



# 中国科技企业如何借鉴和应用 「服务儿童的负责任科技创新」和 「设计保障儿童安全」原则

2026年1月

## 免责声明

本系列案例研究是在中国网络社会组织联合会 (CFIS)、中国传媒大学 (CUC) 与参与研究的中国企业共同开展的“服务未成年人的负责任科技创新”并关注“设计保障儿童安全”的企业案例研究项目框架下完成的,并得到联合国儿童基金会驻华办事处的支持。研究所需信息通过两种方式获取:一是企业自愿提供的披露信息,二是公开可获取的信息。案例研究中作为分析对象的产品,其筛选工作由联合国儿童基金会的咨询合作方——BSR咨询负责开展。

联合国儿童基金会不代言任何公司、品牌、产品或服务。

中国网络社会组织联合会和中国传媒大学不代言任何公司、品牌、产品或服务。

本报告为联合国儿童基金会出版物。对各企业的致谢并不意味着本系列案例研究已获得各企业的认可或支持。文中提及任何特定企业,也不代表联合国儿童基金会认同该企业的政策、产品及实践。这些案例研究旨在探讨中国科技公司如何在实践中应用“设计保障儿童安全”原则。研究结果仅用于展示当前的实践情况,并不旨在对相关主题的其他研究或行业实践提供全面或定论性的观点。

本研究中的引述内容及案例仅用于凸显儿童面临的潜在风险或可获得的机遇。本系列案例研究并非旨在开展正式风险评估,也不代表对相关产品在儿童安全方面所涉风险与机遇的全面呈现。同时,案例研究亦不试图对所涵盖领域内的研究成果或行业实践进行全面阐述或得出定论。此外,本系列案例研究既未阐明,也不代表联合国儿童基金会的任何官方立场。

## 致谢

本出版物由联合国儿童基金会驻华办事处、中国网络社会组织联合会和中国传媒大学合作完成。

共有19家中国科技公司回应了公开征集案例的邀请,其中4家公司完成了自我评估流程。

本系列案例研究报告由来自BSR咨询的J.Y. Hoh撰写。

衷心感谢以下人士为本报告草案提供了有益的意见和建议:联合国儿童基金会因诺琴蒂全球研究与前瞻办公室的史蒂文·沃斯卢 (Steven Vosloo) 和丹尼尔·卡德费尔特 - 温瑟 (Daniel Kardefelt-Winther)、联合国儿童基金会东亚及太平洋区域办公室 (UNICEF EAPRO) 的亚历山德拉·切尔尼亚夫斯卡娅 (Alexandra Chernyavskaya)、联合国儿童基金会儿童权利与商业团队的乔安妮·加利亚·巴伦 (Josianne Galea Baron), 以及其他相关人士。

版权所有 © 联合国儿童基金会驻华办事处, 2026年1月

版权所有 © 中国网络社会组织联合会, 2026年1月

版权所有 © 中国传媒大学, 2026年1月

保留所有权利。

转载本出版物的任何部分, 均需获得许可。

# 目录

案例研究综述 .....	4
百度“文心一言”大模型 .....	9
腾讯未成年人保护中心 .....	18
猿编程APP .....	27
核桃编程3D图形化编辑平台 .....	35
参考文献 .....	45

## 案例研究综述

在本研究报告中,对“设计保障儿童安全”<sup>1</sup>采用了简化定义。“设计保障儿童安全”是指企业在产品开发生命周期的全过程中,主动降低对儿童的负面影响,并最大限度提升积极影响。

本系列包含四项案例研究,探讨了中国科技企业如何将“设计保障儿童安全”原则付诸实践,并提供了切实可行的洞见,助力为儿童打造更安全的数字环境。

这些案例研究的目的具有双重性:

- 其一,举例说明如何在产品生命周期的早期阶段就融入儿童安全考量,并在整个周期内持续贯彻这一理念。
- 其二,展示如何运用国际儿童权利与人权标准(例如《联合国工商业与人权指导原则》(UNGPs)<sup>2</sup>以及《服务儿童的负责任科技创新八项原则》(RITEC))<sup>3</sup>识别并缓解儿童安全相关风险。

**需明确的是,这些案例研究系独立开展。虽然该研究项目得到中国网络社会组织联合会和联合国儿童基金会驻华办事处的支持,并受到联合国儿童基金**

**会(UNICEF)的全球“服务儿童的负责任科技创新”(RITEC)研究的启发,但并不是该全球研究项目的一部分。另需说明的是,尽管研究聚焦于“设计保障安全”,但在分析过程中也纳入了更广泛的儿童福祉框架。**

本系列案例研究聚焦中国多家头部科技企业的产品,涉及腾讯、猿编程、核桃编程及百度,具体包括:

- 1.一款通用型生成式人工智能聊天机器人(百度公司“文心一言”)。
- 2.一个集成于各类数字服务(尤其游戏领域)的服务平台(腾讯公司“未成年人保护中心”)。
- 3.一个编程教育平台(猿编程“猿编程App”)。
- 4.一个具备教育与社区功能的在线平台(核桃编程“3D图形化编辑平台”)。

1.关于“设计保障儿童安全”,目前已有多种定义。经济合作与发展组织(OECD)在《面向儿童的数字安全设计圆桌会议报告》(2023年)中对此有所阐述。[https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/towards-digital-safety-by-design-for-children\\_f1c86498/c167b650-en.pdf#page=7](https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/towards-digital-safety-by-design-for-children_f1c86498/c167b650-en.pdf#page=7) OECD将“设计保障儿童安全”定义为:在数字产品或服务的设计、开发及部署全生命周期中,安全均被视为“首要考量因素”。该组织强调,安全的责任“不应由儿童或其家长承担”——“安全是企业责任的核心组成部分,且决策需以儿童为中心”。更多相关内容可参见S. Livingstone与K. Pothong合著的《设计保障儿童权利:儿童使用的数字产品与服务创新者指南》(2023年,数字未来委员会出版)。[https://digitalfuturescommission.org.uk/wp-content/uploads/2023/03/CRbD\\_report-FINAL-Online.pdf](https://digitalfuturescommission.org.uk/wp-content/uploads/2023/03/CRbD_report-FINAL-Online.pdf)

2.联合国(2011),《工商业与人权指导原则:落实联合国“保护、尊重和补救”框架》,联合国人权事务高级专员办事处。[https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr\\_en.pdf](https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf)

3.联合国儿童基金会因诺琴蒂研究办公室(2024),儿童科技负责任创新(RITEC):数字技术、游戏与儿童福祉(第二阶段报告),获取途径:<https://www.unicef.org/innocenti/reports/responsible-innovation-technology-children>

为确定并筛选本系列案例研究的产品，中国网络社会组织联合会 (CFIS) 与联合国儿童基金会驻华办事处及中国传媒大学 (CUC) 合作，面向中国各地科技企业发出公开征集通知。在初审阶段共收到19个产品的申报材料，其中4个产品的申报材料经筛选后进入下一阶段。

本系列案例研究的产品筛选基于多项标准。首先，这些产品覆盖多个拥有大量儿童用户的科技细分领域及应用场景，具体包括生成式人工智能聊天机器人、教育科技、社交媒体以及家长对游戏的管控工具。其次，所选产品中包含生成式人工智能等新技术——这类技术虽有望为提升儿童福祉创造机遇，但也可能给儿童安全带来新的风险。尽管这些案例研究并非正式的风险评估报告，但它们通过典型案例呈现了部分此类风险，同时也介绍了企业为识别并应对这些风险所采取的举措。

产品选定后，研究团队要求各参与企业填写一套自我评估问卷。该问卷旨在收集详细信息，了解各企业的产品在多大程度上符合“设计保障儿童安全”原则，并体现“**服务儿童的负责任科技创新**” (RITEC) 所明确

的儿童福祉核心维度。随后，联合国儿童基金会依据针对各企业产品特性定制的研究框架，邀请每家企业补充提供相关信息。

### 一个借鉴国际儿童权利、设计保障安全及福祉标准的概念框架

本案例研究所采用的概念框架，借鉴了国际组织制定的一系列儿童安全与福祉标准。其中，联合国儿童基金会与乐高基金会依据《儿童权利与商业原则》共同制定了“**服务儿童的负责任科技创新**” (RITEC) 框架，<sup>4</sup> 该框架包含八项与儿童福祉及儿童权利相关的原则，具体为自主性，能力培养，情感关怀，人际关系，创造力，身份认同，多样性、公平性与包容性，安全保障。《联合国工商业与人权指导原则》(UNGPs)<sup>5</sup> 则为企业如何开展人权及儿童权利尽职调查提供了权威方法指南。《儿童权利公约》<sup>6</sup> 及其配套的《关于儿童权利与数字环境的一般性意见》，<sup>7</sup> 详细规定了在儿童权利领域各方应承担的义务。此外，经济合作与发展组织 (OECD) 也发布了多份关于儿童安全的重要报告，<sup>8</sup> 内容涵盖儿童安全风险分类，以及面向儿童的数字安全设计指南。<sup>9</sup>

4. 联合国儿童基金会因诺琴蒂研究办公室(2024), 儿童科技负责任创新 (RITEC): 数字技术、游戏与儿童福祉 (第二阶段报告), 获取途径: <https://www.unicef.org/innocenti/reports/responsible-innovation-technology-children>

5. 联合国 (2011年), 《工商业与人权指导原则: 落实联合国“保护、尊重和补救”框架》, 联合国人权事务高级专员办事处。 [https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr\\_en.pdf](https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf)

6. 联合国 (1989年), 《儿童权利公约》, 联合国大会第44/25号决议。 <https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child>

7. 儿童权利委员会 (2021年), 《第25号一般性意见 (2021年): 儿童权利与数字环境》 (文件编号: CRC/C/GC/25), 联合国。 <https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation>

8. OECD (2021年), 《数字环境中的儿童: 经修订的风险分类》 (《经合组织数字经济论文集》第302号), 经合组织出版社。 <https://doi.org/10.1787/9b8f222e-en>; Livingstone, S., & Stoilova, M. (2021), 《4C框架: 儿童网络风险分类》 (CO:RE核心议题简报系列)。汉堡: 莱布尼茨媒体研究中心 | 汉斯-布勒多研究所 (HBI); CO:RE (儿童网络研究与证据项目)。 <https://doi.org/10.21241/ssoar.71817> SSOAR

9. OECD (2024年), 《迈向“设计保障儿童数字安全”》 (《经合组织数字经济论文集》), 经合组织出版社。 <https://doi.org/10.1787/c167b650-en>

在这些标准的基础上，研究团队制定了一个实用框架，用于指导企业参与及相关分析工作。

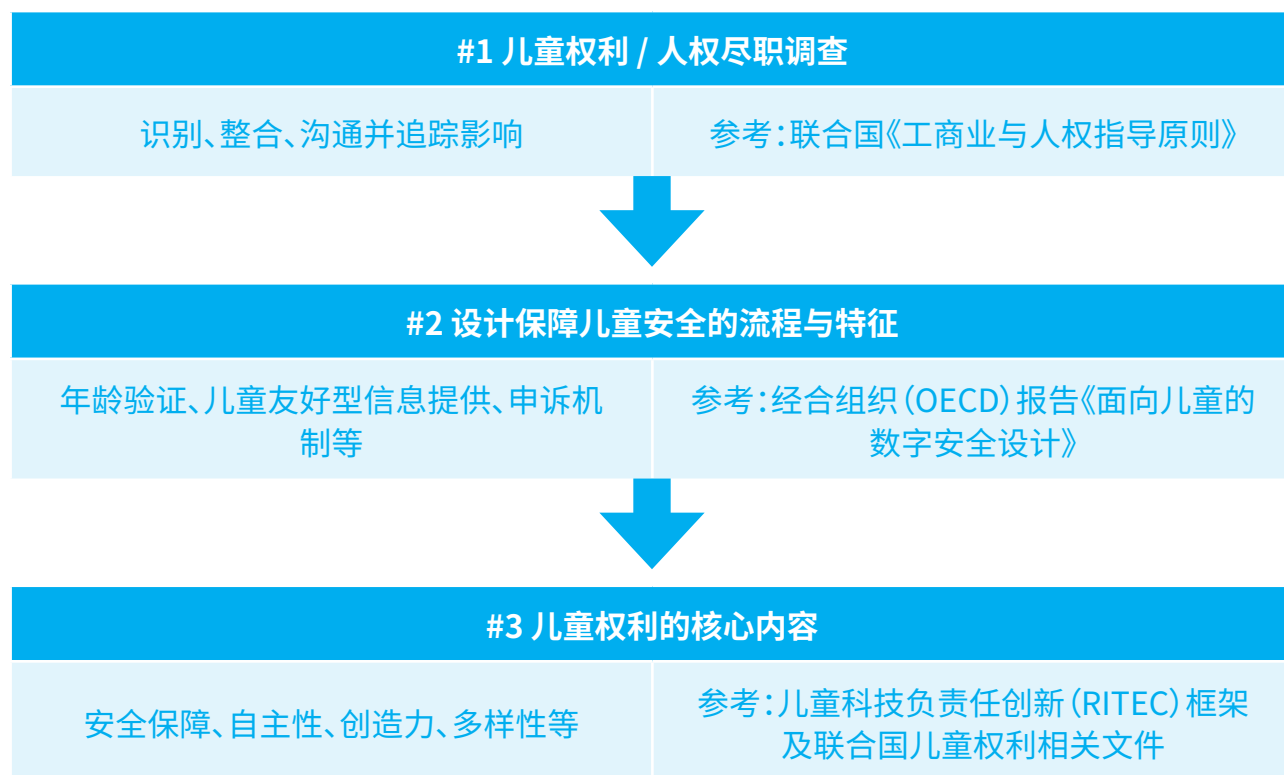
本研究的创新之处在于，将这些国际标准转化为一个包含结构化问题的实用框架，以指导企业的运营行为。该框架的设计思路是：将这些标准中经过调整的核心要素融入结构化研究问题中。在回答这些问题的过程中，以及通过与研究团队进行多轮后续追问，企业能够逐步熟悉这些国际标准的具体内容、要求，以及这些标准在科技产品开发与部署环节的应用方法。

