网易数智 | 网易易盾

# AI加速游戏Secure
# 安全治理生态

网易易盾游戏安全指南

# 序言

在这个由数据驱动的时代，游戏行业正经历着前所未有的变革。技术的飞速发展不仅为玩家带来了更加丰富和真实的体验，也给游戏安全带来了新的挑战。外挂、作弊、黑灰产等安全问题，不仅侵蚀着游戏的公平性，更威胁着整个行业的可持续发展。

网易游戏安全团队深知，游戏安全是游戏体验的基石。为了保护每一款游戏的纯净性和玩家的合法权益，我们不断探索和创新，将人工智能技术应用于游戏安全治理中。《AI 加速游戏安全治理生态 - 网易游戏安全指南》是这一探索的成果，旨在分享我们在游戏安全领域的洞见和实践，为行业提供一份全面的安全治理方案。

本书首先分析了当前游戏安全面临的主要威胁，包括外挂软件的滥用、黑灰产业链的蔓延、账号安全的风险等。我们深入探讨了这些问题的成因和影响，以及传统安全措施面临的挑战。

接着，我们详细介绍了 AI 技术在游戏安全治理中的应用。从行为分析、异常检测到智能识别和自动响应，AI 技术正帮助我们以前所未有的速度和精度，识别和打击外挂行为。我们分享了网易游戏安全团队如何利用 AI 进行实时监控、模式识别和风险预测，构建了一个全面的游戏安全防护体系。

此外，本书还探讨了AI技术在打击黑灰产方面的潜力。通过深度学习和数据挖掘技术，我们能够追踪和分析黑灰产团伙的行为模式，揭露其运作机制，从而采取更有效的打击措施。

我们认识到，游戏安全不仅是技术问题，更是一个需要多方参与和合作的生态系统问题。因此，本书提出了构建游戏安全生态的策略，包括加强行业自律、推动立法保护、提高玩家安全意识等。

最后，我们展望了 AI 技术在游戏安全治理中的未来发展。随着技术的不断进步，我们相信 AI 将为游戏安全带来更多的可能性，为玩家创造一个更加安全、公正的游戏环境。

《AI 加速游戏安全治理生态 - 网易游戏安全指南》不仅是网易游戏安全团队的经验总结，也是对整个游戏行业的贡献。**我们希望通过《指南》，能够启发更多的思考和讨论，促进游戏安全领域的交流与合作，共同推动游戏行业的健康发展。**

在游戏的世界里，每一位玩家都值得享有公平竞技的机会和安全的游戏环境。让我们携手合作，利用 AI 技术，加速游戏安全治理，守护这片充满想象和乐趣的数字天地。

—— 网易易盾游戏安全团队

# 推荐语

"这几年非常感谢易盾技术团队和业务团队的辛勤工作和不懈支持。易盾技术团队开发的加固产品、反外挂措施以及风控产品不仅覆盖面广，兼容性强，而且在稳定性上也有出色的表现，这些都得到了我们趣加项目同事的高度认可和赞赏。
在提供这些高品质的技术解决方案的同时 ，易盾团队也提供了无微不至的支持。专业的技术服务保证了我们的业务流畅运行，同时也帮助我们在市场中建立了良好的声誉。易盾团队对于风险预防和安全保障的专业知识，使我们能够有效地防范各种安全威胁，确保了用户的安全和公司的利益。

我们期待与易盾团队继续保持紧密合作。为用户提供更为安全、稳定的服务。"

——FunPlus 游戏安全团队

"游戏安全就如同铠甲一般，守护着玩家，维持游戏环境的秩序。我们是这场冒险中的铸甲师，不断的提升，为玩家提供更全面的保护，与玩家一起搭建起更加坚实、健康、安全的游戏世界堡垒。"

—— KK 安全团队

"未来，借助 AI 大模型能力，游戏内容审核将更加智能化，不良内容过滤更为有效，从而更好地维护游戏生态环境。"

—— 贾彦龙 游族网络大数据负责人

"游戏黑灰产工作室不仅破坏了游戏内的经济生态平衡，同时也影响了玩家正常获取资源的公平性。易盾通过技术监测游戏规则的执行，为我们《宿命回响》《灌篮高手》《航海王启航》等游戏在对抗黑灰产工作室作弊行为中提供了极大的帮助，从而营造一个公平、健康的游戏环境，让玩家能够享受到纯粹的游戏乐趣。"

—— 陈永华 DeNA 中国 游戏技术总监

"《战意》一直将玩家的游戏体验放在首位，重视玩家反馈、优化玩家体验。特别是在游戏公平性，我们的运营团队本着绝不姑息在游戏中使用外挂、脚本等第三方非法软件玩家的原则，第一时间对扰乱游戏环境的相关玩家进行了封号处理。打造公平、健康、绿色的游戏对战环境是我们对玩家的首要承诺。"

——《战意》运营团队

"紫龙游戏在过去的经营中，坚持与侵犯知识产权的私服、破解支付协议、代充、游戏内打金、刷号工作室、游戏内聊天拉人等黑产灰产进行斗争，在这个过程中，黑灰产的人不断的变换方法，我们也随着这些变化不断调整策略进行分析和打击。在可以遇见的将来，由于这其中依然有很大的利润空间，这些黑灰产依然会不断的变换方法进行试探，我们仍将需要不断升级方法，把一些 AI 技术和合作伙伴的工具也用在里面，将他们对游戏的伤害，对玩家利益和我们厂商利益的侵蚀压缩在一个尽量小的空间，为玩家创造一个更加健康和谐的游戏环境。"

—— 侯志芳 紫龙游戏 CTO

# 目录

# PART. 01 Chapter

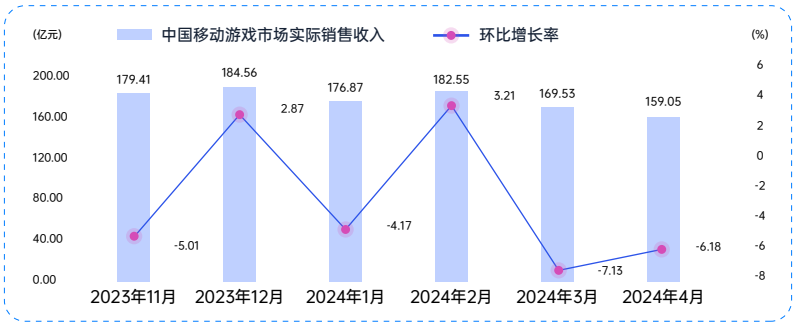
# 困局：上半年游戏产业发展概览

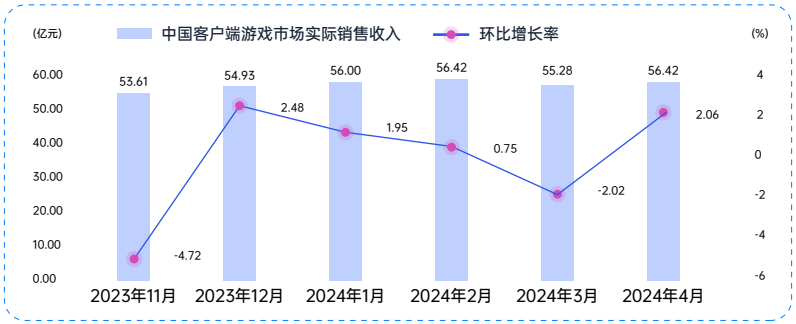
- 2024年上半年中国游戏市场现状
- 2024年上半年主要游戏安全风险挑战

# 2024年上半年中国游戏市场现状

**国内游戏市场波动性增长，移动游戏实消仍占主导地位**：根据中国音像与数字出版协会主管的中国游戏产业研究院的战略合作伙伴，伽马数据发布的《2024 年 1—3 月中国游戏产业季度报告》数据显示，中国游戏市场实际销售收入 726.38 亿元，同比增长 7.60%，环比下降 2.46%。

2024 年 4 月，中国客户端游戏市场实际销售收入达 56.42 亿元，环比上升 2.06%，同比上升 1.47%，中国移动游戏市场实际销售收入为 159.05 亿元，环比下降 6.18%，同比下降 1.93%。





