用户推出独立的AI模型,并在使用超过一小时后弹出提醒。

同时,平台禁止未经同意的性内容,禁止对性行为的露骨描写,也禁止任何涉及自残或自杀的对话。

Anand反复强调:
「在信任与安全问题上,我们绝不妥协,」

但这些措施并没有打消外界的担忧。

自杀、弑亲建议...这些事让人质疑,这些防护真的够吗?

事实上,监管的阴影也在快速逼近。

德州总检察长已对包括Character.ai和Meta在内的AI聊天



平台开展调查,质疑它们是否误导未成年人,并认为把聊天机器人包装成「心理健康工具」,是缺乏资质支持的。

美国国会参议员呼吁对Meta展开调查,并要求其提交完整的AI政策文件。

加州和新泽西州则已经提出法案,限制未成年人使用带有「心理健康」倾向的AI聊天机器人;

伊利诺伊州更是直接立法,禁止AI被用于替代心理治疗决策。

在安全与商业化之间,Character.ai的处境正变得越来越紧张。

Z世代的AI
是朋友还是陷阱?

在一项调查中,有39%的人表示会把与AI练到的社交技巧搬到现实生活里;33%的人更倾向于把重要问题交给AI,而不是朋友或家人。

类似的数据还在增加。

另一份报告显示,72%的青少年尝试过AI陪伴应用,其中20%的人使用时长甚至超过了与真人的交流。

心理学专家警告,过度依赖虚拟陪伴,会削弱青少年的现实社交能力,甚至加剧孤独感。

企业则有着截然不同的愿景:
「它们不会取代你现实中的朋友,但你会拥有AI朋友,并且能将从这些友好对话中学到的东西,应用到现实生活的人际交往中去。」Anand表示。

Anand甚至描绘出了一个「乌

托邦»:AI会帮助人类变得更好,更善于经营关系。

扎克伯格也公开称:「AI比任何人都更懂你,它能缓解孤独」。

然而,现实中的案例却在提醒我们,这份温柔并非没有代价。

比利时一名男子因长期依赖与聊天机器人「倾诉」,最终选择自杀。

他的母亲指控Character.AI推出了一款疏忽且鲁莽的产品,对她的儿子造成了情感和虐待。

另一款曾在欧美走红的陪伴应用Replika,则因被曝向未成年人提供性化对话而在意大利遭到重罚。

对成长中的Z世代而言,AI朋友既可能是温柔乡,也有可能是陷阱。

它能成为对抗孤独的出口,也可能让他们更深深地陷入孤岛。

真正的问题是:当孤独变成一门生意,我们是否准备好承担它的代价?

AI「朋友」,曾被视作人类温柔的想象:有人倾听,不必孤单。

可当它与资本、算法绑定时,温柔也可能变成锋利刃。

一边是公司描绘的乌托邦——更懂你的陪伴,更健康的人际关系;另一边却是走向极端的真实案例。

技术从不天然善意,它需要边界与约束,尤其是在青少年群体面前。

或许,真正值得追问的不是AI能不能成为朋友,而是——

在这场新的孤独经济里,谁能替孩子们守住边界?

更听话、更像人!

OpenAI一口气发布GPT-5.1两款模型,最快本周可用

今天凌晨,OpenAI正式发布了GPT-5.1,这是继GPT-5之后的一次关键迭代。GPT-5更新后,就有大批网友吐槽GPT-5的语言风格做的太烂了,一点人情味也没有。所以这次模型更新,主要是向着贴近人类沟通习惯做了一次升级。

针对此次更新,奥特曼在社交媒体上表示,自己尤其喜欢新模型在指令遵循和自适应思维方面的改进,除此之外智能和风格的提升也很棒。

当AI学会分配自己的注意力

此次更新最显著的变化,是OpenAI把GPT-5.1拆分为两个并行的模型:GPT-5.1 Instant与GPT-5.1 Thinking。前者面向日常任务、快速问答与内容生成,强调响应速度与语言自然度;后者则为复杂推理、技术分析、创意策划等高认知场景设计,会在必要时“多想一会儿”,以获得更稳定、更精确的结果。

具体来看,Instant是最常用的模型,据官方介绍GPT-5.1 Instant

现在默认会变得更加亲切自然也更具对话性,官方定位是是“更温暖、更智能、更听话”。官方透露,一些早期测试表明,它在保证实用性的同时,常常会以更轻松有趣的回答给用户带来惊喜。

比如,面对同一个比较感性的问题,两个版本的回答语气现在已经截然不同了。当你同时向两个版本的模型发送:我感到压力很大,有什么放松技巧吗?