企业的答复已整理提炼为所附的四份案例研究。这些案例研究既展示了应用该框架如何帮助企业实践与国际标准接轨，也呈现了风险识别、风险缓解的实际案例，以及有待改进的领域。其他企业亦可运用这一概念框架评估自身当前状况，并制定进一步与国际标准接轨的计划。

### 基于国际标准的三部分框架

概念框架第一部分：聚焦企业尽职调查流程。通过联合国儿童基金会与BSR咨询设计的结构化问题，企业需思考如何纳入《工商业与人权指导原则》(UNGPs)中明确的人权尽职调查四大核心要素。《工商业与人权指导原则》是关于负责任商业行为与人权的权威国际标准，在通过“设计保障安全”机制维护儿童权利的过程中，该原则同样适用。其中第17条原则明确了企业尊重人权义务的四大要素，具体如下：

1. 识别并评估风险。
2. 整合风险评估得出的结论。
3. 追踪并衡量风险缓解措施的有效性。
4. 就风险及缓解措施进行沟通。



在简化后的产品生命周期的三个阶段中,每个阶段均对上述各项要素进行了评估,<sup>10</sup>具体如下:

- 1.产品构思阶段
- 2.产品开发与测试阶段
- 3.销售、营销与部署阶段。

因此,该概念框架的结构要求企业确保将儿童权利尽职调查的核心要素融入产品全生命周期。这种覆盖全生命周期的方法对于“设计保障儿童安全”至关重要,因为该理念强调主动而非被动地采取风险缓解措施。

概念框架第二部分聚焦于确保企业已考量经济合作与发展组织(OECD)2024年报告《面向儿童的数字安全设计》中明确的“设计保障儿童安全”特定要素。<sup>11</sup>这些要素涵盖产品功能与企业流程两类范畴。企业需说明其产品或流程如何纳入以下要素:

- 1.年龄验证机制
- 2.投诉与补救机制
- 3.儿童友好型信息提供
- 4.企业决策过程中儿童参与或听取儿童意见的机制。

通过与联合国儿童基金会和BSR咨询开展又一轮迭代式追问,企业逐渐熟悉了这些理念,以及在实际运营中落实这些理念所涉及的复杂性。例如,实施有效的年龄验证机制,需要在准确识别用户年龄与尊重用

户隐私之间取得平衡。案例研究既展示了企业如何尝试实现这一平衡,也针对“为与国际标准接轨需进一步考量的领域”提供了建议。

概念框架第三部分聚焦“设计保障儿童安全”的核心内涵,以及儿童福祉相关的风险与机遇。其中,“核心内涵”指设计保障儿童安全及儿童福祉所保护的特定权益,例如自主性或安全保障。框架要求企业识别其产品可能带来的风险,尤其是对儿童用户安全保障构成的风险。同时,企业还需说明其产品或服务在安全保障之外的领域(如儿童的自主性、创造力或多样性发展)可能产生的作用。

本节的整体框架借鉴了“服务儿童的负责任科技创新”(RITEC)框架中提出的八个维度。<sup>12</sup>RITEC项目由联合国儿童基金会、乐高基金会,以及多家学术与研究中心组成的联盟共同发起;其中,联合国儿童基金会多年来在推动“设计保障儿童福祉”方面发挥了主导作用(点击此处查看其他出版物)。

“服务儿童的负责任科技创新”(RITEC)框架的核心贡献之一,是明确了八个可融入产品及体验设计的维度——通过数字玩乐促进儿童福祉的设计需涵盖这些维度,具体如下:

- **自主性**——儿童能体验到掌控感与能动性,并享有选择自由。
- **能力培养**——儿童能感受到自身的掌控能力,且确信自己有能力达成目标。

10.这是对科技公司产品开发生命周期各阶段的简化概括。不同公司对生命周期阶段的划分方式可能有所不同。

11.OECD(2024年),《迈向“设计保障儿童数字安全”》(《经合组织数字经济论文集》),经合组织出版社。<https://doi.org/10.1787/c167b650-en>

12.联合国儿童基金会因诺琴蒂研究办公室(2024年),《儿童科技负责任创新:数字技术、玩乐与儿童福祉(第二阶段报告)》。获取途径:<https://www.unicef.org/innocenti/reports/responsible-innovation-technology-children>

- **情感关怀**——儿童能体验、感知各类情感,并具备调节这些情感的能力。
- **人际关系**——儿童能体验并维系社交联结,拥有归属感,且能关注他人。
- **创造力**——儿童愿意接纳多样体验,能想象不同可能性,会依据原创想法采取行动,并动手创造事物。
- **身份认同**——儿童能探索、构建并展现自身及他人的多方面特质。
- **多样性、公平性与包容性**——儿童能感受到自身的多样性在包容的环境中得到尊重。
- **安全保障**——儿童能得到有力保护,免受各类安全风险的侵害。

尽管“服务儿童的负责任科技创新”(RITEC)框架最初是在数字玩乐领域提出的,但本研究中的案例将其原则应用到了不同的科技细分领域。例如,人际关系、创造力与自主性这三个维度,对于社交媒体、游戏及教育科技应用而言均具有高度相关性。

研究通过结构化问题引导企业说明:在上述八个维度中,与其产品相关的部分维度里,其产品如何产生风险或创造机遇。案例研究特别强调了“安全保障”维度的重要性——鉴于该维度与“设计保障儿童安全”理念存在独特的关联性。

该概念框架对“安全保障”维度的探讨,进一步采用了索尼娅·利文斯通<sup>13</sup>提出的四分类法进行细分——该分类法亦被收录并概述于经济合作与发展组织(OECD)2021年报告《数字环境中的儿童:风险分类修订版》中。<sup>14</sup>此分类法不仅对数字风险领域进行了宏观概述,还明确了四大风险类别,具体包括:接触风险、行为风险、内容风险及消费风险;同时,它还识别并分析了横跨这四大类别的交叉风险。目前,全球科技公司的信任与安全团队广泛运用该分类法,为制定和完善儿童保护政策及项目提供依据与支持。在本系列案例研究的撰写过程中,企业需考量部分此类风险(例如其产品可能引发的接触风险与行为风险),并说明已采取何种措施缓解这些风险。

13.OECD (2021年),《数字环境中的儿童:风险分类修订版》(《经合组织数字经济论文集第302号》),经合组织出版社。<https://doi.org/10.1787/9b8f222e-en>; Livingstone, S., & Stoilova, M. (2021).《4C框架:儿童网络风险分类》(CO:RE核心议题简报系列)。汉堡:莱布尼茨媒体研究所 | 汉斯-布吕多研究所(HBI); CO:RE——儿童网络研究与证据项目。<https://doi.org/10.21241/ss0ar.71817> SSOAR

14.OECD (2021年),《数字环境中的儿童:风险分类修订版》(《经合组织数字经济论文集第302号》),经合组织出版社。<https://doi.org/10.1787/9b8f222e-en>

## 百度“文心一言”大模型

百度推出的“文心一言”是一款生成式人工智能聊天机器人，其主要面向中国内地使用普通话的用户群体。该机器人采用先进的多模态人工智能技术，能够对文本、图像、视频及代码类输入作出响应。

本案例研究探讨了百度如何将儿童福祉理念融入“文心一言”的设计中，并重点关注以下三个维度：

- **安全保障**：确保儿童使用该机器人时能感受到安全
- **创造力**：支持儿童的好奇心与富有想象力的学习过程
- **自主性**：让儿童能够自主做出选择，并掌控自身的学习体验

本案例研究重点阐述了百度如何采用儿童安全影响评估、内容过滤与隐私保护措施，以及面向儿童的交互式教育工具。研究还探讨了百度在平衡丰富内容与安全性方面面临的挑战，并提出了改进建议。<sup>1</sup>

### 免责声明：

联合国儿童基金会不代言任何公司、品牌、产品或服务。

本报告为联合国儿童基金会出版物。对各企业的致谢并不意味着本系列案例研究已获得各企业的认可或支持。文中提及任何特定企业，也不代表联合国儿童基金会代言该企业的政策、产品及实践。

## 概述

生成式人工智能借助机器学习从现有数据中挖掘规律，并生成新内容。它能够生成类人文本、图像、音频与代码，还可支持复杂任务的完成。

生成式人工智能可为儿童学习提供支持、培养其创造力并提升沟通能力。然而，儿童也可能面临更高风险，例如接触有害或不当内容、在人际互动中遭受操

15. 本研究借鉴了另一专项项目——“服务儿童的负责任科技创新”（Responsible Innovation in Technology for Children, RITEC）。该项目由联合国儿童基金会（UNICEF）与乐高集团（The LEGO Group）合作开展，资金来源于乐高基金会（The LEGO Foundation）。项目明确了八项原则（自主性，能力培养，情感关怀，人际关系，创造力，身份认同，安全保障，多样性、公平性与包容性），这些原则可用于评估儿童福祉与儿童权利。更多相关信息详见：<https://www.unicef.org/childrightsandbusiness/workstreams/responsible-technology/online-gaming/ritec-design-toolbox>

纵或侵害，或是个人数据被滥用。因此，在设计人工智能产品时将儿童安全纳入考量，对于充分发挥其积极效益至关重要。

本案例研究聚焦百度“文心一言”与儿童福祉相关的三个核心维度：

- **安全保障：**百度通过儿童权利影响评估，识别儿童使用过程中可能面临的风险，例如接触暴力内容、歧视信息及由人工智能“凭空捏造”的虚假信息。百度借助内容过滤与数据隐私协议应对这些风险，同时结合人工监督与人工智能驱动的检测技术，形成风险管控机制。

- **创造力：**“文心一言”为儿童提供虚拟实验与创意活动，助力儿童在实践中学习，进而培养其学习能力与创新思维。
- **自主性：**“文心一言”通过个性化指导，为儿童的自主学习提供支持，帮助其养成独立学习习惯。

百度在平衡内容丰富性与简洁性、安全性方面面临挑战。由于生成式人工智能采用概率算法，有时可能生成不当内容，因此内容过滤与审核机制需持续优化；同时，儿童需求的动态变化也要求安全措施不断接受评估与调整。对此，百度可通过以下方式进一步完善其策略：与儿童安全领域专家及其他外部相关方开展合作，并让儿童直接参与产品设计过程。

## 儿童福祉维度的实践应用

### 维度1:安全保障——儿童能够感受到安全,且其安全得到保障.

本案例研究探讨了百度为确保儿童安全保障所实施流程中的关键要素：

安全保障流程	
风险评估	识别儿童安全影响，进而整合、跟踪风险防控措施并就其开展沟通。
年龄验证	估算或核实用户年龄的流程，以便数字服务能根据儿童需求进行定制，并落实符合其年龄段的保护措施。
投诉与补救	儿童可通过该途径举报任何侵犯儿童权利的行为，并申请补救。
儿童友好型信息提供	以儿童易于接受的语言及方式传递信息（例如使用清晰、易懂、简短的语句，仅在需要儿童做出决策时提供必要信息，且采用交互式工具），提升儿童的关注意愿。
儿童参与企业决策	咨询儿童意见，并将其观点纳入数字服务的设计与部署过程。

## 风险评估

风险评估应在产品生命周期的各个关键阶段开展：<sup>16</sup>

1. 产品构思阶段：企业对产品进行概念构想、调研并明确产品价值主张，同时制定产品设计方案与概念的阶段。
2. 产品开发与测试阶段：企业进行产品开发与迭代、制作原型，并在产品上线前开展测试的阶段。
3. 销售、营销与部署阶段：企业将产品推向市场、推动产品部署，并在部署后监测产品使用情况以进行后续迭代的阶段。

百度在产品构思阶段便引入“设计保障儿童安全”理念。产品经理、儿童安全专家与合规团队会协作开展儿童权利影响评估，识别可能影响儿童用户的实际风险与潜在风险。评估内容包括产品功能分析等，以此研判隐私泄露、成瘾、接触不当内容等潜在风险，并特别关注儿童的认知水平与差异化需求。

根据评估结果，“文心一言”的拟人化（模拟人类）沟通风格可能导致儿童对人工智能产生情感依赖或过度信任，且儿童在与该聊天机器人互动过程中可能无意间泄露个人信息。基于上述评估结果，设计团队与产品开发团队共同制定了设计规范，将风险防控措施融入产品设计。这些措施包括符合年龄段的内容过滤功能、标注人工智能生成内容的水印功能，以及隐私保护功能。此外，还通过模拟测试与专家评审，对设计方案中风险防控措施的有效性进行初步评估。

百度表示，“设计保障儿童安全”的理念在开发与测试阶段持续落地：开发团队与儿童安全专家会开展代码审查及隐私影响评估，识别儿童使用产品时可能面临的风险（如数据泄露、产品滥用等）。基于评估结果，开发团队会在产品中集成年龄验证、隐私保护等功能，以防控潜在风险。同时，本阶段还会开发并调试内容审核系统，确保其能够识别并过滤对儿童不适宜的内容。用户可通过拨打百度举报热线、发送邮件或点击应用内的在线举报按钮，反馈违规问题；百度客服提供24小时全天候服务，处理用户反馈。

百度指出，在销售、营销与部署阶段，营销团队与儿童安全专家会评估营销策略可能对儿童权利产生的风险。例如，他们会核查广告是否包含涉性暗示的画面或语言，也会评估广告是否采用针对儿童的诱导性手段（如使用动画角色、名人形象，或引导儿童用户要求家长购买产品等）。此外，数据隐私团队会评估产品部署过程中个人数据泄露的风险。针对风险防控措施的有效性，团队会通过自动化测试及模拟儿童用户行为的方式进行评估。

### “设计保障儿童安全”理念在其他流程层面的体现

仅通过单一举措，难以实现面向儿童的“设计保障数字安全”；相反，这需要多项互补且相互关联的举措——这些举措协同作用，方能助力为儿童营造安全且有益的数字环境。

——迈向“设计保障儿童数字安全”<sup>17</sup>

16. 这是对科技公司产品开发生命周期各阶段的简化概括。不同公司对生命周期阶段的划分可能有所不同。

17. OECD(2024),《迈向“设计保障儿童数字安全”》(经济合作与发展组织《数字经济文件》), 经济合作与发展组织出版, <https://doi.org/10.1787/c167b650-en>, 第30页。

风险评估是“设计保障儿童安全”理念的核心环节,在风险识别、风险防控以及防控措施有效性的跟踪与沟通方面发挥关键作用。然而,正如经济合作与发展组织(OECD)所指出的,企业还需实施其他流程,以确保风险能够得到恰当识别与应对。本案例研究重点介绍了其中四类流程:<sup>18</sup>

