



How Chinese Technology Companies Apply Principles of Responsible Innovation for Children and Child Safety by Design

January 2026

Disclaimer

The case studies were produced in the context of an ongoing research program on 'Responsible Innovation for Children' and safety by design between the China Federation of Internet Societies (CFIS), China University of Communication (CUC) and the participating Chinese companies, with the support of UNICEF China. The information was documented through a combination of voluntary disclosures from the companies, as well as publicly available information. Selection of the products used as the subject of the case studies was conducted by BSR, a consulting partner of UNICEF.

UNICEF, CFIS, and CUC do not endorse any company, brand, product or service.

This report is a UNICEF publication. Acknowledgements of company representatives do not imply a company's approval or endorsement of the report, and any reference made to a specific company does not imply endorsement by UNICEF of the company's policies and practices. These case studies have been drafted to explore how child safety by design principles are put into practice by Chinese technology companies. Research findings are used solely to highlight current practices and are not meant to offer a comprehensive or conclusive view of other research or of industry practices in any of the topics covered.

Quotes and examples from the research are used solely to highlight potential risks or opportunities for children. These case studies are not intended to be a formal risk assessment, nor do they represent a comprehensive view of the products' child safety risks and opportunities. The case studies are also not meant to offer a comprehensive or conclusive view of research or industry practices in any of the areas covered. These case studies do not articulate, nor does it represent, any official position of UNICEF.

Acknowledgements

This publication results from a collaboration between UNICEF China, China Federation of Internet Societies (CFIS), and Communication University of China.

In total, 19 Chinese technology companies responded to the open call for case study proposals, and 4 of them completed the self-assessment process.

These case study reports were authored by J.Y. Hoh (Business for Social Responsibility – BSR).

Sincere appreciation is extended to Steven Vosloo and Daniel Kardefelt-Winther, UNICEF Innocenti Global Office of Research and Foresight; Alexandra Chernyavskaya, UNICEF EAPRO; Josianne Galea Baron, UNICEF Business Engagement and Child Rights Programme Group Leadership Team, and others who provided helpful inputs and comments on the draft.

Copyright © UNICEF China, January 2026

Copyright © China Federation of Internet Societies (CFIS), January 2026

Copyright © Communication University of China, January 2026

All rights reserved.

Permission is required to reproduce any part of this publication.

TABLE OF CONTENTS

OVERVIEW 2

THE ENHANCED REPRESENTATION THROUGH KNOWLEDGE INTEGRATION (ERNIE) BOT
BY Baidu..... 7

TENCENT MINOR PROTECTION CENTRE 17

THE YBC APP 27

WALNUT PROGRAMMING 3D GRAPHICAL EDITING PLATFORM 36

REFERENCES..... 48

OVERVIEW

In this research document, a simplified definition of “child safety by design”¹ has been adopted. “Child safety by design” refers to companies proactively mitigating negative impacts on children, and maximizing positive impacts, throughout the product development lifecycle.

This series of four case studies explores how child safety by design principles are put into practice by Chinese technology companies, providing actionable insights to help create safer digital environments for children.

The purpose of these case studies is twofold:

- First, to give examples of how to embed child safety early and consistently throughout the product lifecycle.
- Second, to demonstrate how international child and human rights standards, such as the UN Guiding Principles on Business and Human Rights (UNGPs)² and the eight principles of Responsible Innovation in Technology for Children (RITEC),³ may be used to identify and mitigate risk around child safety.

It is important to clarify that these case studies were initiated independently and are not affiliated with UNICEF’s global RITEC research,

although they were supported by the China Federation of Internet Society and UNICEF’s China Country Office and inspired by RITEC. It is also important to note that, although the focus is on safety by design, a broader child wellbeing framework is considered in the analysis.

These case studies focus on products from several of China’s largest technology companies: Tencent, YBC, Walnut Coding, and Baidu:

- 1.A general-purpose generative AI chatbot (Baidu’s ERNIE Bot).
- 2.A service platform that is integrated into various digital services, particularly gaming (Tencent’s Minor Protection Centre).
- 3.A coding education platform (YBC’s YBC App).
- 4.An online platform with educational and community features (Walnut Coding’s 3D Graphical Editing Platform).