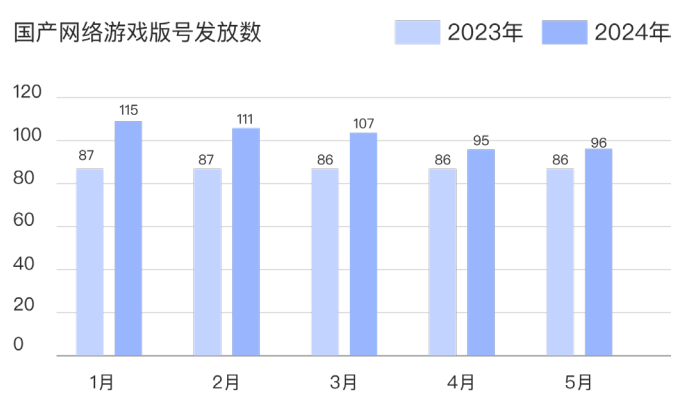

2024 年 4 月，中国自主研发游戏海外市场实际销售收入为 14.18 亿美元，尽管实际销售收入环比略有下降，但同比增长 7.03%，表明海外市场仍具有增长潜力。

从全球区域上来看，据全球游戏市场研究公司 Newzoo 数据。去年亚太、北美和欧洲地区位列游戏消费排名前三。其中亚太市场达到全球游戏市场 46% 的收入份额，以中国、日本、韩国、印度等区域庞大的游戏用户量因素贡献收入为主。同时，中东与非洲地区以及拉丁美洲地区虽占据市场份额不足，但显示出了良好的增长态势，是未来许多游戏厂商着眼挖掘的新兴市场。

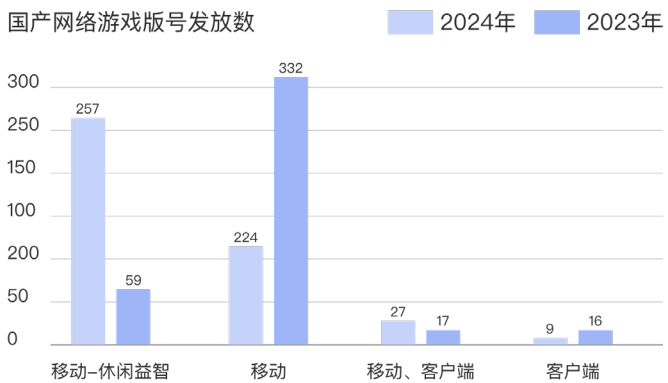
整体而言，2024 年上半年中国游戏行业整体呈现波动性增长，同时，游戏市场人口红利减弱，用户增速放缓，优良的游戏制作和玩家体验是新游迈进市场的决胜利器。

**游戏版号稳定核发，移动 - 休闲益智类发行量为同期 4 倍**：据国家新闻出版署公示结果，2024 年首批国产游戏版号发放数量达到了 115 个，这是近两年来单月发放量最高的一次。2024Q1 每月版号发放数均破百，后续数月虽略呈下降趋势，但月均发放数仍居于高位，高于往年同期平均水平。

截至 2024 年 5 月，国产网络游戏已累计获批 524 款，发放版号数较去年增长 21%。版号批次数量进一步稳定发行，同时各大游戏发行方响应活跃，向整个游戏产业发展



截至 5 月，单移动端版号发行数占 9 成。其中移动 - 休闲益智游戏产品获批 257 款，而 2023 年获批仅 59 款，版号发行数量较同期增长 4 倍。上半年小游戏发展势头依旧猛烈，游戏市场已从高投入、长周期的中重度产品转向发行效率更高、更灵活敏捷的轻量化游戏。随着市场的迅速扩张和增长，小游戏安全风险问题日渐瞩目。




| | 2024年 | 2023年 |
|---|---|---|
| 移动–休闲益智 | 257 | 59 |
| 移动 | 224 | 332 |
| 移动、客户端 | 27 | 17 |
| 客户端 | 9 | 16 |

- 国产网络游戏审批类别

- 来源：国家新闻出版署 - 国产网络游戏审批信息



# 2024年上半年主要游戏安全风险挑战

随着游戏行业的蓬勃发展，安全问题也日益成为行业关注的焦点。面对 2024 年上半年的游戏安全风险挑战，游戏行业需要不断加强技术能力，完善安全策略，与各方共同努力，打造一个更加安全、公平的游戏环境。

### 01 AI 外挂：技术与对抗的较量

AI 技术在提升开发效率，为玩家和开发者带来革命性的体验的同时，也正被不法份子用于制作更高级的外挂。AI 游戏外挂的制售团伙在市场上日益活跃，其功能性和隐蔽性不断升级，为网络安全和游戏公平性带来新的挑战。

5 月 6 日，由某射击类网络游戏公司协助警方侦破的全国首例“AI 外挂”案进行了一审公开宣判。AI 外挂中“cvc”等程序对多款游戏中游戏画面数据进行了未授权获取，对游戏中处理的鼠标数据指令进行了未授权的修改，增加了游戏中“自动瞄准”和“自动开枪”的功能，干扰了游戏的正常运行环境，属于破坏性程序。该团队利用 AI 外挂手段侵入、非法控制计算机信息系统的程序、工具，并获取巨额利润，严重破坏游戏的公平性，损害游戏方合法权益。

### 02 黑代充篡改数据非法获利

同年 5 月，上海市松江警方成功捣毁一个游戏黑代充犯罪团伙。某网络公司旗下两款手机游戏后台出现游戏异常充值数据。该游戏黑代充犯罪团伙利用游戏程序漏洞，使用技术手段篡改认证信息，以低价方式获取官方服务器正价道具商品，造成该公司重大经济损失。

### 03 模拟器助长盗版肆虐

2 月 26 日，任天堂起诉 Switch 模拟器 Yuzu 制造商，指控其违反了《数字千年版权法》（DMCA）的反规避和反交易条款，通过使用非法获得的解密密钥，允许用户在未经授权的情况下玩 Switch 游戏。最终 Yuzu 方全面接受了任天堂提出的条件，同意支付 240 万美元的赔偿金以达成诉讼和解，并全面下架 Yuzu 模拟器。

模拟器游走在游戏产业的灰色地带，其本身并不违法。而如何明晰模拟器合法使用与侵犯游戏版权的边界，依旧是留给所有相关方需长远解决的课题。尽管事件中双方和解协议已经达成，但维权并非易事，模拟器开发者和游戏公司之间的博弈仍在继续。

### 04 DDoS 攻击：网络安全的严峻考验

2 月 29 日，某 RPG 游戏在开服前夕遭受 DDos 攻击，导致服务器稳定性严重异常。且在游戏首发的 8 个小时内，开服时间的延迟、网络波动导致的游戏体验不佳、以及游戏内出现的一系列问题直接影响玩家的初次体验。

黑客组织有计划地将热门小型游戏工作室作为目标，发起 DDoS 攻击以瘫痪其服务器，迫使厂商无法正常为玩家提供服务，进而通过勒索达到非法敛财的目的。

### 05 黑灰产业链的挑战

央视 315 晚会曝光主板机黑灰产业链，只需 20 块手机主板集成的主板机，通过叠加就可以组建成千上万台手机的网络矩阵。网络黑灰产利用主板机可随意改变 IP 地址以逃避监管，进而操纵游戏、刷量、操纵网络投票。这是网络黑灰产群控技术的升级手段。

### 06 未成年人保护监管加强

1 月 1 日起，《未成年人网络保护条例》正式生效。明确要求，建立健全防沉迷制度，网络产品和服务提供者不得为未成年人提供游戏账号租售服务，强调网络游戏应坚持融合、友好、实用、有效的原则，设置未成年人模式，以醒目便捷的方式为监护人履行监护职责提供时间管理、权限管理、消费管理等功能。条例旨从根源上预防未成年人沉迷游戏，健全未成年人网络保护体制机制，培养其网络素养，确保年轻一代在数字世界中安全、健康地成长。

5 月 28 日，中国互联网协会发布《未成年人网络游戏服务消费管理要求（征求意见稿）》。首次明确了网络游戏提供方与监护人在不同情形中的“过错”责任，完善未成年人消费管理和退费规范，巩固防沉迷成果。对网络游戏服务提供者、调节机构等组织提供指导与参考。引导网络游戏产业规范化、有序化发展。

### 07 渠道 CPA 刷假量

在进入存量竞争阶段的游戏市场中买量成本走高，而基于 CPA 模式下，一些渠道商或



推广公司可能会采取刷假量的方式，即通过虚假的数据来提高游戏的下载量、注册量等指标。这种做法不仅损害游戏公司的利益，让大量营销费用换来的是无效用户和虚假数据，更是存在通过刷量手段引发市场不当竞争的作弊行为。