- 1) 年龄验证机制
- 2) 儿童友好型信息提供
- 3) 投诉与补救机制
- 4) 企业决策过程中纳入儿童参与或听取儿童意见的机制

### 1.百度的年龄验证机制

在“设计保障数字安全”理念中,确认用户年龄至关重要:只有明确哪些用户实为儿童,相关服务才能切实保护儿童免受不安全内容与行为的影响。

——迈向“设计保障儿童数字安全”<sup>19</sup>

尽管年龄验证机制是儿童安全保障的重要组成部分,但国际标准要求此类机制在设计时,需充分平衡儿童用户的各项权利(如隐私权、信息获取权等)。百度已在其所有产品(包括“文心一言”)中部署了统一的年

龄验证机制:儿童无法自行创建账号,必须由家长或监护人代为注册。注册时,家长(或监护人)与儿童均需阅读用户协议,协议中列明了将收集和处理的的所有数据类型(下文将进一步说明)。然而,目前尚未对该机制的实际效果进行评估。

要使用“文心一言”,必须先创建账号。在账号创建过程中,百度会收集经过模糊处理的年龄段数据,以判断用户是否未满14周岁。例如,百度可能会利用注册用户自愿填写的申报信息,如年龄段或部分出生日期(月份和年份)。百度表示,其不会收集或存储面部特征、指纹等数据;应用后台不会留存数据,且用于年龄验证的用户年龄相关信息在验证完成后即会被删除。

### 2.百度的儿童友好型信息提供

儿童友好型信息提供服务有助于儿童理解:他们所使用的各类数字产品及服务相关的服务条款、这些产品与服务可能存在的风险,以及如何举报伤害行为、标记不安全或非法内容。

——迈向“设计保障儿童数字安全”<sup>20</sup>

在开发阶段,“文心一言”的设计团队对该产品用户界面中的信息呈现方式进行了考量,以确保儿童能够轻松理解相关内容。例如,其语音训练人工智能功能采用清晰直观的界面设计,并借助语音引导与视觉提示,帮助儿童轻松理解训练内容(详见下文)。

18.需说明的是,本案例研究为简洁起见,对这些流程的分类与经济合作与发展组织(OECD)的分类略有不同。OECD的流程分类体系还包含用于判定伤害的技术工具及以儿童为中心的设计;这些内容的相关方面已纳入本风险评估的“实质内容”部分。

19.OECD(2024),《迈向“设计保障儿童数字安全”》(经济合作与发展组织《数字经济文件》),经济合作与发展组织出版,<https://doi.org/10.1787/c167b650-en>,第30页。

20.同上注,第38页。

此外，百度的产品团队与其用户体验团队合作，开展了专为儿童定制的交互设计测试。团队进行了多轮用户测试，以确保信息呈现方式符合儿童的认知能力，并通过互动动画、图标等方式，帮助儿童理解产品的操作流程。



© 图片由百度提供  
“文心一言”语音训练人工智能的用户界面

### 3. 百度的投诉与补救机制

帮助儿童了解如何举报事件与伤害行为、标记不安全或非法内容，这一点至关重要。若投诉机制难以被儿童找到、获取或使用，其实际效用将大打折扣……因此，儿童友好型投诉机制必不可少。

——迈向“设计保障儿童数字安全”<sup>21</sup>

百度表示，其已为“文心一言”部署了举报机制，并提供全天候客服服务处理用户反馈，其中包括涉及不良内容的反馈。举报投诉类别涵盖不良信息、色情低俗内容，以及辱骂、冒犯性行为。针对性侵相关的举报，将启动升级处理流程并予以优先处置。用户只需点击应用内的举报按钮，即可一键提交违规举报。百度称，其客服团队会在3个工作日内对投诉作出回应，并设有申诉渠道，以确保问题得到彻底解决。用户可就百度的初步反馈或处理结果提出申诉，百度会指派专人对该投诉进行跟进并进一步复核，确保问题最终解决。

### 4. 百度在企业决策中纳入儿童参与或听取儿童意见的机制

在涉及儿童自身的事务中，儿童享有表达意见的权利。当儿童(及青少年)参与到影响自身的项目与政策制定过程中时，这些项目与政策更有可能更好地契合儿童的需求、兴趣与背景。

——迈向“设计保障儿童数字安全”<sup>22</sup>

21. OECD(2024),《迈向“设计保障儿童数字安全”》(经济合作与发展组织《数字经济文件》), 经济合作与发展组织出版, <https://doi.org/10.1787/c167b650-en>, 第30页。

22.同上注, 第40页。

百度表示，其目前会通过咨询家长来了解儿童的需求与偏好。然而，尽管家长的反馈具有价值，但与儿童直接沟通才是获取其观点最准确反馈的理想方式。未来在可行的情况下，百度应考虑与儿童开展直接沟通。

### 百度“文心一言”存在的儿童安全风险

本节阐述“文心一言”存在的部分儿童安全风险，以及前文所述的“设计保障儿童安全”流程如何帮助百度防控这些风险。需要说明的是，本节列举的安全风险示例仅具说明性，并非详尽无遗。

生成式人工智能聊天机器人会给儿童带来多方面的安全风险。这类聊天机器人可能生成有害内容，例如包含暴力、血腥元素的图像或文本，以及涉性内容；还可能被用于生成儿童性侵材料。此外，生成式人工智能聊天机器人可能产生有害偏见——由于会吸收并再现训练数据中存在的偏见，使用这类应用的儿童可能会无意识地内化机器人所传递的偏见或刻板印象。

聊天机器人还可能通过提供危险指令带来额外风险：它们可能教唆儿童伤害自己或他人，在某些极端案例中，甚至曾诱导用户自杀。

百度试图借助“文心一言”等大型人工智能模型的能力，对旗下产品中涉及儿童安全的有害内容进行审查与过滤，以此防控上述风险。尽管大语言模型仍存在局限性，但其自然语言处理能力可用于检测和过滤某些可能对儿童造成危害的内容与行为。下文示例将展示，百度的生成式人工智能系统会尝试识别并过滤不良分子用于伪装诱骗儿童意图的各类暗语。然而，如何应对这些风险，对百度及其他科技公司而言，仍是一项重要挑战。

**安全保障：运用生成式人工智能模型检测儿童性侵行为**

**标题：**回复：求助一下吧

**thread\_content：**找一个温柔，可以长期的zhu（三声）本人07，0爱慕

**回复内容：**1111

**一级评论：**

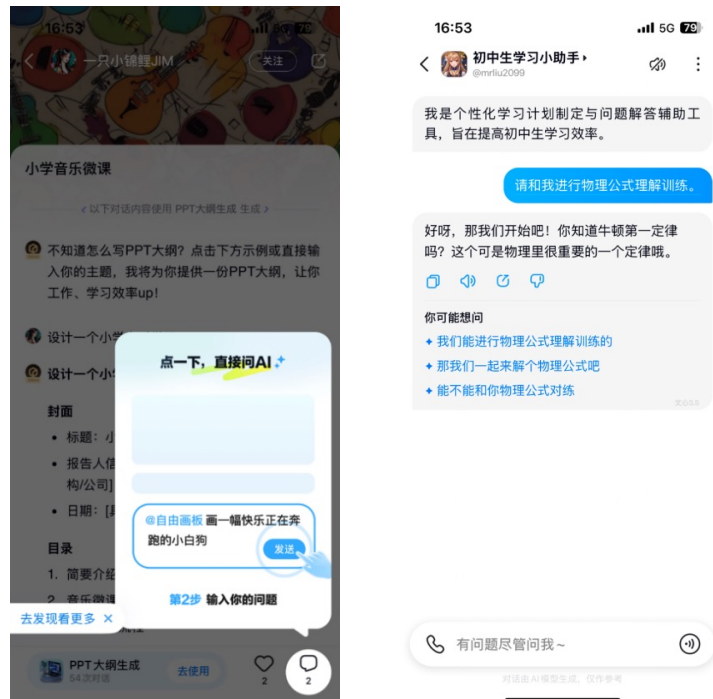
**备注说明：**1170:首先，文本中提到了“07”，暗示了年龄小于18岁的学生身份。其次，文本中提到了“zhu（三声）本人”，其中的“zhu（三声）”可能是指“主-仆”关系，涉及到特殊行为。因此，根据审核规则，该文本内容存在风险。

© 图片由百度提供  
此图展示了大语言模型如何检测网络上性犯罪者所使用的暗语

## 维度2:创造力——儿童用户会怀着好奇心参与其中,并运用想象力进行构建、创造与尝试。

### 在“文心一言”的设计中融入创造力

百度表示,“文心一言”包含专为儿童设计的互动工具,例如虚拟实验和创意游戏。这些功能旨在助力儿童参与科学及艺术相关主题的活动。用户可使用题库、概念解析、个性化学习计划等功能,这些功能或能契合不同用户的学习偏好。



© 图片由百度提供  
“文心一言”的用户界面

## 维度3:自主性——儿童用户可自由选择参与方式,并能体验到自主能动性、选择权与自由度带来的感受。

### 在“文心一言”的设计中融入自主性

百度表示,“文心一言”支持儿童自主发展:儿童可自行设定学习目标,并根据个人需求调整难度级别,这有助于提升他们的自主学习能力。此外,“文心一言”应用还为中小學生提供了种类丰富的教育类人工智能助手,儿童可据此选择符合自身认知水平的学习或娱乐选项。

百度称,“文心一言”还通过自动提供引导式辅助,进一步支持儿童用户的自主性。该聊天机器人可进行个性化设置,在提供反馈时避免直接向学生灌输答案,从而让学生在遇到困难时仍能自主提升解决问题的能力。这种方式不仅可能提高学生的学习积极性与自信心,还有助于培养他们在专门课堂学习之外的自主学习能力。

## 保护儿童福祉与安全所面临的挑战及建议

在撰写本案例研究的过程中，我们识别出了若干关键挑战，并针对性地提出了建议，以供相关企业参考。需说明的是，这些问题在其他致力于保护儿童福祉与安全的科技企业中也普遍存在。同时必须指出，目前并不存在万能解决方案。

**挑战：**“文心一言”需在维持丰富多样功能的同时，确保为儿童提供的信息具备简洁性——而信息简洁正是“设计保障儿童安全”理念的核心要素，这一平衡存在实施难度。

### 建议：

- 考虑通过调研、咨询专家与儿童，以及反复测试和调整，全面了解儿童的认知水平与心理特征。
- 基于实证，考虑实施分层信息设计：根据不同年龄段调整内容复杂度；增设自适应模式切换功能（例如，专门的儿童模式提供简化的交互过程，而高级模式需经家长授权方可开启）；采用任务导向型交互方式，优先满足儿童的核心使用场景（如作业辅导、事实查询等），减少不必要的复杂度。

**挑战：**尽管大语言模型生成的内容能提供知识与信息，但生成式模型的概率算法可能在无意间生成不当内容。偏见相关问题便是例证之一。例如，选择年轻且符合传统审美标准的女性形象作为人工智能助手，或许能获得用户的正面反馈，但也可能强化“女性助理”相关的刻板印象。

### 建议：

- 确保训练数据在性别、年龄、种族等维度上具备多样性与平衡性；将人工智能助手的拟人形象设

计为中立或多样化风格，避免刻板化；同时加强实时内容过滤与人工审核流程。

- 定期对人工智能输出内容进行审计，有助于发现新出现的偏见或问题内容；而面向儿童及监护人的用户反馈机制，可提供额外参考信息，为模型的持续优化提供指导。
- 对当前所采用的技术方法及用户体验开展更多评估工作。

**挑战：**另一风险与拟人化相关：低龄儿童可能会认为有拟人形象的人工智能助手更可信，进而降低对其输出信息的验证意愿。儿童尤其容易将人工智能聊天机器人视为“近似人类的知心伙伴”。但在需要同理心与理解力的场景中，聊天机器人可能无法识别危险情况，或提供不当建议；它们也可能无法给出符合儿童发展阶段的适宜建议。这可能导致儿童产生困惑、沮丧情绪，甚至遭受伤害。

### 建议：

- 采用非拟人化的形象，搭配客观、中立的语言，降低其被感知到的人类属性。
- 另一种有效方法是植入透明度提示，定期提醒儿童该人工智能并非人类，且可能出现错误。
- 加入鼓励验证信息的提示，例如“向可信赖的成年人确认”；或提供信息来源、配备家长指导工具，帮助儿童了解人工智能的局限性，以及交叉验证信息的重要性。

**挑战：**在安全性与创新性之间取得平衡。企业需保护儿童免受有害内容侵害，同时也要支持开放的学习与探索过程。

**建议:**

- 需认识到儿童安全并非单纯的技术问题,而是需要包括企业员工、用户在内的广泛利益相关方共同承担责任并高度重视。企业应优先推动跨团队协作、加强与家长及儿童的沟通,并严格遵守相关政策法规。
- 考虑纳入预测性策略(如未来情景分析法),同时参考同行企业的政策与实践经验。
- 制定政策时,需考虑到儿童用户可能因好奇而故意忽视平台规则、尝试挑战或测试系统边界的情况。对此,可针对这类儿童用户,结合其理解水平,专门解释平台规则设立的原因。
- 在年龄验证流程中,需兼顾安全性与隐私保护的平衡。例如,企业可考量:除账户行为评估外,若采用语音分析(一种生物识别数据形式),是否符合数据最小化原则及隐私保护要求。
- 现有的风险防控措施,如向儿童用户及其监护人告知数据处理情况、对数据进行去标识化处理及删除数据,仍将是实现上述平衡的核心工具,需持续沿用。

需重点关注的是,人工智能驱动的聊天机器人存在不可预测性,这导致了诸多安全风险,且目前普遍尚未找到一刀切的解决方案。尽管百度已采取初步措施防控有害内容及行为相关风险,但仍可考虑对现有方法及用户体验开展更多评估。例如,可探索采用“系统卡片”(记录企业如何测试产品及测试结果的文档)等方式。

