任务**1**：挑选一个要调查的技术案例研究。描述该技术、目的、范围、需求和好处。（**500 words**）

- 技术目的：该技术的目标/目的
  - 强化Ai的技术,解决人工智能问题，以实现广泛的社会效益，特别是在人工智能技术独特适合的应用领域，例如推进科学发展以及应对气候和可持续性问题,
  - 围棋的变化（10^170）大于宇宙中原子数（10^80）所以是一个巨大的舞台，我们可以在这里深度训练和研究还有实践和验证机器学习和神经网络的技术
  - 指导人类棋手更近一步。提升人类的潜在能力，提升自己的实力
  - 探索未知的领域，从而开阔眼界，这本身是一件好事，人工智能就好像打开了一扇'天窗'，让我们可以进一步挖掘围棋所蕴含的广阔天地。"罗超毅说，过去人们如何学围棋、下围棋、享受围棋，现在其实并没有太大改变，无非就是多了人工智能，"可我们下棋的意义，从来都不是为了去战胜人工智能。

- 范围（背景、时间、地点、具体使用情况）
  - 面向职业选手和初学者。
  - 可以用来进行入门的围棋教学和进价的围棋水平强化
  - 任意时间任意地点，好处是可以随时打开软件进行学习，不需要请围棋老师。
  - 可以帮助职业选手或者想要更近一步的玩家复盘。给出整句游戏中的失误和胜率曲线。帮助选手了解不同的下发分支。

- 需求:该技术的动机？为什么需要它？

长期以来，人工智能一直被用来学习和理解如何更好地下棋。但是最近，考虑到他们过去的比赛历史，它也被成功地用于检测一些参赛者是否打得比他们应该做的更好。

尽管游戏可能不再像以前那样，但对于下棋的人来说，有很多好处。这些包括提高智力、同理心、记忆力、解决问题的能力和创造能力。围棋带来的认知和理论挑战使这些能力得以发展。因为大脑的许多部分都在协同工作，所以棋手能够保留所有这些好处。由于国际象棋对大脑的要求，科学家们还发现它可以作为痴呆症的保护因素。

此外，计算机最初是为了在围棋比赛中击败人类而制造的。现在，这些引擎对用户来说变得更具协作性和吸引力。康奈尔大学计算机科学教授乔恩·克莱因伯格（ Jon Kleinberg ）建立了一个像人类一样下棋的国际象棋引擎。这让围棋选手有更好的国际象棋体验，研究人员一直在深入了解人类在不同级别的国际象棋中犯了什么错误。这为他们提供了数据，向用户展示如何改进他们的游戏以成为更好的棋手。在某种程度上，像这样的举措一直在重振吸引许多人的游戏的美感和艺术。以色列大学研究机器学习驱动的围棋引擎的研究员 Eli David表示，"与其让计算机围棋变得更强大，让人类变得更糟，我们可以专注于将国际象棋作为一种游戏形式的艺术。" 允许人类和机器之间的这种共生关系，游戏带来的情感和技术体验可以重新焕发活力，并且在许多情况下得到加强。

任务**2**：考虑你的用户和利益相关者。把他们都列出来，描述他们的属性和特征、需求和技能，为什么他们被认为是用户或利益相关者。**(500 words)**

- 属性:如年龄组（青少年等），性别（如大部分是女性等）。
- 特质：（只有在有具体的相关特质时才提及），如忙、时间少、外向等。
- 需求/目标：他们与这项技术互动的具体需求/目标是什么，如果没有，就说没有（他们可能不是直接用户，但他们可能会受到部署的技术的积极/消极影响）。
- 技能:对技术的理解（如高技术专家，或对技术的低接触）。
- 利益相关者
    - 直接:这些使用技术和/或技术用于它们。
    - 间接：这些感受技术的影响。
    - 特殊人群:范围广泛的类别，包括那些不能使用或选择不使用该技术的人，以及可能对其部署感兴趣的组织，如宣传团体或政府。

任务**3**：使用价值敏感的设计，围绕技术中的人类价值进行概念调查。从你的可能的用户/利益相关者名单中挑选出三个主要的用户/利益相关者（每个小组成员一个），进行想象的角色扮演，讨论技术对他们可能产生的影响，在每个伦理框架下。理想情况下，在进行下一步的研究之前执行这项任务。你可以使用判断游戏（可打印）作为工具，从利益相关者的角度审查这些技术。记录你的讨论（例如拍照）并总结概念调查。（**1000 words**）