据 3 月上海市第二中级人民法院公布一则典型事例，某游戏公司为其掼蛋游戏推广冲榜与某网络公司签订服务合同，预定 1.8 元 /CPA。该游戏因被认定为存在欺诈、不当、非法或不诚实行为，遭到“苹果”应用商店清榜。双方的交易行为已然损害了公平有序的市场竞争环境。最终判决认定双方的刷量服务合同因有违公序良俗而无效。

### 08 全球化带来的安全挑战：多元文化与法规

游戏的全球化布局带来了多元文化和不同法规的适应问题。了解各地法律法规，尊重文化差异，是游戏全球化过程中必须考虑的安全因素。

**“游戏技术的持续革新与游戏产业的发展相伴而行。与此同时，潜伏在暗中的安全威胁也更加隐蔽，作恶技术手段更加高级。游戏安全是持久而艰巨的主题，它不仅关系到玩家的游戏体验，更是游戏产业健康持续发展的基石。游戏安全意识还需持续提升，通过加强技术研发和监管力度，护航整个产业的繁荣与发展。”**

# PART. 02 Chapter

# 僵局：上半年游戏对抗风险概览

- 外挂风险实时对抗数据概览
- 不同游戏安全风险对抗概览

# 根据网易易盾游戏安全数据显示

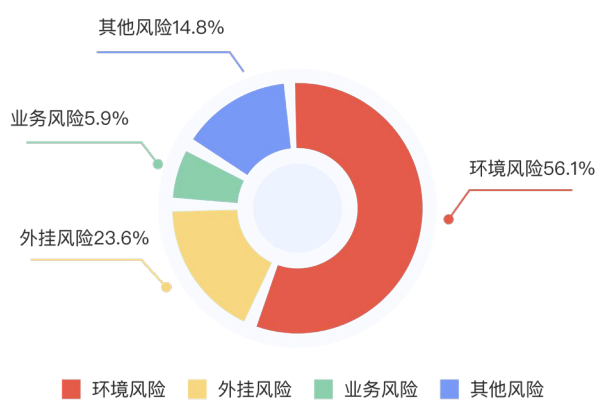
2024 年上半年，游戏总体风险对抗激烈并呈现出上升态势，易盾累计检测安全风险 342 亿次，同比增长 57%，其中环境风险威胁占比最高，检测 192 亿次，占整个检测安全风险类型的 56.1%。外挂危险持续增长，2024 年上半年累计检测 81 亿次，占比 23.6%，增幅明显。

此外，本年度全年累计检测业务风险 20.5 亿次，占游戏安全风险总体的 5.9%，同比略有下降。其他风险呈现出指数级增长态势，累计检测 48.6 亿次，占比 14.8%，同比增长 50%。

**占比图**

| 环境风险 | 外挂风险 | 业务风险 | 其他风险 |
|---|---|---|---|
| 56.1% | 23.6% | 5.9% | 14.8% |

网易易盾识别重点游戏安全风险对比



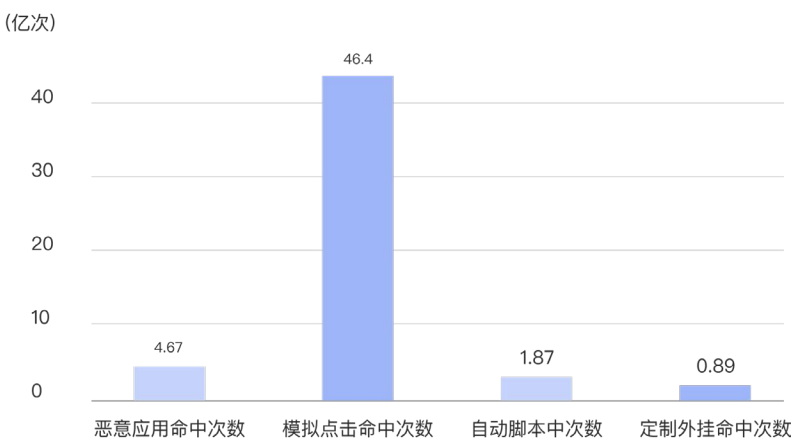



# 外挂风险

2024 年上半年，网易易盾围绕游戏外挂特征，通过定量分析手段，检测得出模拟点击类外挂 46.4 亿次，占比 34.8%；检测恶意应用 4.67 亿次，占比 5.7%；检测自动脚本 1.87 亿次，占比 2.3%；定制外挂 0.89 亿次，占比 1.09%。

| 恶意应用命中次数 | 模拟点击命中次数 | 自动脚本中次数 | 定制外挂命中次数 |
|---|---|---|---|
| 4.67亿次 | 46.4亿次 | 1.87亿次 | 0.89亿次 |

外挂检测数据柱状图




在整个外挂命中数据中，手游外挂（含小程序）风险命中 58.83 亿次，占比 66%，同比增长 54%，端游外挂风险命中 16.45 亿次，占比 20%，同比增长 17%，此外小游戏 /H5 风险环境风险和业务风险显著增加。

# 移动游戏外挂风险加剧

根据网易易盾 2024 上半年检测数据显示，网易易盾累计检测 iOS 终端受到攻击 6.7 亿次，安卓终端受到攻击 42.8 亿次，虽然两类都是移动终端系统，但是由于安全保护系统的区别，导致游戏外挂存在巨大的差异性。相比于安卓系统，iOS 受到代理、数据异常、群控等风险更大。占比分别为：40%、25%、16%、2.9%。而安卓手游受到攻击的前三种外挂行外为：模拟点击、恶意应用、定制外挂和模拟定位，占比分别为：55%、26%、10%、4%。

| 代理 | 40% | 模拟点击 | 55% |
|---|---|---|---|
| 异常设备 | 25% | 恶意应用 | 26% |
| 数据异常 | 16% | 定制外挂 | 10% |
| 群控 | 2.9% | 模拟定位 | 4% |
| IOS手游常见外挂占比 | | 安卓手游常见外挂占比 | |

不同系统手游外挂攻击对比

网易易盾观察到，近年来，手游外挂和脚本工具越来越倾向于在非真机环境下运行，部分外挂为了躲避游戏方常规的检测，通常会使用隐藏安装和特征的手段对抗风控，使得对抗难度日益增大。



# 端游外挂风险

2024 年上半年，网易易盾检测端游外挂攻击类型中，占据前四的分别是修改器、自动脚本、易语言程序和安装风险应用，占比分别为：20%、17%、8%、5%，而紧随其后的依次为加速器和辅助工具。

| 修改器 | 20% |
|---|---|
| 易语言外挂 | 17% |
| 自动脚本 | 8% |
| 加速器 | 3% |
| 辅助工具 | 1% |

端游不同类型外挂攻击占比

# 小游戏风险

2024 年上半年，网易易盾检测小游戏风险中，占据前四的分别为“环境异常、协议异常、包体异常”，占比分别为：34%、14%、11%。

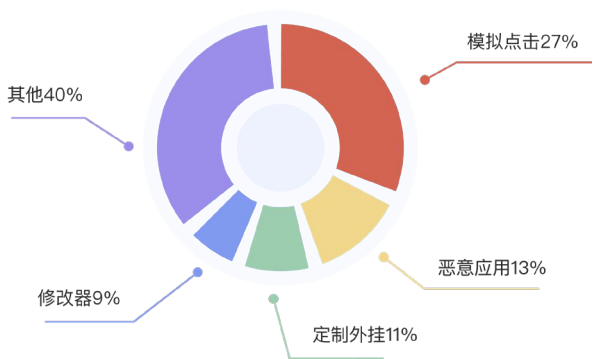
| 环境异常 | 34% |
|---|---|
| 协议异常 | 14% |
| 包体异常 | 11% |

小游戏不同类型外挂攻击占比

# 不同游戏安全风险对抗概览

## MOBA 游戏常见外挂问题

MOBA 类游戏网易易盾 2024 年上半年检测治理累计 2.8 亿次，外挂的攻击类型以通过自动化脚本实现的自动攻击、自动任务等模拟点击脚本挂为主，此外恶意应用和定制外挂占比提升明显。占比分别为：27%、13%、11%、9%