从企业角度来看,“服务儿童的负责任科技创新”的8项原则不仅是履行社会责任的标准,更是塑造积极企业形象、提升用户信任的基石。在设计儿童产品时,企业必须深入理解如何促进儿童能力发展,以及如何支持他们进行情绪调节。这种做法不仅有助于产品获得家长和教育工作者的信任,还能提升其市场竞争力。此外,这些原则为风险识别与防控提供了全面指导,确保产品能有效规避因设计缺陷可能引发的潜在法律风险与社会风险。

——蔡婷婷、马文涛、吴强、赵梦溪(百度)

## 腾讯未成年人保护中心

腾讯未成年人保护中心是一个旨在加强儿童上网安全与家庭教育的综合性平台。其核心是一套内置在腾讯各类应用中的家长控制系统,可帮助家庭管理儿童对游戏、社交媒体及数字内容的使用权限。该中心的人工智能模型不仅为家长工具和教育内容提供技术支持,还通过互动指导和案例分析为家庭提供助力。

本案例分析探讨了腾讯如何着手应对数字时代儿童成长过程中的两大核心挑战:安全保障(涵盖年龄验证、内容过滤及风险评估等方面)和人际关系(通过各类工具,助力构建健康的家庭关系与社交联系)。

腾讯通过整合技术、家庭教育与利益相关方协作,致力于在儿童安全与数字包容之间实现平衡,同时也认识到,随着风险不断演变,对反馈收集、影响分析及策略调整的需求将持续存在。<sup>1</sup>

### 免责声明:

联合国儿童基金会不代言任何公司、品牌、产品或服务。

本报告为联合国儿童基金会出版物。对各企业的致谢并不意味着本系列案例研究已获得各企业的认可或支持。文中提及任何特定企业,也不代表联合国儿童基金会代言该企业的政策、产品及实践。

## 概述

腾讯是一家总部位于中国的全球性科技公司,拥有多元化的科技产品及服务组合,涵盖游戏、社交媒体、音乐流媒体、支付服务及即时通讯等领域。这些产品在为儿童带来多样化机遇的同时,也伴随着各类风险。

为防控部分此类风险,腾讯已设立未成年人保护中心,并搭建了家长服务平台。该家长控制平台已整合到腾讯生态内的各类应用中,其核心功能在于管理儿童对游戏、社交媒体及其他数字内容的使用权

23.本研究借鉴了另一专项项目——“服务儿童的负责任科技创新”(Responsible Innovation in Technology for Children, RITEC)。该项目由联合国儿童基金会(UNICEF)与乐高集团(The LEGO Group)合作开展,资金来源于乐高基金会(The LEGO Foundation)。项目明确了八项原则(自主性,能力培养,情感关怀,人际关系,创造力,身份认同,安全保障,多样性、公平性与包容性),这些原则可用于评估儿童福祉与儿童权利。更多相关信息详见: <https://www.unicef.org/childrightsandbusiness/workstreams/responsible-technology/online-gaming/ritec-design-toolbox>

限。该系统包含多款工具，例如“腾讯公益家庭教育AI模型”，该模型可提供教育内容、问答功能及案例分析。此外，家长控制功能还针对儿童不同成长阶段，定制了相应的安全保护措施。与此同时，“设计保障儿童安全”的理念至关重要：这一理念既确保平台具备安全防护能力，又避免过度限制平台的开放性，从而保障儿童能切实从数字服务中获益。

本案例分析聚焦儿童福祉的两个关键维度——安全保障与人际关系。腾讯方面表示，未成年人保护中心将“设计保障儿童安全”理念置于优先地位，通过年龄验证、内容过滤及游戏时长限制等功能落实这一理念。腾讯指出，其安全保障措施不仅包括产品开发阶段的风险评估，还涵盖与儿童群体的直接沟通互动。此外，未成年人保护中心在“家庭守护计划”框架

下，提供由人工智能驱动的教育工具，这些工具可支持家长开展个性化学习，并助力解决各类家庭问题。

腾讯在平衡商业责任与社会责任方面面临挑战，尤其是在防控儿童消费风险这一问题上。家长控制平台可助力保障儿童安全，例如平台内置的年龄验证机制、内容过滤功能以及对使用游戏时间过长的限制措施。这些举措虽能防控部分儿童安全风险，但仍需对这些方法产生的实际效果开展更多分析。

此外，平衡风险防控与产品服务的开放性也至关重要，唯有如此，儿童在数字世界中的各项权利才能得到充分尊重。未来的发展方向包括拓展外部利益相关方的协作范围，例如与儿童群体及儿童安全领域专家开展更直接的咨询沟通，同时在产品全生命周期内规范尽职调查流程。

## 儿童福祉维度的实践应用

### 维度1:安全保障——儿童能够感受到安全,且其安全得到保障.

本案例分析探讨了腾讯在保障儿童安全方面相关流程的核心要素，具体包括：

安全保障流程	
风险评估	识别儿童安全影响，进而整合、跟踪风险防控措施并就其开展沟通。
年龄验证	估算或核实用户年龄的流程，以便数字服务能根据儿童需求进行定制，并落实符合其年龄段的保护措施。
投诉与补救	儿童可通过该途径举报任何侵犯儿童权利的行为，并申请补救。
儿童友好型信息提供	以儿童易于接受的语言及方式传递信息（例如使用清晰、易懂、简短的语句，仅在需要儿童做出决策时提供必要信息，且采用交互式工具），提升儿童的关注意愿。
儿童参与企业决策	咨询儿童意见，并将其观点纳入数字服务的设计与部署过程。

腾讯在产品全生命周期内开展的安全风险评估

设计保障儿童安全，应在产品全生命周期的关键阶段落实：<sup>24</sup>

1. 产品构思阶段：企业对产品进行概念构想、调研并明确产品价值主张，同时制定产品设计方案与概念的阶段。
2. 产品开发与测试阶段：企业进行产品开发与迭代、制作原型，并在产品上线前开展测试的阶段。
3. 销售、营销与部署阶段：企业将产品推向市场、推动产品部署，并在部署后监测产品使用情况以进行后续迭代的阶段。

腾讯表示，其“设计保障儿童安全”的理念从产品构思阶段便开始落实：在该阶段，腾讯会开展初步影响评估，以识别产品可能对儿童权利产生的潜在影响。随后，公司会召开跨部门团队会议，审核评估结果，并确定所需采取的风险防控措施。此阶段会有专门团队协调不同团队开展工作：家庭教育咨询团队负责监督教育模型的构思与设计，家庭教育团队负责内容运营管理，产品团队则承担产品开发工作。

腾讯表示，在产品开发与测试阶段，公司会开展更为详细的影响评估：根据风险的严重程度与发生概率对其进行评估和优先级排序，核心目标是优先解决高风险问题。此阶段会制作产品原型并进行测试，且会重点针对儿童用户开展测试，以确保产品符合安全与伦理标准。同时，家庭教育咨询团队会开展实地调

研，通过问卷形式直接从社区、学校及家庭收集数据。此外，一支涵盖产品开发、合规法务及社会责任部门成员的跨部门团队会协同合作，共同评估产品对儿童权利产生的影响。

腾讯表示，公司会借助数据工具开展定期风险评估并识别潜在问题。同时，腾讯还会收集并分析家长及儿童用户的反馈，进一步优化产品。在整个阶段，腾讯会与政府监管部门保持沟通，确保合规运营；此外，还会充分利用学术合作及非营利组织合作所获取的洞见，优化教育模型。公司倡导对教育模型进行迭代改进，重点在于增强其类人化特征，从而实现与用户的高效互动。

腾讯表示，在市场推广、销售及部署阶段，公司会收集用户反馈，以识别并解决产品上线后出现的各类风险与挑战。同时，腾讯还会接收来自家长及儿童用户的反馈，用于产品的持续优化。此外，腾讯还会主动与非营利组织及学术机构开展沟通，分享相关洞见，并依据儿童安全原则对产品产生的影响进行评估。

### “设计保障儿童安全”理念在其他流程层面的体现

儿童友好型信息提供服务有助于儿童理解：他们所使用的各类数字产品及服务相关的服务条款、这些产品与服务可能存在的风险，以及如何举报伤害行为、标记不安全或非法内容。

——迈向“设计保障儿童数字安全”<sup>25</sup>

24. 这是对科技公司产品开发生命周期各阶段的简化概括。不同公司对生命周期阶段的划分可能有所不同。

25. OECD(2024),《迈向“设计保障儿童数字安全”》(经济合作与发展组织《数字经济文件》),经济合作与发展组织出版,<https://doi.org/10.1787/c167b650-en>,第30页。

风险评估是“设计保障儿童安全”理念的核心环节,在风险识别、风险防控以及防控措施有效性的跟踪与沟通方面发挥关键作用。然而,正如经济合作与发展组织(OECD)所指出的,企业还需实施其他流程,以确保风险能够得到恰当识别与应对。本案例研究重点介绍了其中四类流程:<sup>26</sup>

- 1) 年龄验证机制
- 2) 儿童友好型信息提供
- 3) 投诉与补救机制
- 4) 企业决策过程中纳入儿童参与或听取儿童意见的机制

## 1. 腾讯的年龄验证机制

在“设计保障数字安全”理念中,确认用户年龄至关重要:只有明确哪些用户实为儿童,相关服务才能切实保护儿童免受不安全内容与行为的影响。

——迈向“设计保障儿童数字安全”<sup>27</sup>

腾讯表示,其年龄验证系统整合了强制性的实名认证与随人脸识别核验功能。该公司指出,所有用户均需通过中国的公安身份识别系统完成实名认证,该系统会自动判定用户是否为未成年人。据介绍,这一判定结果会触发未成年人保护中心针对不健康游戏使用行为的应对流程。

人工智能人脸识别技术还用于核验疑似使用成人账号的儿童用户身份。腾讯表示,公司不会存储完整人脸图像;相反,会对人脸信息进行进一步处理,并将其与身份标识信息分开存储,以防止数据泄露。腾讯称,上述信息会在限定期限后被删除或匿名化处理,且数据的存储与传输均采用加密方式。

### • 分级管控策略

在时间管理方面,未成年人仅可在周五、周六及法定节假日的20:00至21:00期间登录游戏(此为中国相关政策规定)。<sup>28</sup>当儿童使用支持直接扣款支付的数字产品时,平台会对其每日总消费金额设置限制:8周岁以下用户禁止充值;8-12周岁用户单次充值最高不超过50元,每月累计充值最高不超过200元;13-18周岁用户单次充值最高不超过100元,每月累计充值最高不超过400元。此外,若用户单日消费总额超过500元,系统将触发异常支付预警,并向家长发送短信提醒。

### • 家庭协同工具

腾讯表示,该平台支持家长关联孩子的游戏账号,以便家长实时查看游戏时长与消费记录。家长还可设置个性化的时间限制(例如,仅允许孩子在周末玩游戏),从而更好地管理孩子的互联网使用情况。

26.需说明的是,本案例研究为简洁起见,对这些流程的分类与经济合作与发展组织(OECD)的分类略有不同。OECD的流程分类体系还包含用于判定伤害的技术工具及以儿童为中心的设计;这些内容的相关方面已纳入本风险评估的“实质内容”部分。

27.OECD(2024),《迈向“设计保障儿童数字安全”》(经济合作与发展组织《数字经济文件》),经济合作与发展组织出版,<https://doi.org/10.1787/c167b650-en>,第30页。

28.此举遵循了国家新闻出版署《关于进一步加强管理切实防止未成年人沉迷网络游戏的通知》中所阐述的方法。

## 2. 腾讯的儿童友好型信息提供机制

儿童友好型信息提供服务有助于儿童理解：他们所使用的各类数字产品及服务相关的服务条款、这些产品与服务可能存在的风险，以及如何举报伤害行为、标记不安全或非法内容。

——迈向“设计保障儿童数字安全”<sup>29</sup>

腾讯表示，其运用游戏、动画、绘画等互动式媒体形式，让信息传递过程更具趣味性，从而吸引儿童的注意力(见下方图片)。



© 图片由腾讯提供  
腾讯制作的用于向儿童宣传网络安全的动画视频

此外，腾讯表示，其平台提供专门为儿童设计的互联网安全主题动画视频。例如，《未成年人防诈骗图解指南》将常见网络诈骗类型融入童谣中，并搭配动画场景，以吸引儿童用户的注意力。据介绍，这类动画视频的播放量已突破100万次，且已进入校园播放。

## 3. 腾讯的投诉与补救机制

帮助儿童了解如何举报事件与伤害行为、标记不安全或非法内容，这一点至关重要。若投诉机制难以被儿童找到、获取或使用，其实际效用将大打折扣……因此，儿童友好型投诉机制必不可少。

——迈向“设计保障儿童数字安全”<sup>30</sup>

腾讯平台设有多个举报渠道，供儿童用户反馈技术问题，或举报不良信息及非法活动。腾讯表示，目前平台共设有四大投诉渠道，具体如下：

- 1.通过平台直接反馈与投诉
- 2.向举报邮箱发送问题反馈
- 3.通过腾讯专属客服渠道进行举报
- 4.向腾讯数据隐私保护部门邮寄反馈意见。

腾讯表示，在收到反馈意见后，承诺将在15个工作日内完成审核并给出反馈。

29.OECD(2024),《迈向“设计保障儿童数字安全”》(经济合作与发展组织《数字经济文件》),经济合作与发展组织出版,<https://doi.org/10.1787/c167b650-en>,第38页。

30.同上注。

腾讯指出，儿童用户可使用的上述四大投诉渠道均需监护人协助并同意方可使用。在这四大渠道中，部分渠道对儿童而言使用难度较高，例如通过邮件向腾讯反馈问题——这类渠道往往需要监护人提供大量协助或指导。而其他渠道（如通过平台内置的一键举报功能）则更易于儿童操作，仅需少量协助即可使用。



© 图片由腾讯提供  
腾讯平台内用于直接反馈与投诉的界面

#### 4. 腾讯在企业决策过程中纳入儿童参与或听取儿童意见的机制

在涉及儿童自身的事务中，儿童享有表达意见的权利。当儿童（及青少年）参与到影响自身的项目与政策制定过程中时，这些项目与政策更有可能更好地契合儿童的需求、兴趣与背景。