- 3个利益相关者:
    - 开发围棋AI的公司和相关技术人员
    - 围棋职业选手
    - 接受AI的围棋爱好者
- 三个场景：
    - 老师对学生进行围棋教学的场景
        - 技术人员：安全性和隐私性】【4】-- 我们设计的AI系统应该既安全又尊重隐私。因为一旦私密对局，或者老师使用AI教学的细节被外泄，可能会导致对老师未来比赛的不利。在比如网络在线AI教学平台。如果不注重安全性和隐私性,也有可能会导致老师和学生的个人信息外泄。后果是需要开发这项技术的公司负责的。
        - 职业选手：包容性】【4】-- 在职业选手看来，ai应该可以针对不同水平的学生给出不同的教学意见，有利于老师进行针对性教学，并且ai给出的建议应该更加多元化，避免了职业选手可能无法有精力给出更多解法，单一化教学。并且职业选手对于一些高级的棋阵无法快速给出解法，效率不如ai。
        - 爱好者：【可靠性】【5】-- 一些爱好者，例如学生，在被教学的时候，可以更明确的看出这一步棋对后续胜率的影响，以及可以通过AI对每个棋盘位置所预测出的胜率来让学生更好的理解为什么要下这一步棋，AI就类似于一个标准答案，可以避免学生学习到错误的思路和知识

- 观看围棋比赛时候的场景
  - 技术人员：【可靠和安全】【5】-- 在大众观看人类比赛的时候，我们的技术可以提供实时的胜率曲线，以及帮助解说员分析当前棋局可能的后续分支。在这一点是我们的AI技术是很可靠的。以及我们必须确保在赛事进行中我们的技术能够可靠且安全地运行。还可以为不同的用户提供个性化定制服务。在大家在线观看比赛的时候，可以自由的自己尝试下棋查看后续的胜率。
  - 职业选手：【公平】【1】-- 职业选手可能会担心有选手利用ai技术来进行作弊，从而影响比赛的公平性。不利于比赛的观赏性，也会影响了比赛的公平性，从而会打击大众对于围棋的积极性。或者有不法人员利用ai来预测比赛的胜率从而牟利。
  - 爱好者：【透明度】【4】-- 玩家可以实时看到Ai所推荐的下一个最高胜率的落子位置，同时也可以通过胜率曲线和其他落子位置的胜率来知道为什么这个点是AI所建议的，虽然看不到具体的理由，但是也让爱好者在看比赛的时候更加的了解。

- 玩家与AI下棋的场景
  - 技术人员：【包容】【5】-- 我们的AI围棋技术必须要满足范围广泛的人类需求和体验。具体来说需要为不同的用户提供不同的难度选项。以及可能的对于残障人士的额外优化与个人定制化。确保所有人群都能公平的使用我们的技术。
  - 职业选手：【可靠性】【3】-- 从职业选手的角度，可以利用 ai 来进行个性化训练提升自己的技术从而在比赛中取得更好的成绩，另外职业选手可以通过ai复盘自己失利的局来找出自己的不足，职业选手在平时复盘中可能无法全面找出对手的弱点,同时也能通过ai分析对手的弱点，从而进一步在职业大赛中获得优势。
  - 爱好者：【包容】【4】-- 爱好者和AI下棋来训练自己，AI并不一定都是不可战胜的难度，可以选择从基础难度来慢慢训练和提升自己，找到最适合自己水平的AI，这就展现出来AI强大的包容性，可以包容各种水平的围棋爱好者，这是真人所不能比拟的

任务**4**：对类似技术的使用情况进行研究。挑选两个利益相关者讨论过的与该技术有关的问题或担忧，其中可能包括隐私和安全、可靠性和安全性、透明度、包容性和问责制的道德原则。用类似或相关的现实世界技术的例子，对其影响进行批判性分析。**(1000 words)**