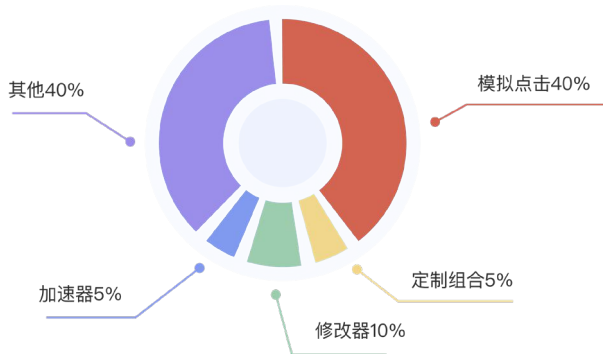

MOBA 游戏常见外挂占比

上文提到的定制外挂，是通过找到游戏内存中所有玩家的数据或相关函数并 HOOK 修改其中的可见属性，以破坏游戏的平衡性和公正性。自动化脚本则是通过外挂程序自动执行一些操作，例如自动使用技能或攻击敌人，而无需玩家手动操作。此外，MOBA 游戏内还存在大量的黑灰产工作室，他们利用自动化脚本掠夺游戏资源，破坏游戏经济平衡。



## SLG 游戏常见外挂问题

2024 年上半年，网易易盾累计检测外挂 5000 万次，其中模拟点击是该类游戏最为常见的外挂攻击形态，它帮助游戏玩家自动执行游戏中的一些任务，例如自动收集资源、自动建造建筑、自动攻击敌方玩家等。该类外挂网易易盾累计检测发现 2000 万次，占比达 40%，其中超过 80% 的脚本是定制类外挂。除此之外，还有如定制组合、修改器、加速器等外挂形态，分别占比 5%、10%、5%。




SLG 游戏常见外挂占比

模拟点击按键脚本可以帮助游戏玩家省去很多重复性的操作，快速升级自己的城市和部队以取得战斗胜利，因此备受游戏作弊玩家青睐。

除了模拟点击脚本以外，许多 SLG 游戏存在协议破解的风险。协议破解主要是外挂作者通过破解游戏协议，可以获取到游戏服务器上的关键数据，而这些数据可以被玩家用来刷分、刷经验等。同时，这类外挂程序还可以修改游戏客户端的数据，例如修改角色属性、修改游戏物品、修改游戏地图等等，从而使游戏变得更加容易或者获得其他不公平的游戏优势。

## RPG 游戏常见外挂形态

根据网易易盾游戏安全部门检测数据显示，网易易盾黑产研究院上报，2024 年上半年，该品类游戏共识别检出 2.9 万批次团伙，涉及 3869 万个黑产账号，占全品类游戏黑产检出的 93%，占 RPG 游戏外挂比例高达 80%。

从工作室作弊形式上看，很多工作室团伙已从过去的多设备、多开模式转为了单设备、单开群控模式。在作弊工具方面，手机移动端多采用虚拟空间、云真机等工具，或者利用 PC 模拟器实现群体多开。而在激烈的端游和手游市场竞争中，一些工作室团伙甚至利用外接硬件设备进行辅助多开行为，如同步器和主板机群控，再配合自动按键脚本实现批量获利。

近年来，综合网易易盾打金工作室治理实践发现，黑产在躲避检测方面的手段也在不断升级，从普通的联网到修改 IP 再到本地组建局域网。在这个过程中，黑产团伙不仅会摸索游戏中的变现方式，还会想方设法绕开游戏公司现行的检测手段，从而压缩普通玩家的生存空间，造成游戏内通货膨胀和价格波动，进而影响游戏公司收益。

## 棋牌游戏外挂形态

2024 年上半年，网易易盾累计检测棋牌类游戏脚本挂 1188658 次，其中首充号问题较为严重，占比高达 80%。除此之外，棋牌类游戏还存在计分牌等外挂风险。

首充号主要是一些不良商家会利用首充号的优惠政策，通过虚假账号进行大量充值，从而获得高额的返利或奖励。

因多数棋牌类游戏都会推出小金额首充活动，以此让玩家在游戏中获得大量的游戏币。黑产分子正是利用这种低价的首充机制，通过自动注册和创建大量重复的首充号，再通过挂机输给指定的聚金号来倒卖获利。



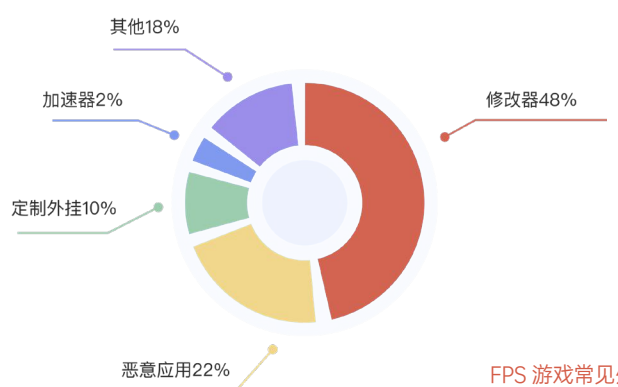
整个过程中，作弊者通常会借助账号注册机、打码平台和自动脚本等辅助工具来完成以上操作。而这种大规模的首充号注册攻击不仅会给游戏造成大量的小额付费，还会对游戏的正常运营造成影响。

重复账号的大批量注册，对游戏服务器造成巨大负担的同时，大量首充号注册会导致游戏通道拥堵，影响真实的玩家进入游戏。尤其在游戏开服前期，等待时间过长会造成大量玩家的流失。

## FPS 游戏安全问题

由于射击类游戏本身需要大量数值计算，游戏方会将部分计算存放于本地客户端，而这为外挂攻击者提供了攻击的温床。可以说，射击类游戏是所有游戏中被外挂攻击最为频繁的游戏类型。

根据网易易盾游戏安全部门检测数据显示，FPS 游戏网易易盾累计检测修改器类外挂 2199 万次，占比达 48%。除此之外，还有如恶意应用、定制外挂、加速器等形态，分别占比 22%、10%、2%。




FPS 游戏常见外挂占比

修改类外挂在功能上，最常见的是自动瞄准、透视功能、无限弹药、生命值和速度加速等。

除了修改类外挂以外，经外挂衍生出来的第三方黑产同样需要重视，比如对游戏影响较大的＂坐挂车＂、外挂护航等违规行为。

在＂坐挂车＂中，游戏玩家通过跟外挂服务者组队，依靠外挂服务者击杀其他玩家以获取高分。外挂护航则是由服务方开启大量外挂角色，与购买者同时匹配，以同样的清扫战场的方式，保送购买者获得高排名。这些外挂者角色价格本身较为低廉，即使受到游戏方处理，损失也会较小，因此需要对购买方做出一定的处罚和限制。

此外，射击游戏中还存在恶意组队、演员等违规现象，这些现象会影响整体游戏平衡和体验，游戏方需要重视并制定相应的制度策略进行管控。

由于射击类游戏主要以即时对战的形式进行，游戏玩家对外挂攻击的感知度更强，因此游戏方需要更高效、更准确地处理外挂攻击。

## 小游戏安全风险

受全球疫情影响，自 2020 年春节以来，小游戏用户数量显著增长。这主要是因为小游戏能够满足用户在碎片化时间内的娱乐需求。此外，小游戏还具有促进社交互动的功能，这不仅在微信平台上得到体现，国内其他平台如快手、抖音、百度等也相继推出了自己的小游戏平台。与大型手游相比，小游戏具有体积小、快速加载或无需安装、占用手机资源少等优点，通常以休闲益智为主题，特别受到特定场景下用户的青睐。

然而，随着小游戏种类、数量和形式的增多，它们也吸引了一些不法分子的注意。这些不法分子可能会通过窃取代码或资源、植入广告或恶意链接等方式对小游戏进行破坏。



因此，小游戏领域存在多种潜在风险，需要用户和开发者保持警惕。目前市面上小游戏面临的风险点主要在于以下几种：

**01 面临层出不穷的盗版风险**

外挂作者通过非法手段获取小游戏的 .wxapkg 安装包，这是微信小程序的专用格式，然后使用市面上反编译工具进行解包。在解包后，他们对游戏的源代码进行修改，可能增加恶意功能或修改核心数值，接着进行所谓的＂换皮＂操作，即更换游戏的界面元素和图片资源，以掩盖其盗版本质。