——迈向“设计保障儿童数字安全”<sup>31</sup>

腾讯表示，其通过访谈、调研和小组讨论的形式与儿童直接沟通，以了解他们对数字服务的期望与需求。该公司会观察儿童在使用数字服务时的行为与反应，从而明确他们的兴趣点及面临的问题。儿童参与此类调研需遵循自愿原则，且须获得家长或监护人的同意。通过该类调研收集的数据会进行匿名化处理，遵循数据最小化原则，并依据中国数据保护相关法律开展收集工作。

为提高公众的网络安全意识，腾讯以儿童易于接受的形式（如漫画、互动游戏）普及常见诈骗类型及防护方法。同时，该公司还会组织路演与课程（主要在城市地区开展），向青少年宣传上网安全知识。每年，腾讯都会邀请儿童参观腾讯未成年人保护中心。中心内设有专门的“网络安全屋”，通过寓教于乐的沉浸式互动体验，让儿童在游戏过程中学习网络保护知识。



© 图片由腾讯提供  
腾讯组织的针对青少年的网络安全宣传路演活动

31.同上注，第40页。

## 腾讯产品面临的安全风险及未成年人保护中心

本节将阐述腾讯产品中存在的部分儿童安全风险(这些风险已通过未成年人保护中心得到部分防控),以及上文所述的“设计保障儿童安全”流程如何帮助腾讯防控此类风险。需要说明的是,本节列举的安全风险案例仅为示例,并非详尽无遗。

### 电子游戏的不良使用

游戏产品的潜在风险之一是儿童用户对电子游戏产生不良使用行为。这种不良使用行为的特征是挤占睡眠、体育锻炼等必要活动的时间,进而可能导致健康状况不佳。根据腾讯未成年人保护中心2024年的数据,9-12岁男性儿童(占总体的82%)是电子游戏相关问题的高风险群体,而家长普遍缺乏应对该问题的有效教育方法。在存在游戏不良使用倾向的家庭中,39.3%的家庭采用放任型教养方式,20.36%的家庭存在过度保护行为(防护措施过当),14.98%的家庭则处于教养缺位状态(对孩子的行为缺乏了解)。

腾讯表示,其通过一套游戏前、游戏中及游戏后的组合措施,致力于应对游戏不良使用风险。据悉,这些措施包括教育工具、一对一专家指导,以及家长通过平台对未成年人游戏时长进行管控。

### 应对与内容及行为相关的风险

腾讯旗下的系列产品可能使儿童面临与内容相关的风险,此类风险会威胁儿童的安全。在游戏社区中,儿童可能接触到不符合其年龄段的内容或露骨内容(包括含暴力、性主题的内容),这些内容可能出现在聊天信息及用户生成内容中。

此外,还存在诸多与行为相关的风险:其中一个重要例子是,施害者常利用网络游戏诱导儿童以实施性侵害。在这些网络生态系统中,由于实时互动并非总能得到全面监管,网络欺凌的发生概率以及儿童接触潜在施害者的风险也会随之增加。

为防控上述风险,腾讯已尝试制定多项应对策略,例如出台安全政策、推出未成年人模式、实施技术监测、开展人机协同审核,并建立反馈举报机制等。这些策略整合了自动化审核与人工审核流程。举例来说,若某用户发布内容为“昨天在X区看到有人持刀打架,太吓人了”,系统会通过语义分析识别出“刀”“打架”“吓人”等敏感词汇,将该内容归类为暴力场景描述类信息。随后,人工审核人员会进一步确认该内容是否违反内容规范,确认违规后便会将其删除。

同样,此类自动化或人工审核也可能识别出施害者(如诱导侵害儿童的人员)常用的语言,例如针对未成年人的俚语,或涉及不当关系的表述——即便这些语言较为隐晦或经过加密处理。腾讯指出,其基于人工智能技术构建的内容安全审核系统,融合了大语言模型的多模态能力与人工审核优势,能够精准检测出不良内容及行为(如涉性内容),召回率超过80%。该系统在设计上会对这些不良内容及行为的特征标记,以便开展进一步核查,目标是及时识别潜在不良内容,从而保护易受伤害的用户群体。此类安全风险对科技企业而言颇具挑战性,因此需要对上述策略进行持续审核与调整,以实现风险防控。

## 维度2:人际关系层面——儿童用户可体验与他人的社交联结,并获得归属感。

### 通过以家庭教育为导向的人工智能模型补充人际关系

腾讯未成年人保护中心致力于促进亲子关系,其依托的正是腾讯所定义的中国科技行业内“首个家庭教育人工智能模型”。该模型旨在成为家长的数字化教育伙伴,为家长在教育过程中面临的难题提供个性化解决方案。

该人工智能模型以聊天机器人的形式集成于腾讯平台中。该聊天机器人基于腾讯混元大语言模型构建,家长可通过它使用腾讯平台上的各类工具及应用,例如教育视频、案例研究和志愿服务相关资源。腾讯指出,尽管该聊天机器人仍在持续优化,但它的目标是:在对话场景中,根据用户需求开展精准的内容对话、提供针对性的工具推荐,并提供教育咨询服务。

该模型的训练数据来源于未成年人保护中心处理过的真实案例。这款家庭教育模型不仅致力于应对当前的教育难题、预测未来的教育趋势,还可解答与家庭教育、学校教育及社区教育相关的问题。根据腾讯的基准测试数据,该模型的问题解决准确率达到94.6%。除问答功能外,模型还内置了多媒体教育资源库,例如涵盖各类家庭教育主题的视频资源。

未成年人保护中心还采用了互动式支持方式,例如志愿者问答环节和教育工具。在志愿者问答方面,腾讯会筛选具备资质且经验丰富的志愿者来解答用户咨询,让用户能从背景各异的志愿者那里获取合适的答案。在某些特定的育儿场景中,非即时性的文字及语音指导可能效率较低,家长用户可通过能提供额外信息的教育工具提升自身育儿能力。这些工具借助互动

视频,梳理育儿过程中的常见误区并介绍科学的育儿方法,随后引导用户进入任务实践环节,将所学方法应用于实际育儿场景中。

### 保障儿童福祉与安全面临的挑战及建议

在撰写这些案例研究的过程中,我们发现了若干关键挑战,并已提供相应建议供企业参考。需说明的是,这些问题在其他致力于保障儿童福祉与安全的科技企业中也普遍存在。此外,必须指出的是,目前不存在万能解决方案。

**挑战:**盈利压力与合规要求之间的矛盾。例如,游戏及社交平台作为主要收入来源,其核心目标之一是留住用户,因此会着力打造具有吸引力的使用体验。然而,平台同时必须防止用户(尤其是儿童)过度使用或不当使用产品,并保障用户权益。在动态变化的环境中平衡这些相互冲突的需求,对平台而言颇具挑战性。

#### 建议:

- 考虑设置负责任的使用功能,包括添加时长限制、适龄推荐及休息提醒,在不显著降低用户参与度的前提下,引导用户养成健康的使用习惯。
- 采用双目标框架,将盈利目标与合规要求相协调:将儿童保护及儿童福祉纳入产品关键绩效指标(KPI),确保衡量成功的标准既包含收入指标,也涵盖安全成果。
- 扩展并规范化相关流程,在产品构思与开发阶段即对风险防控措施进行跟踪评估;同时建立

产品上线后的工作流程,定期将风险评估结论整合至“成长守护平台”。

**挑战:**尽管实名认证以及限制时长、限制消费等遏制过度或不当使用的系统,体现了对儿童网络保护的投入,但当前挑战在于,如何在落实这些防护措施的同时,既不损害用户体验,也不影响数字包容性。要实现合理平衡,不仅需要有效执行相关措施,还需填补家长监护环节的空白。此外,年龄验证与认证机制也引发了质疑,有观点认为这类机制可能侵犯儿童的隐私权,以及他们在数字环境中切实参与的权利。

**建议:**

- 考虑采用替代方法(例如通过儿童自我声明结合验证检查,如简短问卷、任务或理解能力测试等,以评估用户是否可能为儿童),此类方法无需收集和生物识别信息,以规避人脸识别等侵犯隐私的做法。
- 组建跨职能监督团队,纳入合规、产品及伦理领域的专家,持续审查平台运营行为,确保商业目标的实现不会损害儿童权利。
- 通过预警提示与指导建议,加强向家长提供信息的力度,助力家长在不过度干预的前提下,监督和管理孩子的游戏活动。

“服务儿童的负责任科技创新”的8项原则对于思考儿童福祉所受影响具有重要作用。该框架从多个维度保障儿童福祉,规避了单一维度分析的局限性。在安全层面,它促使人们关注产品的物理安全与数据安全;在内容层面,它保障了内容的质量与适宜性,同时防范不良信息。这些原则如同一个完备的过滤器,能够系统性地筛选出可能危害儿童福祉的各类因素。

——腾讯

## 猿编程App

猿编程App是猿编程打造的编程教育平台,旨在帮助青少年掌握人工智能时代所需的技能。该产品为儿童对接讲师,提供定制化课程及各类学习工具。猿编程称其会根据不同年龄段儿童的发展需求设计课程,并构建了可供儿童分享编程项目、发表评论的社区。

本案例研究结合RITEC-8框架的两个核心维度——安全保障与人际关系,探讨猿编程如何保障儿童福祉。文中阐述了该产品在为低龄用户提供安全学习环境、促进积极互动方面的核心关注方向。猿编程的案例,揭示了建设兼具吸引力且能保障儿童权利的数字教育平台时面临的机遇与挑战。<sup>32</sup>

### 免责声明:

联合国儿童基金会不代言任何公司、品牌、产品或服务。

本报告为联合国儿童基金会出版物。对各企业的致谢并不意味着本系列案例研究已获得各企业的认可或支持。文中提及任何特定企业,也不代表联合国儿童基金会代言该企业的政策、产品及实践。

## 概述

猿编程打造了猿编程App这一面向青少年群体的编程教育平台,核心目标是帮助青少年做好应对人工智能时代的准备。

猿编程旨在为儿童对接讲师,同时为他们提供定制化课程及其他学习工具。<sup>33</sup>猿编程表示,在设计课程时,会对不同年龄段儿童用户的发展特点进行评估。

该产品还具备社区功能,儿童用户可在其中查看彼此的编程项目并发表评论。

教育平台能够为儿童用户提供教育支持、培养他们的创造力,还能帮助他们提升沟通效率。然而,儿童也可能通过这类平台面临更高的风险:例如,在线互动可能伴随有害内容及行为风险;竞赛或游戏化功能可能

32.本研究借鉴了另一专项项目——“服务儿童的负责任科技创新”(Responsible Innovation in Technology for Children, RITEC)。该项目由联合国儿童基金会(UNICEF)与乐高集团(The LEGO Group)合作开展,资金来源于乐高基金会(The LEGO Foundation)。项目明确了八项原则(自主性,能力培养,情感关怀,人际关系,创造力,身份认同,安全保障,多样性、公平性与包容性),这些原则可用于评估儿童福祉与儿童权利。更多相关信息详见:<https://www.unicef.org/childrightsandbusiness/workstreams/responsible-technology/online-gaming/ritec-design-toolbox>

33. 数字教育平台可为儿童学习提供补充支持,但无法替代正规教育体系。

引发心理健康问题；儿童的数据也可能在未获得其知情同意的情况下被处理。此外，平台提供的交易功能还可能使儿童面临合同风险，如过度消费等。

确保产品遵循“设计保障儿童安全”原则进行安全设计，对于最大化产品效益、最小化潜在风险至关重要。本案例研究深入探讨了猿编程为应对这些风险所采取的安全防控措施。

本案例研究围绕两个借鉴自国际标准的维度展开，即安全保障与人际关系。猿编程指出，在产品构思、开发及部署阶段，均会开展结构化安全评估。具体措施包括

安全的数据处理规范、完善的隐私保护机制，以及用于应对新型风险（如人工智能生成内容带来的风险）的用户反馈机制。此外，猿编程表示，通过互动编程项目、直播课程及线下活动，该产品的社区化功能能够促进青少年用户之间的协作、沟通与人际关系构建。

猿编程在应对儿童安全风险动态变化（如新兴的人工智能风险）方面面临挑战。其有待改进的方向包括：拓展风险评估流程，以覆盖更广泛的儿童权利领域；在产品开发过程中，更广泛地联动儿童、学术研究者及非营利组织等外部利益相关方。

## 儿童福祉维度的实践应用

### 维度1:安全保障——儿童能够感受到安全,且其安全得到保障.

本案例研究重点阐述了猿编程在保障儿童安全方面所实施流程的核心要素：

安全保障流程	
风险评估	识别儿童安全影响,进而整合、跟踪风险防控措施并就其开展沟通。
年龄验证	估算或核实用户年龄的流程,以便数字服务能根据儿童需求进行定制,并落实符合其年龄段的保护措施。
投诉与补救	儿童可通过该途径举报任何侵犯儿童权利的行为,并申请补救。
儿童友好型信息提供	以儿童易于接受的语言及方式传递信息(例如使用清晰、易懂、简短的语句,仅在需要儿童做出决策时提供必要信息,且采用交互式工具),提升儿童的关注意愿。
儿童参与企业决策	咨询儿童意见,并将其观点纳入数字服务的设计与部署过程。

## 猿编程在整个产品生命周期内开展的安全保障风险评估

设计保障儿童安全应贯穿产品生命周期的关键阶段<sup>34</sup>:

1. 产品构思阶段:企业对产品进行概念构想、调研并明确产品价值主张,同时制定产品设计方案与概念的阶段。
2. 产品开发与测试阶段:企业进行产品开发与迭代、制作原型,并在产品上线前开展测试的阶段。
3. 销售、营销与部署阶段:企业将产品推向市场、推动产品部署,并在部署后监测产品使用情况以进行后续迭代的阶段。

猿编程表示,其“设计保障儿童安全”理念从产品构思阶段便已启动。安全设计评估覆盖多个工作流程,包括平台技术功能、课程开发、讲师招聘,以及平台内容相关的安全保障。例如,猿编程指出,已为获准在平台授课的讲师制定了结构化评估流程,确保通过审核的人员既具备安全资质,又掌握儿童教学所需的专业能力。

