

AI/ADM Case study : Go AI Technology

Wenqian Zhang, Ziyi Liang, Zhangpeng Chen

University of New South Wales

1 Introduction

围棋是一种策略棋类，使用格状棋盘及黑白二色棋子进行对弈。它起源于中国，是世界上最古老的棋类运动之一。也被认为是人类棋牌游戏的巅峰。按照状态空间复杂度来衡量，围棋的变化可以达到10的172次方。按照博弈树的复杂度，是10的300次方 远远大于宇宙中原子数 10^{80} 。从公元前23世纪到如今没有一盘重复的对局，这也是它令人着迷的地方。至今全世界还有上千万的围棋爱好者，就单是在中国每年新接触围棋的儿童青少年大概能达到两三百万人(Xinhua 2020)。

围棋更为准确的定义是完全信息博弈游戏，指每一参与者都拥有所有其他参与者的特征、策略及得益函数等方面的准确信息的博弈。截止2015年，全世界6000多种完全信息博弈游戏中，只有围棋程序无法抗衡相关领域的职业选手。而在2015年10月，AlphaGo成为第一个无需让子即可在19路棋盘上击败围棋职业棋手的电脑围棋程序。因此我们要调查的AI技术是这个依靠在近年来发展起来的机器学习和神经网络技术诞生的AlphaGo围棋人工智能。

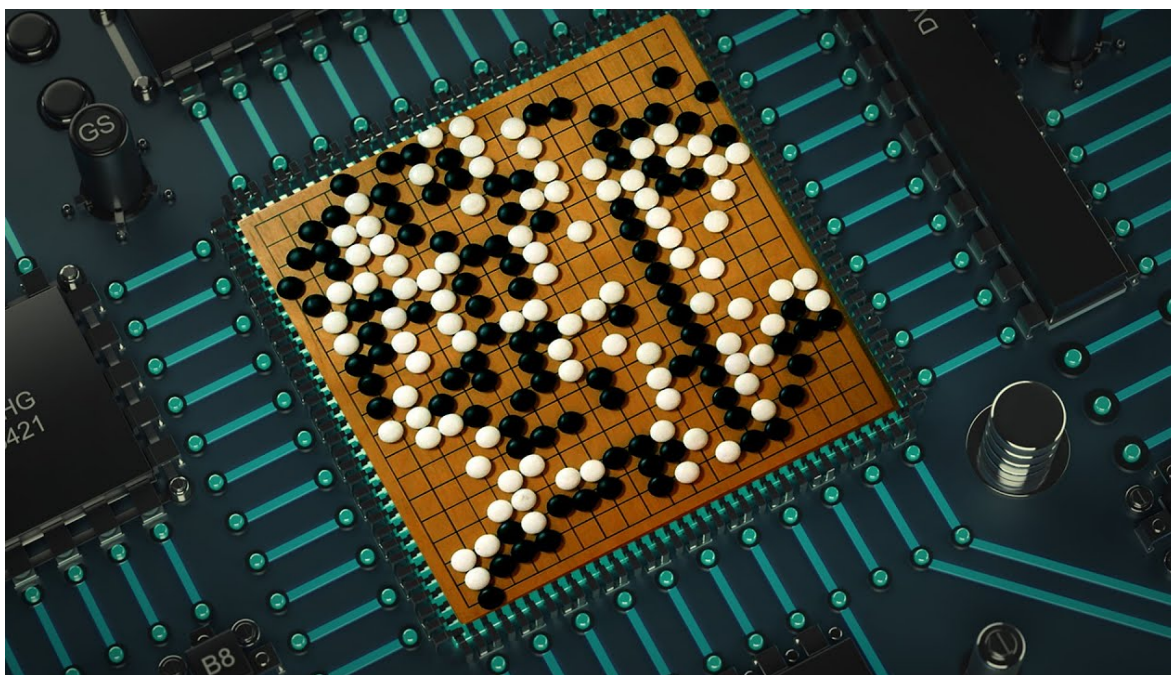


Figure 1. Abstract image of the fusion of Go and computer technology

2 Purpose, Benefits

AlphaGo是由英国伦敦的Google DeepMind团队开发的人工智能围棋软件。它可以像人类的大脑一样自发学习进行训练，以千万亿次的自我博弈训练提高下棋实力。这项技术的目的有：