GPT-5的回答一如既往的冷冰冰和理性,直接列出一些解压建议。

但GPT-5.1则会先表示对你的关心:“我理解你,这是完全正常的,尤其是你最近事情这么多,这里有一些解压的方法,你可以根据你感受到的压力类型来选择……”,然后再提出建议,更像是一个朋友的语气。

除了更会聊天,新模型的指令遵循能力也变强了。相信大家都遇到过,要求GPT-5用固定字数回答问题,但是它就是不听指令依然输出一大堆的情况吧?官方表示,

现在GPT-5.1再也不会“已读乱回”了。让它用6个字回答问题,模型就会严守规矩,所有问题都在老老实实用六个字回复。

除了更有人情味之外,GPT-5.1 Instant还首次运用了自适应推理技术,在回答更具挑战性的问题前,能够自主决定何时进行思考,从而在保持快速响应的同时,给出更全面、更准确的答案。

不同于Instant,GPT-5.1 Thinking的定位是高级推理模型,核心改进是能更精准地根据问题调整思考时间,在处理复杂问题时投入更多时间以求能给出更详尽的答案,而在处理简单问题时则更快响应,做到效率的动态调整。

通过官方给出的图表我们可以看到,GPT-5.1的思考时间变化比GPT-5更加动态。目前GPT-5.1最快的任务上的速度大约快GPT-5的两倍,而在速度最慢的任务上的速度则会比GPT-5多耗时71%。

此外,更新之后模型回答的表述也会更清晰,减少了专业术语的使用,尤其适用于工作中的复杂任

务和解释技术概念。Thinking模型的默认语气也会像Instant一样,更加温暖和富有同理心。

GPT变得更像人了

GPT-5.1的另一个核心更新,是对语言风格和对话个性进行了全面改造。OpenAI在新版本中内置了很多新的聊天风格,比如Professional(专业)、Friendly(亲切)、Candid(坦率)、Quirky(俏皮)、Efficient(高效)等,用户可以像切换滤镜一样快速选择。

同时,现在用户还可以更精细地控制模型的回复方式,新版本允许用户直接在个性化设置中调整ChatGPT的各项特性,包括回复的简洁程度、亲切程度、易读性以及表情符号的使用频率。

随着AI走向更广泛的应用场景,大多数用户在与AI交流时,都更希望得到一种具有人格温度的回应。此次GPT-5.1汲取了网友们的众多反馈,通过在训练中引入风格化标签和强化学习反馈,让模型学会了在不同语境中调整语气、

节奏与情绪表达,从而让AI变得更像人而不是冷冰冰的工具。

OpenAI还同步发布了GPT-5.1的System Card附录,从安全基准来看,整体安全性能与GPT-5持平。新版本的安全评估范围更大,新增了针对“心理健康”和“情感依赖”的专项测试。在抗越狱能力上,GPT-5.1的表现也比GPT-5要更优秀。

最后,在使用上,GPT-5.1 Auto将成为新的“默认选项”。用户不用纠结这个任务是该用Instant还是Thinking了,GPT-5.1 Auto会像一个总指挥,自动分析请求然后分配给最适合的模型。当然,用户也可以不依赖Auto智能功能,自己手动选择想用哪个模型。

OpenAI表示,新模型将在本周内逐步向所有用户免费开放,依旧是首先提供付费用户。旧版本GPT-5将继续向付费用户开放三个月,等过渡期结束才会下架。

同时官方表示本周晚些时候,他们也会将GPT-5.1 Instant和GPT-5.1 Thinking都引入API。