- 对于人工智能来说，1997 年是历史性的一年。IBM的超级计算机“深蓝”（Deep Blue）经过六场比赛，成为世界上首个击败世界象棋冠军的机器。
- 全世界下象棋的人过亿。它是人类智慧的结晶，玩这种游戏需要有一定的策略、远见和逻辑等。因此，象棋对于衡量人工智能发展程度具有重要意义。
- 看着象棋这类游戏，我们一般会说，“当然，电脑确实下象棋很厉害，因为它的游戏规则、动作以及目标都界定明确。”这是一个你了解所有信息的约束性问题。然而，尽管简化了很多，国际象棋仍然是一个极其复杂的游戏，这也是为什么我们要花50年的时间才最终打败世界象棋冠军的原因。
- 虽然，计算机已经主宰了象棋，让人类惴惴不安。但另一方面，它也让更多人对国际象棋产生兴趣。不像30年前，当卡斯帕罗夫对弈卡尔波夫（Anatoly Karpov，俄罗斯国际象棋大师）的时候，没人敢对我们的走棋指手画脚，哪怕我们出现了失误。现在，你可以一边看棋，一边听着机器的解说。所以，从某种意义上来说，机器让国际象棋变得更加大众化了。现在他们可以跟上对弈，可以理解国际象棋的语言。因为有人工智能充当翻译的媒介。
- 如果我们开发出强大的国际象棋机器，那么国际象棋对弈就会变得乏味——会有大量平局出现、太多技巧或者一场对弈有走1800或1900多步最后陷入死局等等——抱有这样的想法其实是错误的。AlphaZero完全相反。与我而言，这更像是一种互补，因为你找到了一个更强大的对手！AlphaZero会牺牲一些棋子来换取更有效的进攻。这不是什么创新，而是AlphaZero学会了这种模式，知道计算其中的几率。但也正是这些让国际象棋充满魅力。
- 马格努斯·卡尔森（Magnus Carlsen，现任国家象棋世界冠军）曾说，他研究过AlphaZero的对弈，他还发现了其中的一些元素和联系。他说，有些走棋他设想过，但未曾真正实践过；现在我们可以知道这些假设的最终结果。
- 人们总说，我们得建立符合道德规范的人工智能。一派胡言。人类才是邪恶的主谋。问题不是出在人工智能身上。问题出在人类使用新的技术去伤害他们的同胞这里。
- 人工智能就好比一面镜子，它照出了好，也照出了坏。我们必须正视这个问题，然后弄明白如何解决这个问题，而不是整天抱怨说：“哦天呐！我们创造的人工智能总有一天会超越我们自己。”不知怎的，我们已然陷入两个极端。人工智能不是魔杖，也不是终结者。它也不是乌托邦或反乌托邦的预兆。人工智能就是一个工具。没错，它是一个很特别的工具，因为它可以扩大我们的思维，但它终究仍只是一个工具。但不幸的是，不管是在自由世界之内还是之外，我们有太多的政治问题，如果不能正确使用人工智能，局面只会更糟。
- 我们学到了很多经验，其中之一就是多种角度看待复杂问题。举个例子，在国际象棋中，人类的下棋方式通常是基于模式识别和直觉，而机器则是通过密集检索数百万乃至数十亿的可能性。这些方法其实彼此互补。国际象棋中如此，许多现实世界中的问题同样如此，电脑和人脑的结合都要胜过其中任何一方。例如，我们肯定不希望计算机全权接管病患的诊断和治疗，因为病情诊断涉及大量无法数据化的信息，但是在提供诊疗建议上，计算机系统还是非常有用的。
- 我们现在采用的是先进的基于人工神经网络的系统——有点像黑箱——它们并不怎么擅长解释自己为何要给出某项建议。但是如果它自己都解释不清原因，又让人们怎么相信这个系统的建议呢？况且，未知的神经网络系统拥有百万参数，极其复杂。不过，处理部分问题还是可以照着给出的优秀范例来训练系统，

> 这在医疗保健领域，计算机进行诊断或给出治疗建议时尤为明显。如果能给出合理解释的话，我想我们开发出来的机器在医生诊断时或许会有更大的发言权。

马斯克和比尔盖茨的警告

对于机器人及人工智能的恐慌，不仅来源于科幻电影。

特斯拉董事长、首席执行官马斯克说："开发人工智能，相当于在召唤恶魔。我认为我们应该非常小心谨慎，如果让我猜测什么最有可能对我们的生存造成威胁，我会说可能是人工智能。我越来越倾向于人工智能应该被监管这个观点，包括国家层面和国际层面上的，以确保我们没有在干蠢事。"