1. 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。
2. 由于围棋棋子能产生近乎无穷的变化和分支。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

3. 指导人类棋手更近一步。提升人类的潜在能力。
4. 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。

AlphaGo这项技术涉及和能影响的范围有：

1. 面向职业选手和初学者。
2. 可以用来进行入门的围棋教学和进价的围棋水平强化
3. 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。
4. 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法会产生什么样的分支和结果。

对人工智能围棋这项技术需求来自于与人们希望能有一种更好的方式来学习和理解如何更好地下棋。比如，这项技术可以帮助用于检测一些参赛者是否能打得比他们之前更好，包括之前犯了什么错误。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。此外，AlphaGo最初是为了在围棋比赛中击败人类而制造的。现在，这个技术对用户来说变得更具协作性和吸引力。允许人类和机器之间的这种共生关系，围棋游戏带来的情感和技术体验可以重新焕发活力。

3 Stakeholders

能被人工智能围棋技术影响的人群可以大致分为以下几类：

1. 利益直接相关者，例如开发围棋AI的公司和相关技术人员。
2. 直接被这些技术影响的围棋职业选手，赛事主办方，围棋协会。
3. 数量众多的围棋爱好者，也是这项技术的直接利益相关者。

3.1. The first category of stakeholders

这类人群是高技术专家，对于AI技术的接受度极高，也是围棋AI技术的创造者和改进者。他们的需求和目标是希望围棋AI能够被更多人接触和互动。希望获得更多数据用于改进自己的技术以及发表相关论文促进整个学界的发展。可以预料的是，他们会受到该项技术部署的积极影响，包括经济上的收入以及扩大从业人员带来的对整个行业的好处。以及未来依靠技术发展，可能的，将这项技术移植和转移到其他人类复杂任务领域，为AI技术整体带来更多可能性。

3.2. The second category of stakeholders

如果说第一类人群是生产者，那么第二类人群就是最直接的消费者或者说是使用者了。尤其是围棋职业选手，他们靠着围棋游戏生存，很多人从小就开始学习围棋。很多人在读完初中后就直接进入国家棋院，把一生都奉献给围棋，立志要成为职业棋手。他们对于这项技术的潜在需求在于可以帮助他们更深的理解自己的围棋策略缺陷和陪伴他们

日以夜继的训练。这类人群对于技术的理解不高，对于技术的接受度也参差不齐。例如有些选手对于这项技术非常抵触，认为围棋AI可能毁掉这个游戏的乐趣。认为AI使得这个历史悠久的游戏失去了未来，一切都是冰冷冷的数据，都是可被计算的。大部分棋手表达了自己对这项技术的担忧：害怕会导致自己失去工作，毕竟人类再怎么努力也不可能赢过每秒计算上亿次的计算机。而反对者则认为这项技术可以让人类的围棋进入一个新时代，在平时的对局中，有时候人类下出了AI没有推荐的手段之后，胜率反而会上升的情况。这也证明了AI其实也有计算的“盲点”，AI看不到的好棋，其实人类是可以看到的。棋手们现在努力寻找围棋AI找不到的盲点。许多人在挑战AI的过程中都体验到了快乐。总结一下，AI围棋技术的部署对这类人群会同时产生积极和消极的双重影响。

3.3. The third category of stakeholders

这类人群的数量众多，其中年龄跨度甚至可以从6岁到90岁。而他们对这项技术的看法也有好有坏。其中按照对于AI技术的接受度可以粗略的分为两组：

1. 体验过AI好处的人，也就是这项技术的使用者。对于这组人群，他们的目的和需求和第二类人群大致一致，都认为AI可以帮助自己的围棋技术更进一步。而且对于他们来说，围棋AI作为“教师”的角色更为突出。
2. 这项技术的抵抗者，也就是拒绝使用围棋AI技术的人们。这组人群更多是年龄较大的人群，不仅仅对于围棋AI的接受度低，甚至对于现如今的其他互联网技术使用度也很低。他们倾向于认为围棋AI会导致围棋失去“人类博弈的乐趣”。由于不会使用新技术，所以AI围棋的部署会对他们产生消极的影响。

