



人工智能与信息社会

人工智能与人类社会未来：挑战：技术视角

陈斌 北京大学 gischen@pku.edu.cn

挑战：技术视角

- › 机器人会因程序出错“发疯”吗？
- › 机器人会不会和主人有一样的脾气？
- › 能用人类的伦理道德约束机器人吗？



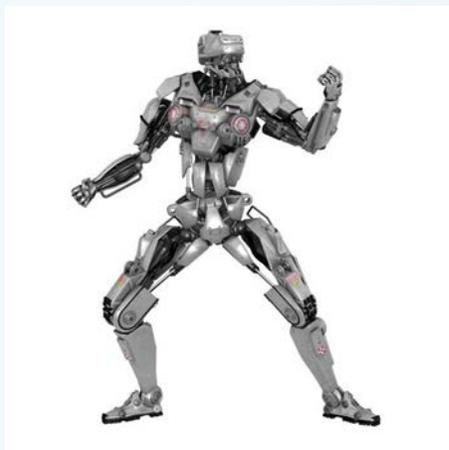
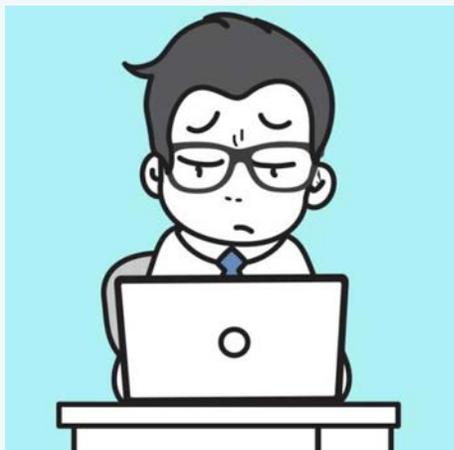
挑战：技术视角

- 挑战之一是能否保证人工智能的应用开发被用于正确的目标。如果用于错误目标，由于人工智能的强大能力，其产生的负面效果可能是快速而大规模的。



挑战：技术视角

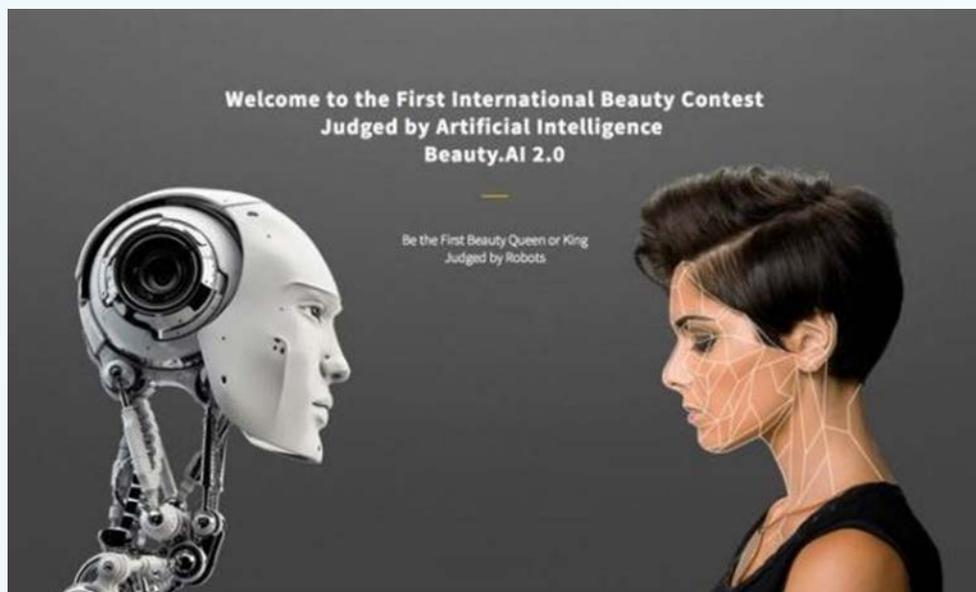
- 挑战之二是即使人工智能是为了合理的目标而开发的，但智能系统开发时存在严重的缺陷（技术人员能力不足或者偶然的疏忽），会产生不可预测的后果。