猿编程称,猿编程的评估流程整合了多项以安全保障为核心的风险防控功能及工作流程。例如,该产品设有安全、实时的内容审核机制与用户举报流程,以避免用户接触有害内容。为保障儿童用户的隐私,猿编程在设计涉及数据收集与处理的产品功能时,采用了必要性与数据最小化的数据保护原则。为防止

儿童使用猿编程时误操作付款,平台设置了密码验证、手机确认及家长模式等功能,确保付款操作由家长而非儿童执行。

猿编程表示,在产品开发与测试阶段,已建立多项安全保障相关的风险防控工作流程。内部团队会对课程内容安全进行监督,参与内容开发及安全监督的团队包括课程内容研究院、教学评估研究院、创新产品研究院、AI课堂中心及体验课中心的成员。安全团队还会对产品开展多轮以安全为导向的测试(包括对抗性测试),以保障数据安全。为尊重用户的隐私关切,猿编程规定,儿童需在家长审阅并同意其隐私政策后,方可使用该产品。

猿编程指出,在销售、营销及部署阶段,平台依托用户反馈机制持续完善猿编程的安全保障水平。用户可通过电话、邮件等渠道提交对产品的反馈。猿编程会定期汇总用户反馈,并开展定量与定性分析,旨在为用户提供更优质的服务体验。为应对新兴风险(如人工智能引发的风险),猿编程已制定产品部署后的应对措施,例如定期审阅并更新隐私政策。

### “设计保障儿童安全”理念在其他流程层面的体现

仅通过单一举措,难以实现面向儿童的“设计保障数字安全”;相反,这需要多项互补且相互关联的举措——这些举措协同作用,方能助力为儿童营造安全且有益的数字环境。

——迈向“设计保障儿童数字安全”<sup>36</sup>

34.这是对科技公司产品开发生命周期各阶段的简化概括。不同公司对生命周期阶段的划分可能有所不同。

35. OECD(2024),《迈向“设计保障儿童数字安全”》(经济合作与发展组织《数字经济文件》),经济合作与发展组织出版,<https://doi.org/10.1787/c167b650-en>,第30页。

36.需说明的是,本案例研究为简洁起见,对这些流程的分类与经济合作与发展组织(OECD)的分类略有不同。OECD 的流程分类体系还包含用于判定伤害的技术工具及以儿童为中心的设计;这些内容的相关方面已纳入本风险评估的“实质内容”部分。

风险评估是“设计保障儿童安全”理念的核心环节，在风险识别、风险防控以及防控措施有效性的跟踪与沟通方面发挥关键作用。然而，正如经济合作与发展组织(OECD)所指出的，企业还需实施其他流程，以确保风险能够得到恰当识别与应对。本案例研究重点介绍了其中四类流程：

- 1) 年龄验证机制
- 2) 儿童友好型信息提供
- 3) 投诉与补救机制
- 4) 企业决策过程中纳入儿童参与或听取儿童意见的机制

### 1. 猿编程的年龄验证机制

在“设计保障数字安全”理念中，确认用户年龄至关重要：只有明确哪些用户实为儿童，相关服务才能切实保护儿童免受不安全内容与行为的影响。

——迈向“设计保障儿童数字安全”<sup>37</sup>

猿编程表示，猿编程并未设计任何不适用于18岁以下用户的内容。猿编程的辅导老师会核实学生的年龄、发展阶段及兴趣爱好，以便为其提供符合年龄段的内容。猿编程的课程分为八个级别，以满足不同年龄段用户的需求。

### 2. 猿编程的儿童友好型信息提供

儿童友好型信息提供服务有助于儿童理解：他们所使用的各类数字产品及服务相关的服务条款、这些产品与服务可能存在的风险，以及如何举报伤害行为、标记不安全或非法内容。

——迈向“设计保障儿童数字安全”<sup>38</sup>

猿编程表示，猿编程中与安全及教学相关的语言，均采用儿童易于理解的简洁表述。此外，应用内的所有信息均搭配卡通形象、鲜艳的配色方案及可爱的图形元素呈现，以保持儿童用户的参与度。

### 3. 猿编程的投诉与补救机制

帮助儿童了解如何举报事件与伤害行为、标记不安全或非法内容，这一点至关重要。若投诉机制难以被儿童找到、获取或使用，其实际效用将大打折扣……因此，儿童友好型投诉机制必不可少。

——迈向“设计保障儿童数字安全”<sup>39</sup>

猿编程表示，猿编程的官方网站及应用程序均设有一键举报渠道。用户可举报的问题包括抄袭、不当视觉内容、不友好行为、侵权、违反法律法规等各类情况。举报处理结果将反馈给用户。此外，用户及其家长或监护人还可直接向辅导老师提交举报并反馈问题。

37. OECD(2024),《迈向“设计保障儿童数字安全”》(经济合作与发展组织《数字经济文件》),经济合作与发展组织出版,<https://doi.org/10.1787/c167b650-en>,第30页。

38. 同上注,第38页。

39. 同上注。

#### 4. 猿编程在企业决策过程中纳入儿童参与或听取儿童意见的机制

在涉及儿童自身的事务中,儿童享有表达意见的权利。当儿童(及青少年)参与到影响自身的项目与政策制定过程中时,这些项目与政策更有可能更好地契合儿童的需求、兴趣与背景。

——迈向“设计保障儿童数字安全”<sup>40</sup>

猿编程表示,会通过反馈渠道将儿童的意见纳入猿编程的开发周期中。该反馈渠道收集的信息将用于定期更新猿编程的内容设置及各类机制。例如,为确保每节课程都适配特定年龄段的儿童,若猿编程收到儿童或家长反馈某节课程可能存在年龄段不适配的问题,便会对课程内容进行调整。

#### 猿编程存在的儿童安全保障风险

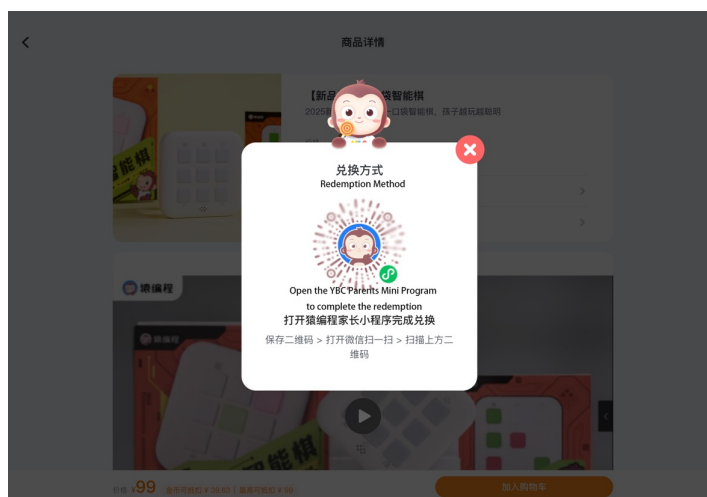
本节将阐述猿编程存在的部分儿童安全保障风险,以及前文所述的“设计保障儿童安全”流程如何帮助猿

编程防控这些风险。需要说明的是,本节列举的安全保障风险案例仅为示例,并非详尽无遗。

#### 应对猿编程的用户风险

猿编程表示,猿编程内不含付费游戏、商品销售及广告内容,且应用内所有场景均不支持使用真实货币交易。平台配套开发了供家长使用的手机端配套应用,真实货币交易仅可在该配套应用内进行。此外,儿童用户可通过积极参与课堂讨论、完成建设性活动等方式获取积分,这些积分可在应用内的积分商城兑换各类物品,例如文具、玩具及礼品,其中书包、儿童书籍和机器人尤其受欢迎。猿编程设有退款机制并提供客服支持,这一举措大幅降低了用户相关风险。

猿编程指出,猿编程设置了密码验证、手机确认及家长端配套应用等功能,以防止儿童在使用应用时误操作付款。这些措施可确保付款操作由家长而非儿童执行。如下图所示,兑换所获积分需经过家长授权,且该授权操作仅可由家长完成。



© 图片由猿编程提供

所获积分的兑换操作仅可由家长在家长端配套程序内完成

40.同上注,第40页。

### 应对猿编程平台互动过程中产生的接触类风险

猿编程的教育服务包含远程会议形式的师生互动这一核心环节,而该环节可能产生接触类风险。此外,猿编程社区是一个开放式交流平台,儿童可在此相互沟通,分享彼此的编程项目并为其点赞。但与所有线上社区一样,该平台也存在风险,例如恶意用户、伪装成儿童的成年人,以及与内容和行为相关的滥用行为(包括诱导侵害、骚扰及欺凌等)。

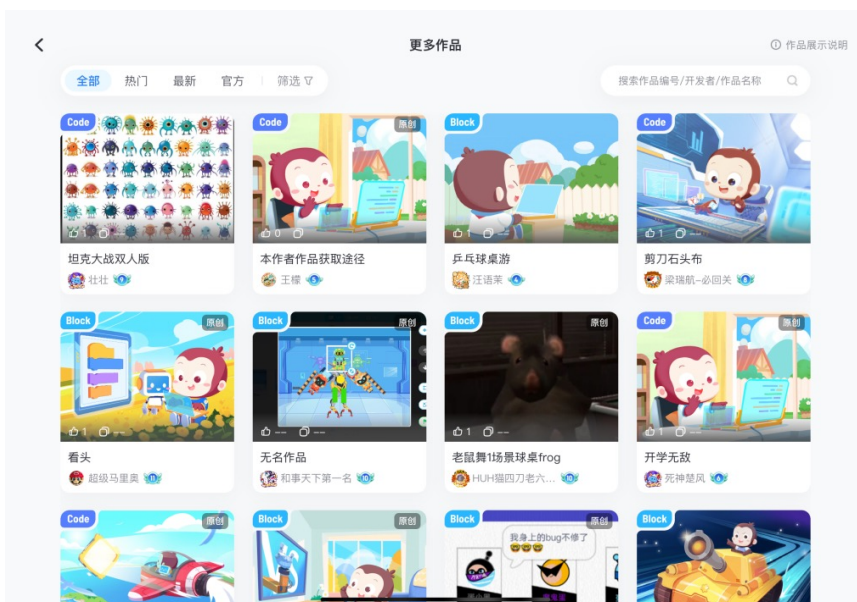
猿编程表示,已采取多项措施防控上述风险。首先,猿编程的所有教师均需通过严格的背景审查,有过伤害儿童记录的教师一律不予录用。其次,与儿童有接触的授课教师及辅导老师需接受持续性评估,包括入职审查与定期检查。第三,猿编程通过实名认证等社区管理措施,防止成年人伪装成儿童。针对不良分子冒充儿童用户的风险,平台运用关键词过滤技术及人工智能图像识别技术,识别不当行为或内容,并标记相关账号以采取管控措施。

## 维度2. 人际关系:儿童用户可体验与他人的社交联结及归属感。

### 人际关系:猿编程的编程教育社区

猿编程表示,猿编程的编程平台为儿童提供了一个展示空间,可向包括同龄人及家长在内的广大受众

展现自己的创意与努力成果。该平台现有10万余名活跃学生用户,这些用户已创作超50万个编程项目,编写的代码行数超过3亿行。



© 图片由猿编程提供  
猿编程创意社区界面

猿编程指出，互动是猿编程的编程社区的核心。学生可对彼此的项目进行点赞、收藏、评分及分享。此外，经其他学生同意后，学生还可查看或复制其作品的源代码，从而实现相互学习。

猿编程进一步表示，猿编程开设了编程直播课程，来

自全国各地的学生可参与小班互动课堂。这些课程包含课堂竞赛环节，通过健康的竞争氛围培养学生的集体荣誉感与同伴情谊。借助这些学习体验，学生能够结识来自中国不同地区的同龄人，进而加强与更广泛学习社群的联结。



© 图片由猿编程提供  
学生分享的编程项目互动界面



© 图片由猿编程提供  
课程期间学生评论的截图

## 保障儿童福祉与安全面临的挑战及建议

在撰写这些案例研究的过程中，我们发现了若干关键挑战，并提出了相应的建议供企业参考。需说明的是，这些问题在其他致力于保障儿童福祉与安全的科技企业中也普遍存在。此外，必须指出的是，目前不存在万能解决方案。

**挑战：**在通过设计保障安全的过程中，核心挑战之一在于儿童安全风险的形式正不断变化。

### 建议：

- 建立实时监控与反馈机制，用以追踪新出现的威胁。这能助力企业快速响应、持续完善安全措施，同时通过更新政策与产品功能，主动防范潜在风险。
- 拓展风险评估流程，覆盖更多主题领域。为在动态环境中实现全面的风险防控，下一步可推动相关流程制度化，确保对儿童权利各维度的影响均纳入评估范畴。
- 扩大产品开发生命周期中咨询外部相关方的范围。除已建立的事后监控反馈渠道外，猿编程还可考虑在整个产品开发过程中，直接与儿童、儿童安全专家、儿童安全领域学者或非营利组织开展合作；在与儿童沟通咨询时，需充分考量伦理因素。

**挑战：**整合安全相关发现通常需要大量资源，包括时间、专业技术及资金投入。在遵守产品开发时间节点的同时合理分配这些资源，往往颇具难度。

### 建议：

- 构建基于风险的框架，优先识别并应对最关键的安全问题，确保时间、专业技术及资金资源集中投入到能产生最大成效的领域。
- 将儿童安全考量融入产品从设计到测试的开发流程的每个阶段，以避免后续产生高昂的整改成本，并确保持续符合伦理准则与监管标准。

在仔细研读“服务儿童的负责任科技创新”框架后，我们发现这8项原则所倡导的诸多理念与我司现行原则高度契合。我们还将依据这些原则的建议，进一步完善产品与相关原则。未来，我们计划将这些原则融入课程与应用程序的设计中，为儿童打造安全、有保障且内容丰富的数字环境。