4 Value Sensitive Design

我们将会使用“Judgment Call”这个游戏来帮助揭示AI围棋技术中的道德问题，预测社会影响。我们将分别角色扮演下面三组具体的利益相关者，以他们的角度撰写产品评论来思考技术可能会产生什么样的危害和影响：

1. 开发围棋AI的相关技术人员
2. 围棋职业选手
3. 接受AI的围棋爱好者

评论格式：【利益相关者】 - 【关注的道德原则】 - 【评级卡(1- 5星)】：具体评论
接下来是我们在三个场景中进行讨论的会议摘要：

4.1. Scenario I

围棋老师正在在线围棋对战平台上对学生进行围棋教学。该在线平台使用了围棋AI技术作为辅助。为不同用户提供了个性化服务。

- **AI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。在比如在线平台。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。后果是需要我们负责的。

- **职业选手 - 包容性 - 4:** 围棋AI技术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。这样的AI技术会对我们在教学上产生很大的帮助。
- **爱好者 - 可靠和安全 - 5:** 希望能在被教学的时候。通过围棋AI技术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。如果这项技术能帮助我们学生避免学习到错误的思路 and 知识，那么我们会很高兴。

4.2. Scenario II

一场国际围棋比赛正在进行中，参加的选手之前都使用了围棋AI作为训练工具。同时比赛解说们也在通过围棋AI技术辅助他们向全世界直播讲解这场比赛。

- **AI技术人员 - 可靠和安全 - 5:** 在大众观看人类比赛的时候，我们的技术应该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技术应该做到很可靠。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变化。
- **职业选手 - 公平和问责制 - 1:** 我们担忧的是可能会有选手利用围棋AI技术来进行作弊。如果出现这样的情况，那么不仅会影响比赛的公平性，而且也会不利于比赛的观赏性，从而会打击大众对于围棋的积极性。另一种情况是可能有不法人员利用AI来预测比赛的胜率从而牟利。而且选手担忧出现事故后应该由谁来负责。AI技术犯的错是否应该由使用者承担还是发明者。
- **爱好者 - 透明度 - 3:** 我们希望可以实时看到AI所推荐的下一个最高胜率的落子位置。同时希望被告知内容是如何个性化的。我们得到的AI辅助信息是否都一样，有没有因为收集我的行为而私自对用户体验进行个性化设置。例如我的水平很低虽然可能看不懂复杂的变化图，但还是希望围棋AI能给我提供所有的局势变化信息。这能让我在看比赛的时候更加的了解。

4.3. Scenario III

一名玩家正在与围棋AI进行博弈。这名玩家可能是任何人，例如职业选手或者是普通爱好者。

- **AI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。例如对视力障碍者进行语言播报。确保所有人群都能公平的使用我们的技术。
- **职业选手 - 包容 - 3:** 从职业选手的角度，希望可以利用围棋AI技术来进行个性化训练。例如这项技术能否收集自己的历史棋谱，用来定制化的设计博弈策略，从而帮助我们修复自己的技术缺陷。职业选手在平时的练习中可能无法全面找出对手的弱点。如果能围棋AI技术可以模仿对手的下棋风格，那么职业选手们会很支持这项技术的部署。
- **爱好者 - 包容 - 3:** 爱好者可能担忧AI都是不可战胜的难度。希望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难度。如果围棋AI技术

能包容各种水平的围棋爱好者。那么这是真人所不能比拟的，专属于AI围棋技术的优势。

4.4. Conceptual Investigation Summary

在三个不同的场景下经过代入不同的利益相关者围绕技术中的人类价值进行概念调查后，我们组对于部署围棋AI技术可能产生的有关问题或担忧得出了以下几点结论：

1. 大众最关注的价值原则可能会是技术的包容性。不希望这项技术成为小一部分人的专属。不希望这项技术成为一种特权，成为少数人剥削多数人的工具。同时也希望这项技术可以满足范围广泛的人类需求。为不同人群包括但不限于残障人士等提供良好的个性化体验。
2. 最直接的利益相关者，也就是职业棋手们，最关注的可能是这项技术的公平和问责制。他们害怕会有人利用这项技术在比赛中获得不正当的优势。
3. 围棋AI技术的可靠和安全以及安全性和隐私性也是一个被重点提到的价值准则。技术公司和相关技术人员们担心这项技术会为他们带来可能的法律上的风险。
4. 在总体上而言，利益相关者们对于这项技术持有正面的评价。在我们模拟的“Judgment Call”游戏中得分也很高。他们对于围棋AI技术的期望高于担忧。