程序员会出错，机器人当然也会出错

挑战：技术视角

- › 例如在深度学习算法中，人无法知道机器“思考”的具体过程，而且机器学习的数据来源和质量可能存在问题，人工智能可能会被教坏。

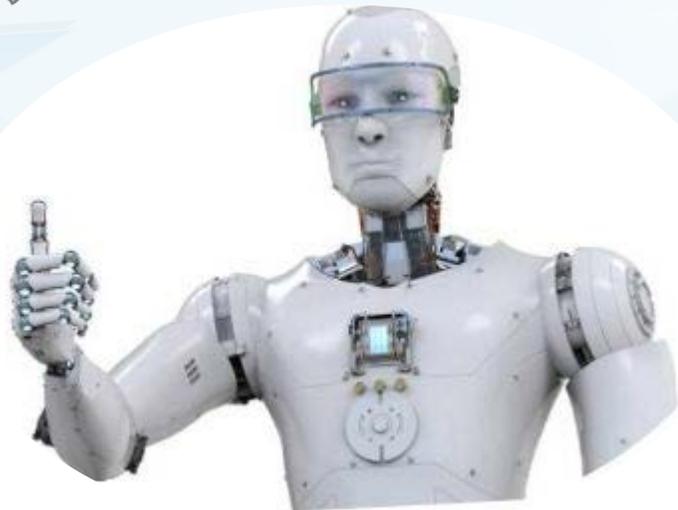


挑战：技术视角

怎么办？

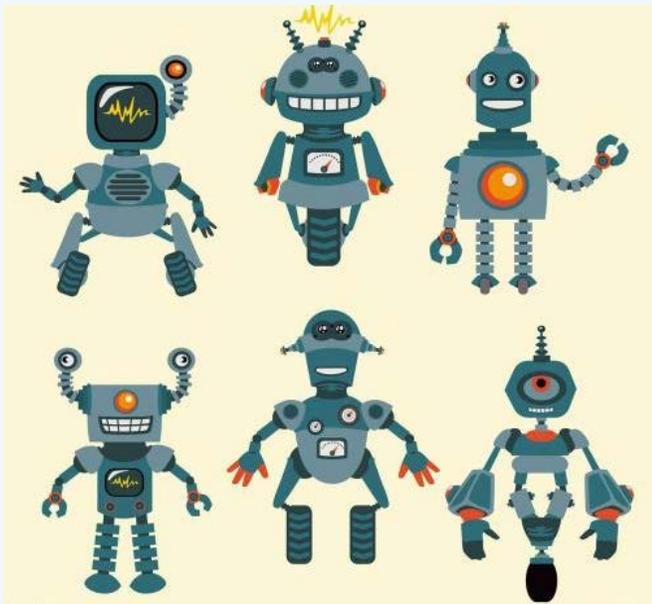
检测风险程序？

人工智能安全管理系统？



挑战：技术视角

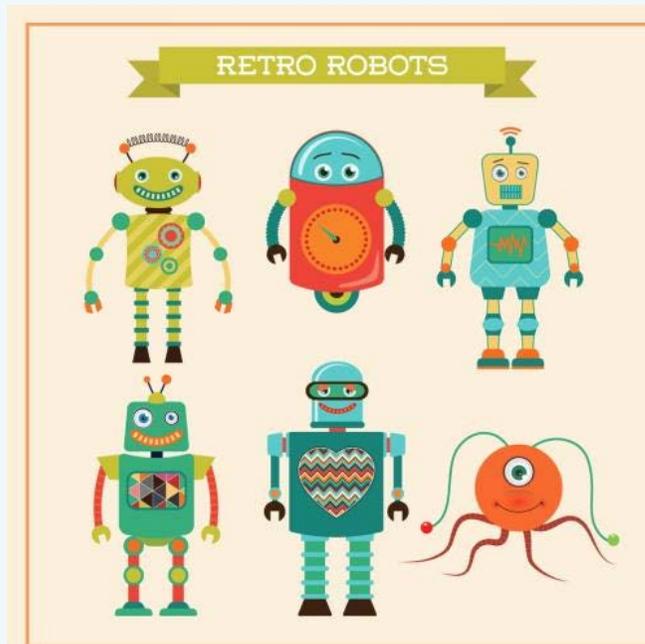
- 挑战之三是人工智能设计者在制作机器人时，当然会将自己的想法加入到机器人的思维系统中，从而使得很多人工智能应用没法同时使所有用户满意。



比如说，如果绝大多数机器人由男性制作，那么这些智能机器人就不一定懂女人心了。

挑战：技术视角

技术人员
多元化!



- › 人工智能研发者的多元化有助于满足不同人群的需求。

挑战：技术视角

- 挑战之四是人工智能毫无疑问具有强大的能力，但是终端用户不一定具有充分地使用该应用的知识 and 技能。如何对人工智能的服务群体进行指引和监管，将是一大难题。



挑战：技术视角

- › 挑战之五就是如何让机器人们能和我们一样遵守伦理道德？



机器人们都是真善美的化身？

挑战：技术视角

› 我们能否在技术上融入伦理道德？

也许机器人们对友情，爱情，亲情并不能完全理解，那就交给机器学习吧！



世界这么大，
我也想去看看

挑战：技术视角

- › 可以把人类的道德准则通过**算法**转化为人工智能的行为准则，让系统做出人类认为正确的道德决策，只是人工智能的道德判断比人类更理性、客观。



挑战：技术视角

› 总结：

人工智能系统当然会出错，会出现单一化、难操控、难预测等各种问题，这都需要我们来不断解决、不断优化，创造一个更加智能的时代！