1.A variety of definitions have been put forward for child safety by design. Organisation for Economic Co-operation and Development (OECD) (2023). *OECD roundtable on digital safety by design for children*. 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 The OECD has defined child safety by design as meaning that safety is seen as being “a paramount consideration through the whole life cycle of a digital product or service’s design, development and deployment.” The burden of safety “does not lie with children or their parents” – “safety forms a key part of corporate responsibility, and decision-making is child-centred.” For more, see Livingstone, S., & Pothong, K. (2023) *Child Rights by Design: Guidance for Innovators of Digital Products and Services Used by Children*. Digital Futures Commission. https://digitalfuturescommission.org.uk/wp-content/uploads/2023/03/CRbD_report-FINAL-Online.pdf

2.United Nations (2011). *Guiding principles on business and human rights: Implementing the United Nations “Protect, Respect and Remedy” framework*. United Nations Human Rights Office of the High Commissioner. https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf

3.UNICEF Office of Research – Innocenti (2024). *Responsible Innovation in Technology for Children (RITEC): Digital technology, play and child well-being* (Phase Two Report). Available from <https://www.unicef.org/innocenti/reports/responsible-innovation-technology-children>

To identify and select products for these case studies, the China Federation of Internet Societies (CFIS) collaborated with UNICEF's China Country Office and the Communication University of China (CUC) to issue an open call for proposals to technology companies across China. A total of nineteen product proposals were submitted for review, and four were ultimately selected to move forward to the next stage.

Products were chosen for these case studies based on several criteria. First, they collectively cover a range of tech sub-sectors and applications with many child users: generative AI chatbots, edtech, social media and parental control of gaming. Second, they include new technologies, such as generative AI, that may create opportunities to promote child wellbeing, but may also pose emerging risks to child safety. While they are not a formal risk assessment, the case studies describe an illustrative sample of some of these risks, as well as company efforts to identify and address these risks.

After the products were selected, each participating company was asked to respond to a set of self-assessment questions designed to gather detailed information on how their products align with child safety by design principles, and reflect key dimensions of child wellbeing identified by RITEC. UNICEF then invited each company to provide

additional information, guided by a research framework tailored to the specific characteristics of their product.

A conceptual framework inspired by international child rights, safety by design, and wellbeing standards

The conceptual framework applied to these case studies is inspired by a range of child safety and wellbeing standards produced by international organizations. UNICEF, working with the LEGO Foundation's Child Rights and Business Principles, has produced the RITEC framework,⁴ a set of eight principles (autonomy, competence, emotions, relationships, creativity, identities, safety and security and diversity, equity and inclusion) related to child wellbeing and child rights. The United Nations' Guiding principles on business and human rights: (UNGPs)⁵ provide the authoritative methodology for how businesses should carry out human rights and child rights due diligence. The Convention on the Rights of the Child,⁶ and its accompanying General Comment on children's rights in relation to the digital environment,⁷ detail the obligations that exist around child rights. The OECD has also produced several key reports on child safety, including a taxonomy of risks to child safety⁸ and guidance on digital safety by design for children.⁹

4.Ibid.

5.United Nations (2011). *Guiding principles on business and human rights: Implementing the United Nations "Protect, Respect and Remedy" framework*. United Nations Human Rights Office of the High Commissioner. https://www.ohchr.org/documents/publications/guidingprinciplesbusinessshr_en.pdf

6.United Nations (1989). *Convention on the Rights of the Child*. United Nations General Assembly, resolution 44/25. <https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child>

7.Committee on the Rights of the Child (2021). *General comment No. 25 (2021) on children's rights in relation to the digital environment* (CRC/C/GC/25). United Nations. <https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation>

8.OECD (2021). *Children in the digital environment: Revised typology of risks* (OECD Digital Economy Papers No. 302). OECD Publishing. <https://doi.org/10.1787/9b8f222e-en>; Livingstone, S., & Stoilova, M. (2021). The 4Cs: Classifying Online Risk to Children (CO:RE Short Report Series on Key Topics). Hamburg: Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI); CO:RE - Children Online: Research and Evidence. <https://doi.org/10.21241/ssoar.71817> SSOAR

9.OECD (2024). *Towards digital safety by design for children* (OECD Digital Economy Papers). OECD Publishing. <https://doi.org/10.1787/c167b650-en>

Building on these standards, the research team developed a practical framework to guide company engagement and analysis.

The innovation in this work is to turn these international standards into a practical framework, with structured questioning, to guide the operational behavior of companies. The framework was designed by incorporating adapted elements of these standards into structured research questions. Through the process of answering these questions, and through iterative follow-up questioning with the research team, the companies become familiar with these international standards, their requirements, and how they apply to technology product development and deployment.