——猿编程

## 核桃编程3D图形化编辑平台

核桃编程的3D图形化编辑平台为儿童提供学习编程、制作交互式项目以及与他人建立联系的空间。该平台有论坛、即时通讯、竞赛、项目分享等社区功能。

本案例研究将探讨核桃编程如何着力解决以下问题：

- **安全保障：**通过风险评估、专家咨询、设计即隐私原则及技术防护措施，保护儿童免受网络风险、不良内容及数据滥用的侵害。
- **多元性、公平性与包容性：**提供多样化的角色工具、面向残障儿童的无障碍功能，以及为服务不足社区群体提供的平价编程教育。
- **能力培养。**

核桃编程正致力于完善内容审核、用户反馈及风险管理体系，旨在打造一个能让所有儿童安全培养技能、发挥创造力并与同龄人在线互动的平台。<sup>41</sup>

### 免责声明：

联合国儿童基金会不代言任何公司、品牌、产品或服务。

本报告为联合国儿童基金会出版物。对各企业的致谢并不意味着本系列案例研究已获得各企业的认可或支持。文中提及任何特定企业，也不代表联合国儿童基金会代言该企业的政策、产品及实践。

## 概述

核桃编程打造了核桃编程3D图形化编辑平台，这是一款专为儿童设计、融合社区与教育功能的3D平台。<sup>42</sup>

该平台提供多种互动形式，包括公共论坛、即时通

讯、排行榜与竞赛功能，以及编程项目分享空间。平台还配备各类创意工具，支持儿童用户开发并分享交互式项目（例如创作故事），同时提供编程教程与教育资源供用户学习使用。

41. 本研究借鉴了另一专项项目——“服务儿童的负责任的科技创新”（Responsible Innovation in Technology for Children, RITEC）。该项目由联合国儿童基金会（UNICEF）与乐高集团（The LEGO Group）合作开展，资金来源于乐高基金会（The LEGO Foundation）。项目明确了八项原则（自主性，能力培养，情感关怀，人际关系，创造力，身份认同，安全保障，多样性、公平性与包容性），这些原则可用于评估儿童福祉与儿童权利。更多相关信息详见：<https://www.unicef.org/childrightsandbusiness/workstreams/responsible-technology/online-gaming/ritec-design-toolbox>

42. 数字教育平台可为儿童学习提供补充支持，但无法替代正规教育体系。

教育类平台虽能为儿童用户提供学习便利、培养创造力,并帮助他们提升沟通效率,但儿童在这类平台上也可能面临更高风险:在线互动可能伴随不良内容与行为风险;竞赛及游戏化功能或引发心理健康问题;儿童的数据还可能在未获得其知情同意的情况下被处理。此外,平台若提供交易功能,可能使儿童面临合同风险(如过度消费)。

确保产品设计符合“设计保障儿童设计”原则,具备安全属性,对于最大化平台效益、最小化风险至关重要。本案例研究深入探讨核桃编程3D图形化编辑平台为应对上述风险所采取的安全防控措施。

核桃编程案例研究聚焦儿童福祉的三个维度:安全保障,多样性、公平性与包容性,以及能力培养。

核桃编程表示,其会针对平台的即时通讯、排行榜等功能开展风险评估,过程中会咨询专家与利益相关方,

并通过强化内容审核、落实“设计保障隐私”原则等调整措施,解决已识别的风险。平台还实施了技术防护手段,如更严格的过滤机制与优化后的举报系统,以防范网络欺凌、网络暴力及数据滥用等问题。在多样性、公平性与包容性方面,核桃编程尝试通过多样化角色创建工具、面向残障儿童的无障碍设计、适应性编程课程,以及为服务不足地区推出的平价编程教育项目,将该理念融入平台运营。

对核桃编程而言,在严格的安全管控与便捷的用户体验之间取得平衡,仍是一项持续面临的挑战。随着数字风险不断演变,核桃编程必须保持灵活性,持续更新其安全框架与方法体系。除采用更具前瞻性的风险评估技术以预判未来挑战外,通过咨询更广泛的儿童权利专家及非营利组织来深化利益相关方参与,也将是至关重要的举措。

## 儿童福祉维度的实践应用

### 维度1:安全保障——儿童能够感受到安全,且其安全得到保障。

本案例研究探讨了核桃编程为确保儿童安全保障所实施流程中的关键要素:

安全保障流程	
风险评估	识别儿童安全影响,进而整合、跟踪风险防控措施并就其开展沟通。
年龄验证	估算或核实用户年龄的流程,以便数字服务能根据儿童需求进行定制,并落实符合其年龄段的保护措施。
投诉与补救	儿童可通过该途径举报任何侵犯儿童权利的行为,并申请补救。
儿童友好型信息提供	以儿童易于接受的语言及方式传递信息(例如使用清晰、易懂、简短的语句,仅在需要儿童做出决策时提供必要信息,且采用交互式工具),提升儿童的关注意愿。
儿童参与企业决策	咨询儿童意见,并将其观点纳入数字服务的设计与部署过程。

## 风险评估：

设计保障儿童安全应贯穿产品生命周期的关键阶段<sup>43</sup>：

- 1. 产品构思阶段：**企业对产品进行概念设想、调研并明确产品价值主张，同时制定产品设计方案与概念的阶段。
- 2. 产品开发与测试阶段：**企业进行产品开发与迭代、制作原型，并在产品上线前开展测试的阶段。
- 3. 销售、营销与部署阶段：**企业将产品推向市场、推动产品部署，并在部署后监测产品使用情况以进行后续迭代的阶段。

核桃编程表示，其“设计保障儿童安全”理念的落地，始于产品构思阶段由公司产品开发团队及法律合规团队共同开展的风险管理工作。该公司会通过平台功能（如即时通讯功能、内容分享功能及排行榜功能）的影响评估，分析儿童权利面临的实际风险与潜在风险。这些评估工作包括查阅数字环境中儿童权利相关的现有研究成果，以及咨询儿童权利专家、家长和教育工作者，以充分了解各类特定风险。在设计阶段或测试阶段中发现的新风险或新问题，会被上报至跨部门审核团队。该团队会根据风险的严重程度与紧急程度进行分类，确保及时采取防控措施。

核桃编程表示，其会通过调整部分产品功能来整合评估结果。例如，该公司称已优化内容审核流程，实现对可能违反平台标准的内容进行自动审核与标记。核桃编程还表示，其应用设计保障隐私原则，旨在最大限度减少个人数据收集并保障数据安全。据

悉，平台会更新平台政策，以应对网络欺凌或个人信息未授权分享等风险。即时通讯功能中融入了包括审核协议在内的技术功能，助力降低用户间不当接触的可能性。作为公司更广泛风险防控工作的一部分，这些措施需接受定期审查。

核桃编程指出，“设计保障儿童安全”理念在产品开发与测试阶段持续推进。该公司开展自动化安全评估，通过压力测试检验系统算法是否存在潜在安全漏洞。若识别出风险（如潜在的数据泄露或不当内容分享相关风险），核桃编程会尝试修改产品功能（如内容审核工具），加入更严格的过滤机制。此外，核桃编程会优化用户举报系统，让儿童和家长能更便捷地标记各类不安全行为或内容。

核桃编程称，其会通过定量与定性两种方式衡量风险防控措施的有效性。例如，公司会分析定量数据，包括标记内容的数量、内容审核的速度与准确性，以及与安全问题相关的用户举报频次。这些指标会在不同测试周期内接受检验：例如，不当内容举报量下降，或有害内容标记速度提升，均表明相关措施发挥了成效。

---

43.这是对科技公司产品开发生命周期各阶段的简化概括。不同公司对生命周期阶段的划分可能有所不同。

## 设计保障儿童安全：儿童与家长利益相关方参与

核桃编程表示，其开展的用户测试与原型评估工作以儿童安全为核心关注点。该公司会邀请一组儿童、家长及教育工作者体验平台的早期版本，通过观察他们的使用模式，识别与隐私、数据安全及接触不当内容相关的各类风险。

此外，核桃编程还会收集参与测试阶段的儿童、家长及教育工作者提供的定性反馈。这有助于公司了解用户对安全功能的看法，以及用户在使用平台时是否有安全感。同时，核桃编程还会审查用户满意度调查结果，并组织焦点小组讨论，从用户视角评估风险防控措施的实施成效。

核桃编程表示，在销售、营销与部署阶段，其会监控平台是否存在各类风险，例如即时通讯中的不当互动、未经授权的内容分享或潜在的隐私泄露。之后，公司可能会收紧内容审核过滤机制，或针对不安全行为增设举报渠道。此外，核桃编程会根据反馈意见及监测到的风险，更新用户指南与平台政策，确保用户行为符合儿童安全原则。例如，作为风险防控策略的一部分，可能会出台更严格的应用内沟通规则，或对内容上传作出更明确的规定。

核桃编程指出，在产品部署后，公司会通过多种渠道披露产品对儿童权利产生的影响，以确保透明度，并促进与利益相关方的联动。核桃编程计划发布年度报告，详细介绍其产品开发流程，重点说明在测试及原型制作阶段如何考量并保障儿童权利。这些报告包含多个详细章节，内容涉及儿童保护政策、内容审核系统的有效性，以及基于用户反馈所做的改进。同时，核桃编程还会在其官网发布新闻稿，向公众及利益相关方告知其儿童安全相关规程。

## “设计保障儿童安全”理念在其他流程层面的体现

仅通过单一举措，难以实现面向儿童的“设计保障数字安全”；相反，这需要多项互补且相互关联的举措——这些举措协同作用，方能助力为儿童营造安全且有益的数字环境。

——迈向“设计保障儿童数字安全”<sup>44</sup>

风险评估是“设计保障儿童安全”理念的核心环节，在风险识别、风险防控以及防控措施有效性的跟踪与沟通方面发挥关键作用。然而，正如经济合作与发展组织(OECD)所指出的，企业还需实施其他流程，以确保风险能够得到恰当识别与应对。本案例研究重点介绍了其中四类流程：<sup>45</sup>

- 1) 年龄验证机制
- 2) 儿童友好型信息提供
- 3) 投诉与补救机制
- 4) 企业决策过程中纳入儿童参与或听取儿童意见的机制

44. OECD(2024),《迈向“设计保障儿童数字安全”》(经济合作与发展组织《数字经济文件》), 经济合作与发展组织出版, <https://doi.org/10.1787/c167b650-en>, 第30页。

45. 需说明的是, 本案例研究为简洁起见, 对这些流程的分类与经济合作与发展组织(OECD)的分类略有不同。OECD 的流程分类体系还包含用于判定伤害的技术工具及以儿童为中心的设计; 这些内容的相关方面已纳入本风险评估的“实质内容”部分。

## 1.核桃编程的年龄验证机制

在“设计保障数字安全”理念中,确认用户年龄至关重要:只有明确哪些用户实为儿童,相关服务才能切实保护儿童免受不安全内容与行为的影响。

——迈向“设计保障儿童数字安全”<sup>46</sup>

核桃编程表示,其设有多步骤年龄验证流程。在账号注册过程中,用户需填写出生日期,儿童用户还需获得家长同意。这些数据仅用于判定用户年龄,将存储在加密数据库中,且仅授权人员可访问。该平台对成人用户的使用权限有限制,仅经过验证的辅导老师可使用;未经验证的成人无法向辅导老师发送消息。不过,年龄验证与身份认证措施引发了相关担忧,即可能对儿童隐私及他们参与数字环境的权利产生影响,因此必须在保障安全与保护权益之间维持平衡。

核桃编程进一步表示,其运用自动化技术检测异常内容,一旦检测到便触发验证流程。例如,若儿童用户试图访问被标记为年龄不适宜的内容,核桃编程的系统会根据初始填写的年龄信息,要求用户进行进一步验证,或直接限制访问。进一步验证手段包括借助外部服务开展的文件验证或人脸识别(需获得家长同意)

## 2.核桃编程的儿童友好型信息提供服务

儿童友好型信息提供服务有助于儿童理解:他们所使用的各类数字产品及服务相关的服务条款、这些产品与服务可能存在的风险,以及如何举报伤害行为、标记不安全或非法内容。

——迈向“设计保障儿童数字安全”<sup>47</sup>

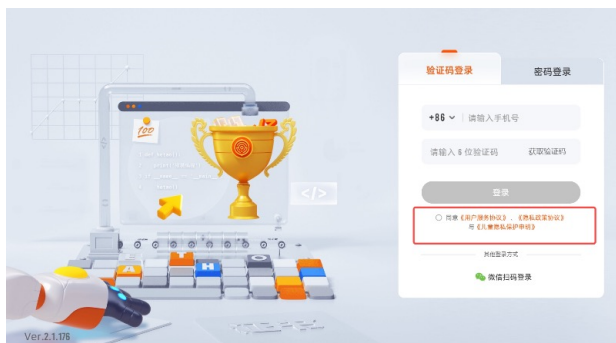
据核桃编程介绍,其在整个平台中均使用与用户年龄相适配的语言,将复杂概念简化为简短、清晰的语句。对于隐私设置、安全指南、编程说明等核心功能,核桃编程会在用户需要做出决策时,有针对性地呈现相关信息。这种信息提供模式旨在避免过多内容使用户感到信息过载,同时在用户需要时提供切实有效的指导。例如,核桃编程的隐私说明如下:

为保障您的账号安全,核桃编程仅向他人共享您的用户名及头像。用户的私人信息(如真实姓名、手机号码)将受到安全保护,未经您或您家长的允许,不会向他人共享。您可随时进入账号设置,修改隐私相关配置。

完成账号注册需提供以下信息:手机号码。若您未提供上述个人信息,我们将无法为您创建账号、完善账号资料或生成用户ID。

46.OECD(2024),《迈向“设计保障儿童数字安全”》(经济合作与发展组织《数字经济文件》),经济合作与发展组织出版,<https://doi.org/10.1787/c167b650-en>,第30页。

47.同上注,第38页。



© 图片由核桃编程提供

核桃编程登录界面，并标注了《用户服务协议》、《隐私政策协议》和《儿童隐私保护申明》的入口

此外，核桃编程表示，其3D图形化编程平台包含多项专为儿童设计的功能，这些功能可帮助儿童熟悉平台操作、理解编程概念，并保障在线安全。具体功能包括操作指南，即一系列关于如何使用平台不同功能的分步教程，例如创建项目、保存进度以及与平台社区互动等操作的指导。