5 Analysis of similar technologies

对于另一个完全信息博弈游戏，国际象棋来说，1997年是历史性的一年。这一IBM的超级计算机“Deep Blue”经过六场比赛，成为世界上首个击败世界象棋冠军的人工智能。本报告将对国际象棋AI技术的使用情况进行研究，并以此为围棋AI技术是否部署提供参考依据。

5.1. Case Study I

最近在莫斯科举办的一场国际象棋公开赛上，就发生了一起严重的机器人伤人事故：由电脑控制的机械臂因为操作速度过快将其对局的一名7岁男孩的手指压成了骨折。



Figure 2. A chess robot broke a 7-year-old boy's finger.

https://www.youtube.com/watch?v=jxS8GmV5hg&ab_channel=AllNews

根据俄罗斯新闻网 TASS(2022)的报道，现场这台执行下棋操作的机械臂出现了移动错误，可能是程序Bug。从现场视频来看机器人“吃掉”该参赛男孩的一枚棋子，并把自己的棋子放在空位上，但小男孩未等机器人完成动作便急忙回击。随后，机器人捏住男孩手指。工作人员急忙帮助小男孩将手指从机器人的“虎钳”中挣脱出来，并送往医院，但小男孩的手指骨折了。这段视频被上传到推特上广泛传播，男孩的父母已经联系当地的检察官办公室提起指控。

这毫无疑问违反了价值原则“可靠和安全”。不过见到详细应用程序前无法判断其具体原因。不过，首先可以排除人为设计机器人“惩戒”程序，因为机器人不能伤害人类是国际上机器人三原则的重要原则之一。很有可能小孩子的手指位置刚好是机器人计算获得下一步需要移动的棋子位置，在机器人形成动作轨迹指令后，小孩子恰巧将手指遮挡了这枚棋子。因为在机器人形成动作指令后，是不会再依据视觉做出判断。也就是说，机器人“以为”夹持的是棋子，而实际是小孩子的手指。这只是一种巧合。这一例子证明了现阶段AI技术是可能偏离人的完全控制和常规理解的。这些涉及到机器人伦理学，也同时对应了价值原则“问责制”，需要人们对机器人的学习演进过程严格监督和把控。因此人们对人工智能的应用风险仍需时刻保持高度警惕。

5.2. Case Study II

2022年9月4日，在美国圣路易斯举办的国际象棋Sinquefeld Cup的第三轮比赛中，世界排名第一的挪威棋手Magnus Carlsen意外输给了19岁美国小将Hans Niemann，结束了53场不败的记录，爆出了今年国际象棋棋坛的最大冷门。根据Davis(2022)的报告，Magnus指控Hans涉及利用国际象棋AI技术作弊，而且网络上也在流传Niemann可能使用了振动的无线肛门珠来指示他应该采取什么行动。

在Carlsen公开斥责Niemann后，Chess.com(2022)也发布了一份72页的调查报告，详细阐述了他们对尼曼此前作弊的调查结果。这份调查报告提供了大量基于数据分析的证据，依靠诸多作弊检测手段来判定是否作弊，包括将棋手的行棋和超级计算机的推荐进行对比，以及采纳了来自国际象棋公平竞争专家们的意见。报告认为，尼曼可能在截至2020年的100多场线上比赛中使用了非法作弊手段，其中包括一些有赢棋奖金的比赛，他的成绩在统计学上异乎寻常。