**02 来自广告的核心收益受损**

在当前的小游戏市场中，广告是游戏开发者获取收入的一种常见方式，但这也带来了一些问题。不法分子利用外挂工具或协议抓包，对广告内容进行篡改，其中包括修改为推广广告：这些广告通常是为了推广其他游戏或产品，有时它们会以跳转链接的形式出现，用户点击后可能会被引导至其他游戏或应用；恶意广告：这类广告可能包含欺诈内容、恶意软件下载链接或其他对用户有害的元素。诱导性广告：这些广告通过诱导用户点击或分享来增加其曝光度或传播范围，有时这些行为可能会侵犯用户的隐私或导致不必要的信息泄露。更甚者会直接跳过广告环节，从而使业务方在广告方面的收益直接受损。

**03 小游戏黑产团伙工作室**

小游戏中的黑产团伙通过运用虚拟设备和分身工具，以较低成本实施大规模群体作弊，这种行为不仅严重破坏了游戏的公平性，还对游戏公司造成了经济损失和声誉风险。由于这些团伙的隐蔽性和运行平台的数据局限，检测和识别作弊行为面临重重困难，导致＂薅羊毛＂、＂团伙打金＂等非法牟利现象频发。

**04 通过各种手段实现外挂功能**

外挂团伙通过在模拟器、云手机和虚拟机中加载特定版本的微信，利用变种的 GG 修改工具来篡改微信小游戏的内存数据，实现游戏内核心数值的非法修改，这种行为不仅破坏了游戏的平衡性，还严重损害了游戏的口碑和玩家的体验。此外，当前较为常见的手段还包括对游戏的数据包体协议进行扒取，从而获取敏感的明文信息，进而通过修改协议来篡改游戏数据和功能。这些非法行为不仅对游戏的正常运营构成威胁，也对玩家的账户安全和隐私保护带来了风险。

# PART. 03 Chapter

# 破局：游戏安全治理驱动游戏健康发展

- 游戏安全治理的价值
- 游戏安全发展概况
- 游戏安全常见的业务场景

# 游戏安全治理的价值

游戏安全治理在现代游戏行业中占据着至关重要的地位。它不仅仅是保护游戏免受黑客攻击和作弊行为的必要手段，更是确保玩家有一个公平、公正、愉快的游戏体验的基础。以下是游戏安全治理的几大核心价值：

**保障玩家权益：** 有效的安全治理措施能够防止玩家账户被盗、虚拟财产被窃取等问题，维护玩家的合法权益。特别是在如今玩家投入大量时间和金钱在游戏中的情况下，保护玩家账户和虚拟财产免受攻击显得尤为重要。

**提升游戏信誉：** 安全稳定的游戏环境能够提升玩家对游戏的信任度，从而增强游戏的品牌形象和市场口碑。玩家往往更愿意投资时间和金钱在一个他们认为安全可靠的游戏中。

**促进游戏生态健康发展：** 通过防止外挂、作弊等行为，确保游戏规则的公平性，促使游戏社区和谐发展，增加玩家的留存率和参与度。健康的游戏生态能够吸引更多的新玩家，同时保持老玩家的积极性和忠诚度。

**确保经济体系稳定：** 游戏内的虚拟经济系统需要依赖安全的交易环境，避免虚拟货币的通货膨胀和经济系统崩溃，从而保持游戏内经济活动的正常进行。一个稳定的经济系统对游戏的长期成功至关重要。

**合规性和法律要求：** 随着全球范围内对数据隐私和网络安全的法律法规日益严格，游戏公司必须确保其安全治理符合相关规定，以避免法律风险和潜在的罚款。



# 游戏安全发展概况

随着游戏产业的迅速发展，游戏安全问题也变得愈发复杂化和多样化。从最初的简单防护措施到如今复杂的多层次安全体系，游戏安全技术经历了显著的演变：

**早期阶段：** 最初的游戏安全主要集中在基础的防火墙和简单的反作弊措施上，主要应对基础的黑客攻击和作弊行为。这些措施虽然初级，但在当时的技术环境下已经能提供一定程度的保护。

**中期发展：** 随着网络游戏的普及，安全问题逐渐多样化，游戏公司开始采用多层次的安全方案，包括用户行为分析、动态防护系统和多因素认证等。这一阶段，游戏安全不仅仅关注外部攻击，还开始注重内部威胁和漏洞修复。

**当前阶段：** 现如今，AI 和大数据技术在游戏安全中的应用越来越广泛。利用机器学习算法，可以实时监控和分析玩家行为，识别和预防潜在的安全威胁。同时，云计算技术的普及也使得游戏安全防护更加高效和灵活。例如，网易游戏通过 AI 技术实现了智能反作弊和自动化威胁检测，极大地提升了安全治理的效率。

此外，游戏安全技术的发展也受到以下几个关键趋势的推动：

**跨平台安全：** 随着跨平台游戏的兴起，游戏安全治理需要应对不同平台间的安全差异，确保无论玩家使用的是 PC、主机还是移动设备，都能享受到同样的安全保护。

**实时响应能力：** 现代游戏安全需要具备实时响应和处理能力，以便在发现威胁的第一时间进行处理，避免造成更大的损失。AI 技术和自动化工具在这方面发挥了重要作用。

**全球化安全策略：** 随着游戏公司全球化扩展，其安全治理需要考虑不同国家和地区的法规和安全需求，制定符合各地要求的安全策略和措施。

# 游戏安全常见的业务场景

在实际运营中，游戏安全治理涉及多个具体的业务场景，每个场景都有其独特的安全需求和挑战：

**账号安全：** 包括账号注册、登录和交易等环节的安全保护。常用的措施有多因素认证、异常登录检测和密码强度要求等。通过生物识别技术（如指纹识别、面部识别等）进一步提升账号安全性，阻止未经授权的访问。

**防作弊与反外挂：** 通过监测玩家行为和游戏数据，识别并阻止作弊行为和外挂使用。AI 技术在这方面有着广泛的应用，可以自动学习和识别新的作弊模式。例如，通过分析玩家操作的异常性和游戏内行为模式，可以有效识别并阻止外挂用户。

**虚拟财产保护：** 保护玩家的虚拟财产不被非法获取和盗取。包括虚拟物品交易的安全监控、交易记录的追踪等。利用区块链技术，可以为虚拟物品和交易提供更高的安全性和透明度。

**聊天及社交安全：** 监控游戏内的聊天和社交互动，防止诈骗、骚扰和不当言论的传播。利用自然语言处理技术和关键词过滤等手段进行实时监控，并通过社区管理和举报机制来维护社交环境的健康。

**支付与交易安全：** 确保玩家在进行游戏内购买和交易时的支付信息安全，防止支付欺诈和虚假交易。通过采用安全支付网关、加密技术和反欺诈系统，可以有效保障玩家的支付安全。

通过全面的游戏安全治理措施，游戏公司不仅可以保护玩家的权益和游戏生态的稳定，还能为游戏的长期健康发展奠定坚实基础。在网易游戏的安全治理实践中，AI 技术的引入和应用进一步提升了安全防护的效率和效果，为游戏行业树立了新的标杆。




网易易盾通过一系列创新性的安全技术和管理措施，成功应对了多种安全威胁，建立了一个安全、可靠的游戏环境。例如，通过大数据分析和 AI 技术，网易游戏能够精准识别和处理异常行为，快速响应和修复安全风险。

# PART. 04 Chapter

# 重塑：AI 风控驱动游戏行业稳健增长

- 外挂治理
- 黑灰产
- 内容合规风险
- 隐私合规
- 未成年治理

随着 AI 技术在制挂、用挂中的使用，游戏厂商对于外挂的打击和治理难度呈现指数级增加。同时，AI 在外挂治理方面的应用又会对传统外挂、AI 外挂的治理创造全新的价值，甚至带来革命性的变革。

AI 的核心优势在于其情报收集、大数据的处理能力和玩家行为识别技术，这些能力使其在外挂检测和预防方面表现出色。

网易易盾认为，AI 技术与传统检测方式的结合，可以实现多维度的外挂打击策略，包括实时检测、精准打击和自动更新对抗策略，有效应对不断进化的外挂技术。同时，AI 在黑灰产团伙检测方面展现出巨大价值，通过数据分析识别团伙行为，构建工作室画像，从而进行有效打击。