核桃编程的用户体验团队与教育专家及儿童心理学家合作，致力于优化这些沟通策略。

### 3.核桃编程的投诉与补救机制

帮助儿童了解如何举报事件与伤害行为、标记不安全或非法内容，这一点至关重要。若投诉机制难以被儿童找到、获取或使用，其实际效用将大打折扣……因此，儿童友好型投诉机制必不可少。

——迈向“设计保障儿童数字安全”<sup>48</sup>

核桃编程表示，其在用户界面中设置了投诉提交表单，儿童或其家长可通过该表单反馈问题。家长也可代表孩子提交投诉。每一份投诉提交都会通过核桃编程的事件管理系统（见下图）



© 图片由核桃编程提供

核桃编程事件管理系统的后台界面

支持团队会对每个案例进行审核，并将其上报至法务合规团队等相关部门。对于紧急问题，儿童或其监护人也可直接与支持人员沟通。

核桃编程进一步表示，其还提供了其他补救机制，例如封禁违规用户、移除不当内容，以及就问题的解决情况提供反馈。在网络欺凌或隐私泄露等严重情况下，该平台允许用户申请全面调查；必要时，平台还会与有关部门及儿童保护组织开展合作。

### 4.核桃编程在企业决策中纳入儿童参与或听取儿童意见的机制

在涉及儿童自身的事务中，儿童享有表达意见的权利。当儿童（及青少年）参与到影响自身的项目与政策制定过程中时，这些项目与政策更有可能更好地契合儿童的需求、兴趣与背景。

——迈向“设计保障儿童数字安全”<sup>49</sup>

48.同上注。

49.同上注，第40页。

核桃编程表示已建立多项儿童参与机制。该公司会组织焦点小组讨论与用户测试活动，参与者涵盖不同年龄段的儿童。此外，核桃编程还设立了儿童咨询委员会，成员为青少年代表，他们会定期与产品开发团队会面。该委员会为儿童提供了一个反馈渠道，便于他们就平台功能、安全措施及教育内容提出意见。核桃编程与教育工作者及儿童心理学家合作，开发符合年龄适配性且易于参与的咨询方式。该公司表示，未来计划发布反馈报告，以说明这些儿童参与活动对平台更新产生了哪些影响。

### 核桃编程3D图形化编辑平台相关的儿童安全风险

本节将阐述核桃编程3D图形化编辑平台存在的部分儿童安全风险，以及前文所述的“设计保障儿童安全”流程如何帮助核桃编程防控这些风险。需要说明的是，以下仅为安全风险的示例说明，并非详尽无遗。

#### 应对3D图形化编辑平台上的消费类风险

儿童可能会使用家长的账号，或在应用内独立购买产品。若缺乏管控或风险防控措施，该应用内消费场景可能给儿童用户带来财务风险，例如因误操作购买，或通过小额交易过度消费。若没有完善的防护机制，儿童可能会在不知情的情况下订阅相关服务，也可能不理解定期付款的概念，进而导致家长的信用卡被持续扣费。

为防控此类风险，核桃编程表示，应用内消费场景中的所有购买行为均需家长明确同意，儿童无法直接完成交易。具体流程为：核桃编程的课程班主任会直接与儿童家长沟通，获取家长许可；在对儿童的编程能力进行测试后，再为其推荐不同难度级别的编程课程。

此外，核桃编程还指出，其工作人员会向用户详细说明付费后可获取的内容与服务。平台通过安全的第三

方支付网关处理交易，以保护财务信息安全，防止未经授权访问。若儿童账号产生任何财务操作或订阅行为，核桃编程会立即向注册用户（家长）发送相关通知。同时，平台会定期开展审计与监控工作，及时识别并处理消费交易及财务管理相关的潜在问题。

#### 识别并防控隐私风险

3D图形编程平台会收集儿童的姓名、年龄、电子邮箱地址等个人信息，用于账号注册及提供个性化服务。例如，前文提及的年龄验证流程会收集儿童用户的出生日期信息。若缺乏适当管控或风险防控措施，此类数据收集行为可能导致敏感信息面临未经授权访问的风险。例如，若安全措施不到位，恶意人员可能获取儿童个人信息，进而实施身份盗用、未经授权联系或其他侵犯隐私的行为。此外，平台可能会收集儿童的位置数据，以提供基于位置的服务或内容推荐；若没有健全的隐私防护机制，这种行为可能引发监控风险——儿童的活动与位置可能在其不知情或未同意的情况下被监控。此类隐私泄露事件可能使儿童面临恶意人员追踪等有害行为的威胁。

为防控这些隐私风险，核桃编程表示，平台采用端到端加密技术，保障所有数据交换与通信的安全。收集姓名、年龄、电子邮箱地址等个人信息前，均需获取家长的明确同意，且这些信息会存储在安全的加密数据库中。核桃编程实施严格的访问控制机制，确保仅授权人员可接触敏感数据。同时，平台会提供隐私政策，向家长与儿童说明数据的收集、使用及保护方式。平台会尽量减少位置数据的收集，在可能的情况下对其进行匿名化处理，且用户可自主选择开启或关闭基于位置的服务。

核桃编程进一步指出，平台通过高级加密技术保护聊天、消息等通信数据的安全，防止其被未经授权访问或

拦截。家长可通过家长控制功能查看并管理儿童的隐私设置, 监督儿童账号的信息共享对象及内容。为进一步加强隐私保护, 平台会定期开展安全审计与漏洞评估, 识别并处理潜在风险, 为儿童用户营造更安全、更透明的使用环境。

然而, 如前文所述, 关键在于平衡各方需求: 年龄验证与身份认证措施已引发相关担忧, 即这些措施可能对儿童隐私及儿童参与数字环境的能力产生潜在影响。

**维度2: 多样性、公平性与包容性: 涵盖不同类型的儿童群体及童年经历, 尽可能满足各类儿童的使用需求。**

**核桃编程在文化、种族及经济层面的多样性实践**

核桃编程表示, 其会开展市场调研, 以确保无论是原创还是授权的卡通角色设计, 都能充分吸引儿童群体。核桃编程计划推出一项角色定制功能, 允许儿童设计自己喜爱的动画角色, 但相关设计需经过内容审核流程。

核桃编程指出, 其致力于提供以公平为核心的编程课程, 这些课程面向所有儿童设计, 无论他们的背景

或先前经验如何。课程包含适应性学习路径与支持资源。来自该编程平台的用户反馈似乎表明, 这一功能提升了教育公平性, 促进了包容性参与。不过, 平台应更多关注残疾儿童群体的需求。

此外, 核桃编程表示, 为考虑到地区经济差异所导致的教育差距, 其平台产品定价约为市场价格的30%, 为中国三至六线城市的儿童提供了接触编程教育的机会。历年销售数据显示, 来自这些低线城市的用户



© 图片由核桃编程提供  
核桃编程课程剧情与关卡设计中的角色形象

数量正逐步增长。同时，核桃编程已与“美丽中国”公益项目建立战略合作伙伴关系，向偏远山区学校捐赠编程产品，并提供配套的教育及技术支持(见下图)。



© 图片由核桃编程提供  
核桃编程向偏远地区学校捐赠的教学物资

核桃编程表示，该项目目前已覆盖广东、福建、云南三省的20多所学校，让核桃编程的产品能够服务更多儿童，并为经济欠发达地区的儿童提供编程教育。



© 图片由核桃编程提供  
核桃编程员工在偏远地区学校开展的现场教学活动

### 维度3:能力培养:儿童用户对自身效能、能力及技能的认知得到提升,并能体验到掌控感。

#### 核桃编程用户能力培养核心功能

据该公司介绍，核桃编程3D图形编程平台旨在通过交互式实践活动助力儿童编程能力发展。其拖拽式界面将编程概念简化为易于操作的组件，儿童可通过组合代码块创建项目。这种循序渐进的设计有助于儿童构建基础知识、建立信心，降低传统编程语言可能给儿童带来的畏难感。指导性教程提供结构化学习路径，从基础任务起步，难度逐步提升，以支持儿童技能的渐进式发展。

核桃编程表示，协作功能是平台助力技能培养的重要环节。通过联合编程项目，儿童不仅有机会练习技术技能，还能参与团队合作、沟通交流及基础项目协调

工作。这种协作模式旨在鼓励创意分享、任务分配与集体解决问题，有助于培养儿童在编程及其他场景中均适用的人际交往能力。

据悉，该平台还包含进度追踪与强化工具，进一步提升儿童的能力认知感。进度追踪面板通过可视化方式反馈已完成任务、挑战及成就，帮助儿童认识自身进步、明确待提升领域。项目式学习活动鼓励儿童将编程概念应用于游戏、动画等现实场景项目中，在此过程中，他们可不断尝试、精进技能，深化对编程的理解。这些功能共同构建了一个支持性环境，助力儿童稳步提升编程领域的技术能力与创新能力。

## 保障儿童福祉与安全面临的挑战及建议

本案例研究的开展过程中,已识别出若干关键挑战。研究同时提供了相应的建议供企业参考,并明确这些问题在其他致力于保障儿童福祉与安全的科技企业中也普遍存在。需重点说明的是,目前不存在万能解决方案。

**挑战:**如何在实施严格安全措施的同时,兼顾良好的用户体验,是一项具有挑战性的任务。尽管内容审核、家长控制等功能至关重要,但必须找到合理平衡,避免影响儿童与平台进行有效互动的能力。

### 建议:

- 实施可根据儿童年龄、使用模式及行为动态调整的内容审核与家长控制机制,在确保提供有力保护的同时,不过度限制儿童的平台参与度或探索空间。
- 积极邀请儿童与家长参与测试、反馈及协同设计过程,确保安全措施兼具易用性与非侵入性,且能支持儿童与平台开展有效互动。与儿童沟通咨询时,需落实伦理考量。
- 提升产品对残疾儿童的适配性。具体措施可包括:通过增大按钮尺寸简化操作,以及兼容屏幕阅读器等辅助技术。
- 扩大外部咨询利益相关者的范围,在“设计保障儿童安全”理念的持续实施过程中,将儿童权利专家或非营利组织纳入咨询体系,同时需兼顾伦理措施。

**挑战:**如何将风险评估系统且无缝地融入产品开发周期,同时不阻碍创新进程,是一项复杂的任务。

### 建议

- 定期监测新兴数字风险与趋势,据此更新安全协议;为儿童和家长提供教育资源,引导他们负责任、安全且理性地使用平台。
- 依据国际标准,在使用非侵入式、人工智能驱动的年龄验证机制与考量隐私保护之间寻求平衡。
- 采取技术措施以防控更广泛的儿童安全风险,包括检测消息和评论中的有害行为、识别平台上的欺诈交易、排查教育材料中的偏见或歧视内容。企业应对其产品开展儿童权利风险评估。

“服务儿童的负责任科技创新”的8项原则为评估我们的3D图形编程平台如何支持和保护儿童权利提供了一种全面的方法。对于核桃编程而言,这些原则提供了一种结构化的方式,能将儿童福祉融入从概念构思到上线运营后的产品生命周期的每个阶段。通过遵循这些原则,我们可以确保隐私、数字包容性、儿童赋权及安全等问题得到系统性解决。

展望未来,我们计划深化这8项原则的应用,具体措施包括完善内部评估和防控儿童风险的流程。这其中涵盖制定更具针对性的影响评估、扩大与儿童权利专家的合作,以及确保与用户建立持续的反馈循环。

——李军,核桃编程

## 参考文献

1. 经济合作与发展组织, 经合组织儿童数字安全设计圆桌会议, 经合组织出版社;2023年。获取途径:[https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/towards-digital-safety-by-design-for-children\\_f1c86498/c167b650-en.pdf#page=7](https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/towards-digital-safety-by-design-for-children_f1c86498/c167b650-en.pdf#page=7).
2. Livingstone S, Pothong K,《设计中的儿童权利:儿童使用的数字产品与服务创新者指南》。数字未来委员会;2023年。获取途径:[https://digitalfuturescommission.org.uk/wp-content/uploads/2023/03/CRbD\\_report-FINAL-Online.pdf](https://digitalfuturescommission.org.uk/wp-content/uploads/2023/03/CRbD_report-FINAL-Online.pdf).
3. 联合国,《工商业与人权指导原则:落实联合国“保护、尊重和补救”框架》,联合国人权事务高级专员办事处;2011年。获取途径:[https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr\\_en.pdf](https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf).
4. 联合国儿童基金会 (UNICEF) 因诺琴蒂研究办公室,《儿童科技负责任创新 (RITEC):数字技术、玩乐与儿童福祉(第二阶段报告)》,联合国儿童基金会;2024年。获取途径:<https://www.unicef.org/innocenti/reports/responsible-innovation-technology-children>.
5. 联合国,《儿童权利公约》,联合国大会第44/25号决议;1989年。获取途径:<https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child>.
6. 儿童权利委员会,《第25号一般性意见(2021年):儿童在数字环境中的权利》(CRC/C/GC/25),联合国;2021年。获取途径:<https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation>.
7. 经济合作与发展组织,《数字环境中的儿童:风险分类修订版》(《经合组织数字经济论文集第302号》),经合组织出版社;2021年。<https://doi.org/10.1787/9b8f222e-en>.
8. Livingstone S, Stoilova M.《4C框架:儿童网络风险分类》(CO:RE核心议题简报系列),汉堡:莱布尼茨媒体研究所 | 汉斯-布吕多研究所(HBI);CO:RE——儿童网络研究与证据项目;2021年。<https://doi.org/10.21241/ssoar.71817>.
9. 经济合作与发展组织,《迈向儿童数字安全设计》(经合组织数字经济论文集),经合组织出版社;2024年。<https://doi.org/10.1787/c167b650-en>.



携手为儿童

## 联系方式

联合国儿童基金会驻华办事处

地址：北京市朝阳区三里屯路12号

邮政编码：100600

电话：+86 10 8531 2600

传真：+86 10 6532 3107

电子邮箱：[beijing@unicef.org](mailto:beijing@unicef.org)

版权所有 © 联合国儿童基金会驻华办事处, 2026年1月

版权所有 © 中国网络社会组织联合会, 2026年1月

版权所有 © 中国传媒大学, 2026年1月