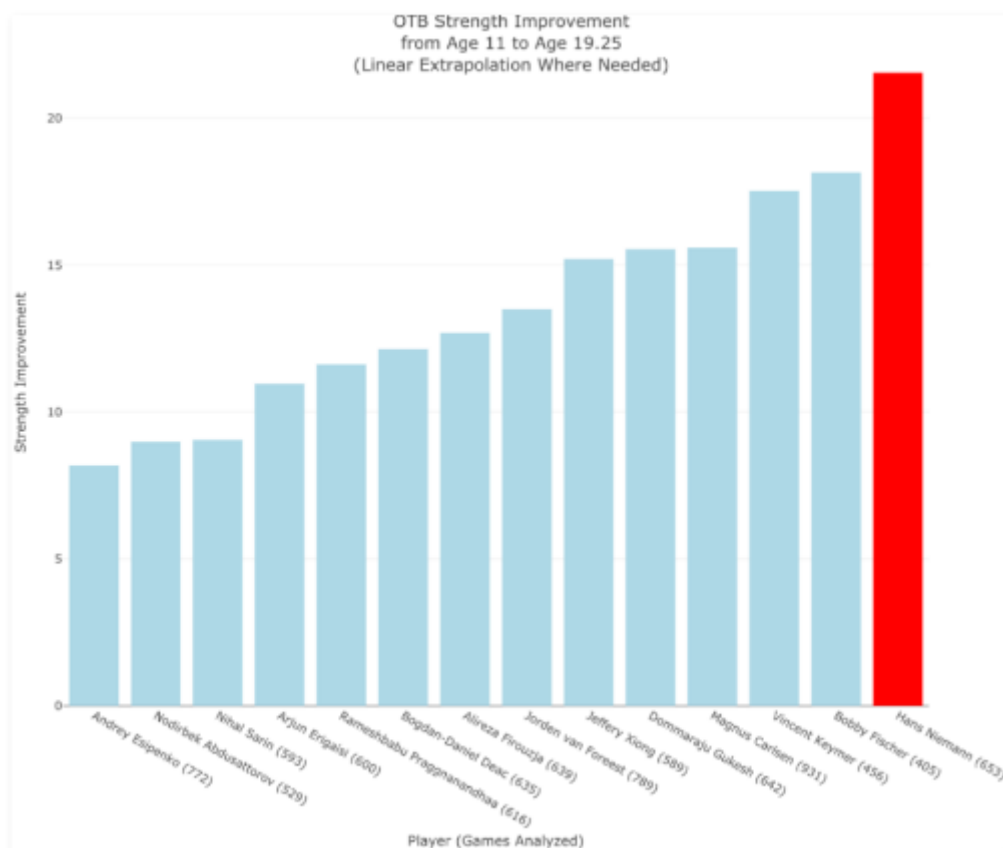


Figure 3. Chess.com(2022) releases heavyweight report accusing Niemann of possibly cheating hundreds of times

这次的事件毫无疑问涉及到职业选手最担心的价值原则“公平性”。选手私自使用国际象棋AI技术进行作弊对其他选手造成了真正的直接伤害，也对整个国际围棋运动带来了长期损害。未来的国际象棋比赛，只能用不断升级的高科技防作弊技术来防范棋手不断升级的高科技作弊技术，来维护这项运动的人类智力纯洁性。一旦棋手利用AI技术作弊成功，比赛就失去了意义。

整个国际象棋界应当重视AI作弊问题。它是对这项运动存在的威胁。赛事组织者应该严肃考虑增加安保措施和作弊检测手段。我们有不同的道德价值观。对于某些人来说，公平是一个非常重要的价值。当有人作弊并创造不公平的竞争环境时，我们认为这是非常不公平的。这虽然是人类的主观恶意带来的结果，但毫无疑问国际象棋AI技术是它们的帮凶。因此AI技术公司也更应当为此负责。在保证自己技术的“包容性”，也就是满足范围广泛的人类需求的同时也需要约束自己技术可能会被不道德利用的可能性。