# 外挂治理

外挂的治理需要多种维度、多种策略的对抗，网易易盾智能风控反外挂系统集成了户端加固、AI 模型分析、行为特征识别等一系列治理手段。

其中，客户端加固技术通过脚本文件保护、资源文件加密和引擎保护等措施，增强了游戏的安全性，提高了破解的难度。

智能风控引擎则利用大数据分析和机器学习，实时监测并识别异常行为，有效区分正常玩家和使用外挂的玩家。

此外，包括按键脚本识别、协议防破解、变速检测、内存读取 / 修改检测等功能，可以全面覆盖外挂行为的各个方面。

## 按键脚本

在游戏领域，按键脚本是一种常见的外挂形式，它们通过自动化操作给作弊者带来不正当优势，破坏了游戏的平衡和玩家体验。

网易易盾结合 AI 轨迹模型，能够精准识别常规和非常规的按键类脚本。这些脚本可能运行于手机、模拟器或 PC 环境中，但无论其运行环境如何，系统都能通过分析屏幕点击坐标和用户交互行为，有效区分自动化脚本操作与正常玩家操作。

此外，通过持续收集和分析定制脚本样本，智能风控系统能够快速响应新出现的外挂威胁，实现对按键脚本的实时检测和有效阻断。

## 协议防破解

游戏外挂经常通过截取和修改网络数据包来实现作弊功能，网易易盾通过在客户端实施严格的加密措施，保护游戏数据包不被未授权的第三方软件截取和篡改。

同时，使用白盒加密算法，系统能够内置多种加密协议，实现一次一密的强校验，有效防止主动调用和数据泄露。此外,系统还结合了身份验证机制,进一步加强了安全性。当游戏数据在客户端被加密后，只有合法的服务端能够解密这些数据。这个过程涉及到算法和时间戳的匹配，确保了数据的新鲜度和安全性。任何非法的篡改尝试都会因为加密算法的复杂性和强校验机制而失败。

另一个关键组成部分是其对异常行为的监测能力。系统能够检测到数据包的不规则波动，这可能表明有外挂软件试图修改游戏速度或执行其他非法操作。通过模拟网络波动，系统能够在变速的前提下进行不规则倍率跳动，从而有效识别和阻止非法行为。

## 变速检测

变速检测专门针对那些尝试通过非法手段改变游戏速度，以获得不正当优势的外挂行为。

该系统通过先进的算法监测游戏运行速度，以识别任何异常的加速或减速行为。当检测到非正常的加速倍率时，系统能够及时响应，采取措施阻止这种行为，确保游戏速度的正常运行，维护游戏的公平性。

一个重要的特点是系统能够模拟网络波动，通过在变速的前提下进行不规则倍率跳动



来模拟真实网络条件下的游戏运行情况。这种模拟不仅增加了外挂检测的复杂性，也提高了检测的准确性，使得潜在的作弊者难以通过常规手段规避检测。

此外，变速检测功能与系统的其他防护措施相结合，如内存读取 / 修改检测和协议防破解等，形成了一个多层次的防护网络。这确保了即使外挂开发者尝试通过多种手段进行作弊，系统也能够从不同角度进行识别和防范。

## 内存读取

内存读取外挂通过非法读取或修改游戏内存数据，破坏游戏平衡，损害玩家体验。

网易易盾采用内存陷阱技术和内存监控机制，设立监测点捕捉异常内存访问行为。当系统检测到非法的内存读取或修改尝试时，能够迅速识别外挂特征，并触发相应的安全响应。例如，通过在游戏内存中布置陷阱，一旦有外挂尝试触发这些陷阱，系统便能立即察觉并记录行为特征。

此外，网易易盾通过分析修改执行特征，能够区分正常游戏过程与外挂行为。系统不仅能识别修改前的状态，还能监测修改后的结果，如增减障碍物等，从而判断是否存在作弊行为。
同时，对修改器启动特征的检测，通过监控游戏运行时的内存变化，及时发现并响应外挂程序的加载和执行。这种实时监控和响应机制，有效提升了游戏安全性，保护了游戏的公平竞争环境。

## 跨端修改

跨端作弊行为通常涉及使用 PC 工具来非法修改移动设备或模拟器上的游戏数据，以获得不正当优势。

网易易盾的跨端修改检测功能通过高级监控和分析技术，能够捕捉到这些非法修改行为的迹象。它利用独特的算法对游戏的网络流量和内存活动进行实时监控，确保任何试图通过外部工具或软件进行的非法数据修改都能被迅速发现。

一个关键的检测手段是内存陷阱技术，它能够在游戏的内存空间中设置陷阱，一旦有外挂尝试进行非法内存读写，系统便能立即识别并记录这些行为。此外，系统还能够检测到修改执行特征，比如外挂启动前后的游戏状态变化，从而判断是否存在作弊行为。

此外，跨端修改检测功能与系统的其他安全措施相结合，如客户端加固、协议防破解、变速检测等，形成了一个全面的防护体系。

## 盗版

盗版软件不仅侵犯了开发者的权益，还可能带来安全风险，如携带恶意软件或病毒，损害玩家利益。

网易易盾基于签名和哈希值的自动化检测机制，能够有效识别未经授权的软件副本。通过内置的算法，系统能够对游戏的数字签名进行验证，确保只有合法版本能够运行。这种技术的应用大大提高了盗版软件的检测和拦截效率。



同时，基于对盗版软件行为模式的分析能力，网易易盾能够识别和记录盗版软件的分发渠道和使用情况，为开发商提供了打击盗版行为的有力数据支持。此外，智能风控反外挂系统还能够与开发者的运营策略相结合，通过技术手段限制盗版用户的某些功能或服务，从而降低盗版软件的吸引力。

在用户体验方面，网易易盾通过精准识别盗版软件，避免了对正常用户造成误伤，确保了所有玩家都能获得公平且一致的游戏体验。同时，系统还能够为开发者提供灵活的策略配置，以适应不同市场和不同用户群体的特定需求。

# 黑灰产

在数字娱乐的繁荣背后，游戏黑灰产悄然滋生。这一非法产业链通过开发和销售外挂软件，破坏游戏平衡，为使用者提供不正当的竞争优势。作弊者利用这些工具，进行自动化操作，如自动打怪、刷金币、模拟点击等，从而在游戏内获得大量非法资源。

此外，黑灰产还包括账号盗窃和非法交易行为，如通过盗取玩家账号进行非法牟利，或在游戏外进行虚拟物品的非法买卖。这些行为不仅侵犯了玩家的权益，也对游戏公司造成了巨大的经济损失。

游戏黑灰产的蔓延，损害了游戏行业的健康发展，引起了社会各界的广泛关注。游戏公司、监管部门和技术提供商需要共同努力，通过法律、技术和教育等手段，打击黑灰产，保护玩家利益，维护游戏生态的公平与秩序。

戏中的黑产场景通常分为三个阶段：创号阶段、囤号阶段和获利阶段，每个阶段都有其特定的操作和目的。

**创号阶段：**在这个阶段，黑产团伙利用自动化工具或服务，如云手机、模拟器和设备墙技术，批量创建游戏账号。他们可能使用虚拟手机号接收验证信息，通过脚本自动化完成注册过程。这些账号随后被用于游戏内的非法活动，如利用新账号的福利进行初始资源积累。

**囤号阶段：**囤号阶段是黑产团伙通过脚本和自动化工具批量登录和管理这些账号的阶段。他们利用这些账号参与游戏内的福利活动，如首充优惠，以囤积游戏币或其他有价值的资源。这些资源随后可以被用于游戏内的交易或在黑市上出售，从而实现变现。



**获利阶段：**在获利阶段，黑产团伙通过多种手段将囤积的游戏资源变现。常见的手段包括送礼 / 买卖账号、倒金（在游戏内通过交易将资源转移给其他玩家，然后通过线下支付获得真实货币），以及出售成品账号等。这些行为破坏了游戏的经济平衡，损害了正常玩家和游戏公司的权益。

针对这些黑产场景，游戏公司和安全服务提供商如网易易盾采取了一系列措施。例如，在创号阶段，可以通过增加注册难度、实施更严格的验证流程来提高黑产团伙的成本。在囤号阶段，可以通过监控异常登录行为、限制账号活动来识别和阻止黑产行为。在获利阶段，可以通过实施交易监控、限制资源转移和加强与支付平台的合作来打击非法交易。

此外，游戏公司还可以通过法律途径对黑产团伙进行打击，与执法机构合作，追踪和起诉这些非法活动。同时，提高玩家对黑产行为的认识，鼓励他们通过官方渠道报告可疑行为，也是维护游戏健康生态的重要一环。

## 打击模拟器黑产行为，提升工作室起号难度

在创号阶段，黑产工作室可能利用模拟器进行多开操作，通过复制、镜像和重装 APP 等手段躲避业务方自建的 IP 聚合，实现单 PC 无限多开。这种行为不仅破坏了游戏的公平性，还对游戏的经济系统造成了影响。