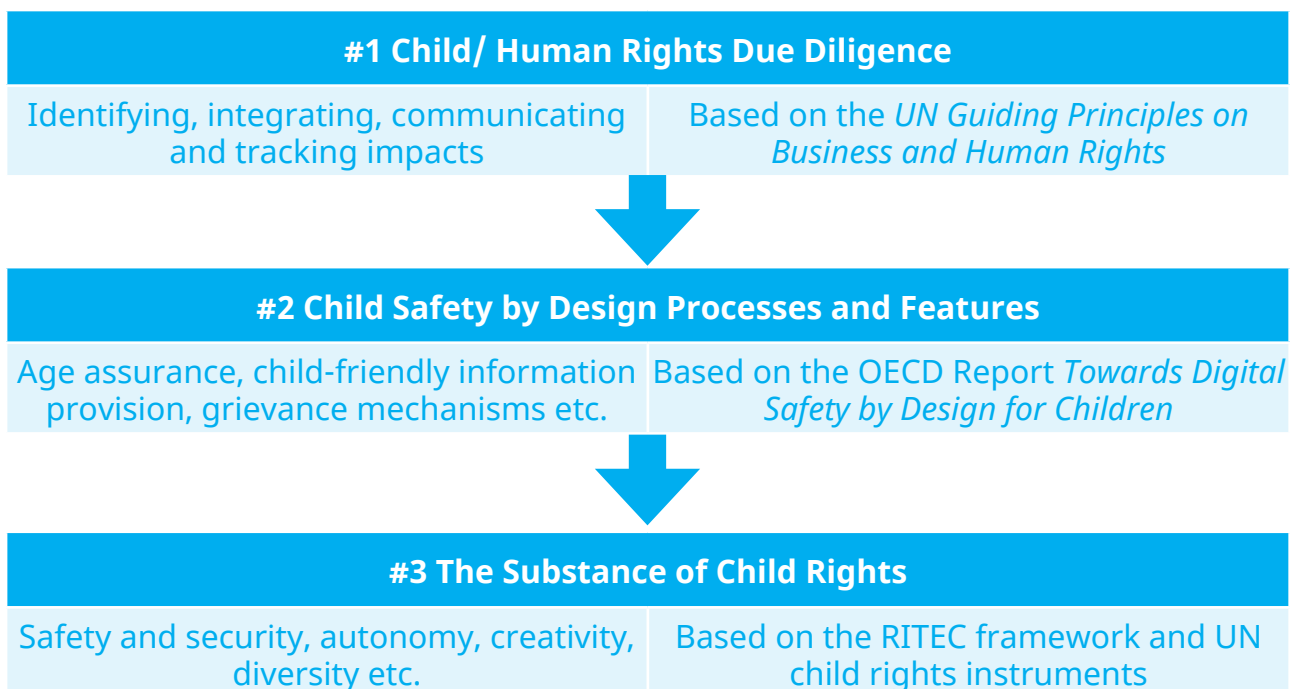
Company answers were distilled into the four attached case studies. These case studies illustrate how applying this framework can help align company practice with international standards, as well as surfacing practical examples of risk identification and mitigation, and areas for improvement. Other companies may also apply the conceptual framework to assess their current

state and chart a plan for further alignment with international standards.

A three-part framework based on international standards

Part 1 of the conceptual framework focuses on the **company's due diligence processes**. Through structured questioning from UNICEF and Business for Social Responsibility (BSR), companies were invited to consider how they incorporated the four core components of human rights due diligence specified in the UNGPs. The UNGPs are the authoritative international standard on responsible business conduct and human rights, and equally apply to the process of securing child rights through safety by design. Principle 17 of the UNGPs lays out the four components of the company's duty to respect human rights:

1. Identifying and assessing risks.
2. Integrating findings from the risk assessment.
3. Tracking and measuring the effectiveness of risk mitigations.
4. Communicating risks and mitigations.



Each of these components was assessed at each of the three phases of a simplified¹⁰ product lifecycle:

1. Product ideation
2. Product development and testing
3. Sales, marketing and deployment.

The conceptual framework’s structure thus requires companies to ensure that the core components of child rights due diligence are incorporated across the product lifecycle. This lifecycle-wide approach is important for child safety by design, which emphasizes proactive rather than reactive risk mitigation.

Part 2 of the conceptual framework focuses on ensuring that companies have considered specific elements of **child safety by design** as identified in the OECD’s 2024 report, *Towards Digital Safety-by-Design for Children*.¹¹ These elements are a mix of **product features or company processes**. Companies are required to describe how their product or processes include the following:

1. Age assurance mechanisms
2. Complaints and redress mechanisms
3. Child-friendly information provision
4. Child participation or consultation in company decision-making.