6 Comparison between Analysis of Similar Technologies and Conceptual Investigation

首先可以发现，在Conceptual Investigation里总结的价值原则“可靠和安全”可能引发的问题都在类似技术的部署后发生了（Case Study I）。这证明我们扮演利益相关者推出来的担忧是有意义的。不过，在实际情况中涉及到的这些原则还是出现了我们之前没有想到的情况。我们之前是以相关技术人员们的视角担忧如果没有考虑“可靠和安全”价值可能会带来可能的法律上的风险。不过现实中，可能用户们是会更应该担心这些价值原则的。例如在Case Study I中国际象棋AI机器人伤害了小男孩的手指，造成了骨折这种严重的伤害。这是我们之前没有料想过的情况。这个实践经验可以给我们的概念性调查带来互补。也就是相关技术人员们不应该只从自己的视角考虑“可靠和安全”这个价值原则，只想着如何能让AI系统避免给自己带来法律风险。也应该为用户们在实际使用中考虑会不会对他们造成危害，AI技术能否在任何时刻都保证可靠且安全地运行。

我们还可以看到在Case Study II中有职业选手使用了国际象棋AI技术来进行作弊，这正涉及到在Conceptual Investigation中扮演职业选手时候所担忧的价值原则“公平性”。AI在这些行业无懈可击的表现给了一些选手通过作弊来提高胜率的方式。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和国际象棋比赛的公平性。同时，国际象棋AI技术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，这些也是我们在部署围棋AI技术之前需要考虑到的一些情况。

总体来说，部署围棋AI技术之前，需要考虑更多的情况。如果有一些需要通过围棋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，来危害社会和扰乱秩序。

7 Deployment Recommendations

在经过详细分析后，围棋ai技术应该可以被部署，不过需要做出一些调整。首先对于技术人员以及围棋爱好者来说，可靠以及安全性是他们最为关心的话题。机器人三原则虽然听起来似乎有道理，但许多论据已经证明了有些情况下它们是不充分的。所以要多采取采取边缘案例来避免，假设AI开发人员编写的代码可以检测其他玩家（假设是人类）是否跳过了机械臂的正确转动。假设机械臂和棋盘在障碍物后面，然后需要及时检测一个人翻过障碍物进入了机械臂所在的位置。夹具应该具有压力感应能力。这可能会提供反馈，即被挤压的东西与棋子的一致性不同。在某个阈值上，运行抓手的AI应该在假设抓到棋子以外的东西的情况下自动释放。这可以覆盖所有可能不正确地进入其抓握器的其他物体或生物元素。我们会在未来根据实践逐步添加更多的边缘案例来使我们的围棋ai更加的安全和可靠。

其次对于围棋职业选手来说，他们会更加关心是否会有人利用ai进行不正当的作弊行为，我们围棋ai的公平性，是他们最为关心的。例如许多大学已经在使用 Turnitin 等基于文本的抄袭检测器来检测是否学生的论文是由AI辅助撰写的。在未来如果我们的围棋ai有风险会被利用于作弊。我们在部署中应该增加更多类型的检测，将做出的动作与引擎推荐的动作进行比较并且将球员过去的表现与他们的历史实力进行比较。从而来判定他们是否使用ai进行作弊，来保护我们的Go AI不会被利用于不正当行为。另外与公平性最为相关的还有一点就是法律。在部署我们的AI技术时，应该推动立法者对于AI制定严密的法律来管理ai的各种用途。但是起草和颁布此类法律的努力是一个渐进的过程，我们会在接下来的实践当中慢慢完善。

Referance

Chess.com Team. (CHESScom. (2022, October 4). *Hans Niemann Report*.

Chess.com.

<https://www.chess.com/blog/CHESScom/hans-niemann-report>

Davis, S. (2022, September 26). *Magnus Carlsen accuses Hans Niemann of cheating at chess*, Sinuefield Cup. Insider; Insider.

<https://www.insider.com/magnus-carlsen-accuses-hans-niemann-of-cheating-at-chess-sinuefield-cup-2022-9>

TASS. (2022, July 28). *Moscow Chess Federation calls for strengthening protection after robot breaks boy's finger*. TASS; TASS.

<https://tass.com/sport/1486181>

Xinhua. (2020, January 11). *Chang Hao: The root of Chinese Go prosperity lies in the masses*. Sina.com.cn.

https://k.sina.com.cn/article_213815211_0cbe8fab02000tn7w.html#/