为了有效打击这种行为，网易易盾采取了一系列措施。首先，建立了模拟器三重指纹验证系统，包括本地 PC 指纹、服务端 PC 指纹和模拟器指纹，以即时计算多开数量。这种验证机制能够快速识别出模拟器的多开行为，从而采取相应的反制措施。

其次，与业务方联合在玩家登录阶段对多开玩家执行踢下线操作。这种操作的整个

识别处理周期控制在 3 秒以内，确保了快速反应的能力。同时，通过这种方式，指纹碰撞率被控制在百万分之三以下，大大降低了误判的可能性。

通过这些措施，网易易盾的快速反应机制显著提升了工作室起号的难度，降低了其效率，增加了设备成本，迫使黑产工作室选择其他游戏牟利。这不仅保障了游戏的 DAU 健康发展，还维护了游戏经济系统的稳定，并保障了普通用户的权益。

## 红包：传感器驱动的AI模型

“传感器数据驱动的 AI 模型”是网易易盾游戏风控解决方案中的一个创新技术应用，专门用于精准打击红包黑产行为。这种技术的核心在于利用移动设备的传感器数据，结合人工智能算法，识别并打击那些利用自动化手段抢红包的黑产团伙。

在某些游戏中，红包作为一种激励机制，本意是为了提升玩家的活跃度和游戏体验。然而，黑产团伙通过使用群控设备和自动化脚本，大量抢红包，不仅破坏了游戏的公平性，还严重影响了普通玩家的游戏体验。这些团伙的行为往往具有高度的隐蔽性和复杂性，传统的打击手段难以有效应对。

针对这一问题，网易易盾的传感器数据驱动 AI 模型通过采集设备的传感器数据，如加速度计、陀螺仪等，获取设备在操作过程中的三维向量集。通过对这些数据进行去噪和归一化处理，AI 模型能够识别出设备在操作过程中的异常模式。例如，正常玩家在操作手机时，设备的移动轨迹会呈现出一定的随机性和多样性，而自动化脚本操作往往具有规律性和重复性。

通过训练有监督的 AI 模型，系统能够学习并区分正常玩家和黑产团伙的操作行为。这种模型不仅能够识别出设备是否处于移动或静止状态，还能够通过聚类分析，识



别出具有群体操作特征的设备，从而精准打击红包黑产行为。

此外，这种 AI 模型还能够适应不同的游戏环境和黑产手段的变化。随着黑产团伙不断更新其作弊手段，AI 模型可以通过持续学习和数据积累，不断提升其识别和打击能力。

## 主板机：多维数据融合与图神经网络

“多维数据融合与图神经网络”技术是网易易盾在其游戏风控解决方案中采用的一种先进方法，专门用于识别和打击主板机黑产团伙。这种技术通过综合分析多种数据源，利用图神经网络的深度学习能力，实现对复杂黑产行为的精准识别。

在游戏行业中，主板机黑产团伙通过将多台手机主板组合起来，组装成具有群控操作功能的设备，从而实施作弊行为。这些团伙利用特殊的硬件配置和网络协议，进行非法操作，如抢红包、刷分等，严重破坏了游戏的公平性和经济系统。

为了有效识别这类黑产团伙，网易易盾采用了多维数据融合技术，结合设备的电池、主板特征等硬件信息，以及其他行为和资源特征，构建了一个全面的设备画像。这些数据被输入到图神经网络模型中，通过模型的深度学习能力，对设备行为进行分析和模式识别。

图神经网络是一种适合处理图结构数据的算法，它能够捕捉设备间的复杂关系和交互模式。在主板机黑产场景中，图神经网络能够识别出设备间的异常连接和行为模式，如设备间的同步操作、非正常的网络流量等。通过对这些模式的分析，图神经网络模型能够对主板机黑产团伙进行聚类和识别。

此外，网易易盾的风控解决方案还包括了快速的响应机制，一旦识别出黑产行为，系统能够立即采取措施，如踢下线、账号封禁等，以最大限度地减少黑产行为对游戏生态的影响。

## 团伙检测：

团伙黑产管控 - 团伙检测”是网易易盾游戏风控解决方案中的一个关键组成部分，专门用于识别和打击游戏中的团伙黑产行为。这些团伙通常运用复杂的手段进行非法活动，如利用模拟器多开、群控设备、恶意篡改设备信息等方式进行作弊，严重影响游戏的公平性和经济系统。

团伙黑产管控的核心在于通过 SDK 数据收集，结合 AI 算法进行深入分析，从而有效对抗各种复杂的工作室团伙黑产场景。这些场景可能包括设备多开、群控牧场、恶意篡改设备信息等，这些行为往往难以通过传统手段检测和防范。

易盾的团伙检测技术利用 60 多个设备因子，通过聚类 AI 算法对行为特征、设备特征和资源特征进行综合分析。这种方法能够识别出异常的设备使用模式和行为模式，从而将正常用户与潜在的黑产团伙区分开来。

在检测到团伙黑产行为后，易盾的风控系统可以采取一系列管控措施。这些措施包括但不限于活体检测、验证码检测、踢下线、账号封禁和收益干扰等，以确保游戏环境的健康和公平。活体检测算法通过创新的“光线深度分析”和领先的“真人认知人脸模型”，结合设备环境多维度分析，识别摄像头注入、底层篡改等风险。

此外，团伙黑产管控还包括对游戏业务的自处理能力，使游戏方能够根据检测结果自行采取相应措施，如踢下线操作，以及根据风险等级和用户信息进行账号封



禁等。这不仅保障了游戏的日活跃用户（DAU）健康发展，还维护了游戏经济系统的稳定，并保障了普通用户的权益。

## 渠道反作弊

手游渠道服是指游戏开发商与特定渠道合作，针对该渠道的用户群体开放的服务器。这种合作模式允许游戏开发商通过不同的渠道来吸引和维护用户，把一部分的运营推广的代理权下放到各大厂商平台，例如百度、华为、九游、vivo 等各种应用商店，同时也为渠道方提供了独特的内容和服务，以增加用户粘性和收益。渠道服是游戏市场推广和运营策略的一部分，通过与不同渠道的合作，游戏可以覆盖更广泛的用户群体，提高游戏的市场渗透率和收益。

手游渠道服服务器两种常见做法：渠道专服和混服。渠道专服是指某些渠道方可能会提供服务器资源，特别是当渠道方拥有强大的技术基础设施时。这种情况下，渠道方负责服务器的运营和维护，渠道用户和官方用户不在一个服务器内，游戏内容不互通，对玩家数据拥有自主权，基本仅同步游戏商玩家加密账号和角色 ID。混服是指开发商管理游戏逻辑服务器，而渠道方管理用户接入服务器或提供特定的技术支持，渠道用户和官方用户在一个服务器内，游戏内容互通，渠道用户数据共享。

# 内容合规风险

近年来，游戏产业已经成为了文化产业的另一种载体，玩家在游戏内通过文字、语音、图片、表情来进行互动聊天，同时在一些开放性的游戏内容植入会让玩家拥有更多自主权的玩法，比如派对游戏自主创建比赛关卡，角色扮演类游戏中基于大语言模型的智能 NPC 等等。

在游戏内，内容演变成了玩家输出和玩家引导输出两种形式。

## 玩家内容：言论自由需要在阳光下

玩家相关的内容涉及方方面面，比如个人头像、昵称、开放角色形象、关卡、互动聊天内容、互传资料等。

但是所有的内容都需要在大的法律、法规的框架下进行，比如玩家内容违反社会伦理法律、暴力、色情、恐怖主义等不良内容被不法分子加以利用后，会在很大程度上影响游戏正向价值与持续运营。

此外，游戏运营者需要确保玩家尊重种族、民族、文化、地域、风俗等游戏繁荣，而不是散播恶搞、挑衅、煽动仇恨言论。

因此，游戏运营者需要对游戏内玩家创作的内容进行谨慎审核和审查，确保内容合法、合规、合理。



## AI 内容：关注引导不良言论与侵权

目前，AI 技术已经被应用到了游戏的方方面面，比如在内容领域包含智能 NPC、AI 生成背景音乐、AI 生成美术、AI 游戏客服等等。

其中，智能 NPC 作为游戏中的非玩家角色，虽然能提供更加真实的互动体验，却也存在被部分玩家利用，并生成恶搞内容从而带来法律及合规风险。

AI 生成的背景音乐和美术作品虽然能节省开发时间和成本，但可能触及版权问题，特别是在 AI 学习了受版权保护的作品后。不仅可能引发法律纠纷，还可能减少音乐和美术产业的就业机会和创作动力。