比尔盖茨也表示："我是对超级智能的发展保持担忧的人之一。目前，那些在我们的位置替代我们劳动、并完成许多任务的机器并不是超级智能。如果我们能够很好地进行管理，像这样的发展会变得越来越好。然而几十年之后，智能会带来新的问题。我十分同意马斯克的观点，但不明白为什么目前人们对这个问题并不担心。"

马斯克还出资1000万美元给一个致力于人工智能领域安全性研究的基金。

据了解，目前国际上达成共识的机器人三原则是在1940年由被誉为"机器人学之父"的科幻作家阿西莫夫所提出的：

第一条，机器人不得伤害人类，或看到人类受到伤害而袖手旁观。

第二条，机器人必须服从人类的命令，除非这条命令与第一条相矛盾。

第三条，机器人必须保护自己，除非这种保护与以上两条相矛盾。

冯华山表示，如果违背"机器人三原则"，将人工智能应用到战争中，将带来恐怖的结果。更重要的是如果机器人在执行自我学习的过程中失去或脱离人的监控，按照程序自行排列、演进，那么就有可能形成超越人类理解的内部逻辑，从而出现缺乏约束、失控、反人类的可能性。所以，坚守原则和国家的规定对人工智能使用的安全性而言至关重要。

人工智能伦理安全如何把控？

从全球范围来看，美国政府较早关注到这一问题，2016年在白宫文件中呼吁开展研究，并在《美国国家人工智能研究与发展战略规划》中明确了提高公平、可解释性，符合伦理道德的设计，确保系统安全可靠等要求。

2017年，霍金、马斯克、DeepMind创始人戴密斯等全球数千名专家和企业代表，签署有关伦理的阿西洛马人工智能23条原则，呼吁业界遵守，共同保障人类未来的利益和安全。

任务**5**：将你的发现与你最初的概念性调查进行比较。你能从这个练习中得出什么结论？**(500 words)**

- 举例一到两个一致或不一致的地方
  - 不一致
  - 一致
    - 问责制
    - 公平
    - 预想的可靠与安全和实际中的不一样
- 调查结果是否有互补性
  - 是的
    - 包容性所带出的问题
- 最后总结得出的结论。

通过上述两个case study， 发现在现实中的事情与之前在Judgment Call中的有很相似的地方

在在比case study I中，可靠和安全价值原则出现了我们之前没有想到的情况，AI伤害了对手的手指，造成了骨折这种严重的伤害，伤害人的AI是不可接受的,这违反了1940年由被誉为“机器人学之父”的科幻作家阿西莫夫所提出的机器人三原则:

第一条，机器人不得伤害人类，或看到人类受到伤害而袖手旁观。

第二条，机器人必须服从人类的命令，除非这条命令与第一条相矛盾。

第三条，机器人必须保护自己，除非这种保护与以上两条相矛盾。

尽管这些行为可能不是机器人有意为之，但是此行为仍然明显的与第一条相对违背，而这些原则应该是制造机器人的时候所遵循的。

在比case study II中如说公平，现实中有职业选手使用了AI技术来进行作弊，这与之前所担忧的是一样的，ai在这些行业无懈可击的表现给了一些选手通过违规来提高胜率的方式，可以想像如果没有ai，那么围棋之类的运动几乎不可能在世界级的比赛中作弊，然而AI的出现却让人开始担忧围棋比赛的公平性。同时，围棋AI过于强大的包容性也带来了我们没有预想到的问题，包容了一些可以被不道德的利用的方面，这些也是我们在设计AI之初所需要考虑到的一些情况。

总体来说，在AI设计之初，需要考虑更多的情况，如果有一些需要通过AI来操控的机器臂或者其他直接与人类接触的机械产品，都需要在任何情况下注意人的安全，同时也需要对AI的用途进行一些限制，防止AI被用来使用在一些非法的场景，来危害社会。

任务**6**：提供你的建议。你应该采用和部署这项技术吗？如果是这样，你是否需要进行一些改变或以特定方式进行调整？**(500 words)**

- 综合所有的道德风险和对利益相关者的影响来推论是否认同。
- 调整部分：简单列举一点需要特别注意的事项就行。