AI 游戏客服提供了自动化的玩家支持，但可能因无法准确理解玩家需求而导致服务体验下降。智能客服系统也可能成为网络攻击的目标，如通过恶意代码获取玩家个人信息。

所以，加强 AI 生成内容的安全性测试势在必行，确保 AI 生成内容的合法性和原创性。

## 拉人引流：热门游戏正成为拉新渠道

随着版号审批与发放的常规化，新游戏在买量方面的竞争将愈加激烈，而传统买量渠道不仅价格高昂，同时可能存在买量欺诈问题，所以当前热门游戏会成为其他平台获取玩家 / 用户资源的一种特殊渠道，他们通过在平台中发送相关的拉人、引流广告，引导玩家到其他平台试玩，而此类广告通畅又会有非常多的变种，较难识别，游戏平台如果不进行监管，则会导致用户逐步流失，进而影响游戏收益。

**违规内容治理难度**

国内游戏中的违规内容由于涉及数量巨大，类型多样，更新快等特点，同时游戏违规内容又涉及各类监管问题，所以在治理难度上相对于其他载体更为困难，具体来说：

- **数量巨大**

违规内容，特别是文本内容的量级非常大，易于传播且影响群体广泛，解决这些问题需要投入大量的人力、资源来处理海量数据和信息，同时还要快速进行甄别。

- **类型多样**

违规内容包括色情、广告、谩骂、违禁等多种类型，甚至还会出现大量不同内容的组合形态，并且游戏场景不同，同一内容的理解也会不同，因此针对不同的游戏场景和违规类型，需要采用不同的处理方法和技术，依赖高效精细的治理体系。

- **更新快**

黑灰产常常采用文字变种、分段发布等方式，以绕开反垃圾策略，这使得平台需要持续攻防，增加了打击违规行为的难度。



# 隐私合规

游戏隐私合规是指游戏开发者和运营商需要遵守相关法规和规定，保护玩家的个人隐私和数据安全。

在游戏开发和运营中，涉及到了大量的个人数据，如玩家的账号、游戏记录、支付信息、聊天信息等，因此需要严格保护玩家的隐私和数据安全。

比如，手游厂商在收集使用玩家个人信息时的“告知 - 同意”模式，需要明确告知数据采集的目的、范围和使用方式。此外，游戏厂商需要严格保护玩家个人数据的安全，采取必要的安全措施，防止数据被盗取、泄露、滥用等。而游戏玩家有权访问、更正和删除自己的个人数据，游戏开发者和运营商需要提供相应的服务和支持。

# 未成年治理

2024 年 1 月 1 日，《未成年人网络保护条例》正式实施，条例中与游戏行业相关的内容包括：

网络游戏服务提供者的责任：网络游戏服务提供者应当建立、完善预防未成年人沉迷网络的游戏规则，对游戏产品进行分类并予以适龄提示。

实名制和身份验证：网络游戏服务提供者应当通过统一的未成年人网络游戏电子身份认证系统等必要手段验证未成年人用户真实身份信息。

游戏时长和消费限制：网络游戏企业向未成年人提供游戏服务的时长，法定节假日每日累计不得超过 3 小时，其他时间每日累计不得超过 1.5 小时。同时，对未成年人的游戏消费进行限制，例如 8 周岁以上未满 16 周岁的用户，单次充值金额不得超过 50 元人民币，每月充值金额累计不得超过 200 元人民币。

内容规范: 网络产品和服务中含有可能影响未成年人身心健康的信息的，制作、复制、发布、传播该信息的组织和个人应当在信息展示前予以显著提示。

网络欺凌防治：网络产品和服务提供者应当建立健全网络欺凌行为的预警预防、识别监测和处置机制，设置便利未成年人及其监护人保存遭受网络欺凌记录、行使通知权利的功能、渠道。

个人信息保护：网络服务提供者为未成年人提供信息发布、即时通讯等服务的，应当依法要求未成年人或者其监护人提供未成年人真实身份信息。

另外，对于违反《未成年人网络保护条例》规定的主体，将依法承担相应的法律责任，包括但不限于罚款、暂停业务、吊销营业执照等。

# PART. 05 Chapter

# 翻盘：游戏安全产品与服务

- 全生命周期游戏对抗
- 游戏风控发展趋势

# 全生命周期游戏对抗

- 游戏风控解决方案


游戏安全对抗是一个多方动态博弈和演进的过程，包含了攻击技术制造者、使用者、防守方、防守技术开发者，而随着技术的进步和市场环境的变化，进攻技术与防守技术的博弈将更加猛烈。

游戏开发者和运营者需要关注游戏不同阶段存在的各类安全隐患，并制定相应的防范和对抗措施，从而保护游戏玩家账号财产安全、游戏对抗公平。

具体来说：

**01 初始阶段：基础防护措施**

- 密码保护：通过设置账号密码，提供基本的账户安全。
- 验证码：通过图形或短信验证码防止自动化脚本注册和登录。

**02 外挂与作弊防护**

- 外挂检测：开发外挂检测系统，通过特定算法识别外挂行为。
- 实时监控：对游戏内的行为进行实时监控，快速响应作弊行为。



**03 风控系统化**

- 集成系统：建立集成的风控系统，集中管理账号安全、交易监控等。
- 数据分析：利用数据分析技术，对游戏内数据进行挖掘，识别异常模式。

**04 内容风控的重视**

- 敏感词过滤：建立敏感词库，自动过滤不恰当的文本内容。
- 图像识别：应用图像识别技术，识别并处理违规图片和视频。

**05 全球化风控挑战**

- 多语言支持：风控系统需要支持多语言，适应不同地区的需求。
- 文化差异考量：考虑不同文化背景下的内容风控策略。
- 法律与合规标准不一：护航游戏出海。

**06 用户体验与风控平衡**

- 误封减少：优化算法减少误封情况，提高玩家满意度。
- 玩家申诉：建立玩家申诉机制，及时处理玩家对风控措施的反馈。

**07 风控定制化与智能化**

- 定制化策略：根据不同游戏类型和玩家行为定制风控策略。
- AI 应用：运用机器学习和人工智能技术，提高风控的智能化水平。

**08 风控信息共享与合作**

- 行业合作：与同行业其他公司合作，共享风控信息。
- 第三方服务：与专业的风控服务提供商合作，提高风控效率。

**09 法规合规与未成年人保护**

- 合规性检查：确保游戏内容和运营活动符合当地法律法规。
- 防沉迷系统：开发防沉迷系统，限制未成年人的游戏时间。

# 游戏风控发展趋势

**AI 技术将被广泛使用：**AI 技术已经被广泛用在作弊和反作弊策略中，可以预见在接下来很长的时间内，作弊的对抗将由人 vs 人向人 +AI vs 人 +AI 过渡。而作为防守者的游戏厂商和游戏安全服务商，必须不断适应这种变化，采用先进的 AI 算法来检测和阻止作弊行为。

**黑灰产团伙隐藏技术将更加高明：**随着作弊者与防守者对抗的持续激烈，作弊者使用高权限隐藏或伪装成正常文件特征，利用加密、混淆等手段增加分析难度，使得游戏安全团队更加难以发现。

**未成年人保护持续加码：**随着《未成年人网络保护条例》的实施，游戏开发者需要加强对未成年人在线时长、社区互动和虚拟币充值的管理，同时探索多种手段保护，以确保未成年人的健康成长。

**用户生成内容（UGC）的风险管理日益重要：**开放世界游戏玩家可以自主创意和产出内容，但这也可能导致不合法或不合规的情况出现。游戏平台需要重视并管理这些 UGC 内容，防止破坏社会风气和影响游戏体验。

**出海内容合规依然是重中之重：**不同国家和地区对游戏内容的审查标准存在差异，同时游戏公司需要根据当地的法规和文化习惯进行内容审核和修改，从而避免造成不必要的损失。

**营销欺诈风险增多：**虚假买量、虚假宣传等问题随着游戏市场竞争加剧会更加频发，游戏公司需要加强第三方买量平台合规性管理以及买量数据真实性审查，确保营销活动的合规性和效果。

网易易盾