Through another round of iterative questioning with UNICEF and BSR, companies became familiar with these concepts and the complexities of operationally implementing them. For instance, implementing

effective age assurance should strike a balance between accurately identifying the age of users and respecting their privacy. The case studies demonstrate how companies have attempted to strike this balance and provides suggestions on further areas to consider in order to align with international standards.

Part 3 of the conceptual framework focuses on the **substance** of child safety by design and child wellbeing risks and opportunities. **Substance** refers to the specific interests, such as autonomy or safety, that child safety by design and wellbeing are intended to protect. Companies are required to identify the risks, especially to the safety and security of child users, that their products might pose. Companies were also asked to explain how their products or services may contribute in areas other than safety and security, such as children’s autonomy, creativity or diversity.

The overarching structure of this section is inspired by the eight dimensions set out in the Responsible Innovation for Technology in Children (RITEC) framework.¹² The RITEC project was co-founded by UNICEF, which has played a leading role over the years in promoting wellbeing by design ([click here](#) to view other publications), the LEGO Foundation, and a coalition of academic and research centres.

One of the RITEC framework’s core contributions is the identification of eight dimensions that can be incorporated in the design of products and experiences to promote children’s wellbeing through digital play:

10.This is a simplified approximation of the stages of the product development lifecycle for a tech company. Each company may structure the lifecycle stages differently.

11.OECD (2024). *Towards digital safety by design for children* (OECD Digital Economy Papers). OECD Publishing. <https://doi.org/10.1787/c167b650-en>

12.United Nations Children’s Fund (UNICEF), Office of Research – Innocenti (2024). *Responsible Innovation in Technology for Children (RITEC): Digital technology, play and child well-being* (Phase Two Report). Available from <https://www.unicef.org/innocenti/reports/responsible-innovation-technology-children>

- **Autonomy** – children experience a sense of control and agency and have freedom of choice.
- **Competence** – children have a sense of mastery and feel that they can achieve goals.
- **Emotions** – children experience, are aware of, and can regulate a range of emotions.
- **Relationships** – children experience and manage social connectedness, feel they belong, and are aware of others.
- **Creativity** – children are open to a range of experiences, imagine different possibilities, act on original ideas, and make things.
- **Identities** – children explore, construct, and express facets of themselves and others.
- **Diversity, equity and inclusion** – children feel that their diversity is respected in an inclusive way.
- **Safety and security** – children are robustly protected from safety and security risks.

While RITEC was originally conceived in the context of digital play, these case studies adapt the RITEC principles across different technology subsectors. For instance, relationships, creativity and autonomy are all highly relevant to social media, gaming and edtech applications.

Companies were asked structured questions to demonstrate how their products create risks

or opportunities in a selection of these eight dimensions that are relevant to their products. The case studies emphasize the importance of the “safety and security” dimension, given its unique relevance to the concept of child safety by design.

The conceptual framework’s exploration of the safety and security dimension was further subdivided using the four-part taxonomy developed by Sonia Livingstone¹³ and reproduced and outlined in the OECD’s 2021 Report, [Children in the Digital Environment: Revised Typology of Risks](#).¹⁴ This provides a high-level overview of the digital risk landscape, and outlines four risk categories: contact, conduct, content and consumers. It also identifies and analyses risks that cut across these four risk categories. This taxonomy is widely used by Trust and Safety teams in global technology companies to inform and strengthen their child protection policies and programs. In the course of producing these case studies companies were required to consider a selection of these risks, such as the contact and conduct risks posed by their products, and how they have mitigated them.

13.OECD (2021). *Children in the digital environment: Revised typology of risks* (OECD Digital Economy Papers No. 302). OECD Publishing. <https://doi.org/10.1787/9b8f222e-en>; Livingstone, S., & Stoilova, M. (2021). The 4Cs: Classifying Online Risk to Children (CO:RE Short Report Series on Key Topics). Hamburg: Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI); CO:RE – Children Online: Research and Evidence. <https://doi.org/10.21241/ssoar.71817> SSOAR

14.OECD (2021). *Children in the digital environment: Revised typology of risks* (OECD Digital Economy Papers No. 302). OECD Publishing. <https://doi.org/10.1787/9b8f222e-en>

The Enhanced Representation through Knowledge Integration (ERNIE) Bot by Baidu

The ERNIE Bot, launched by Baidu, is a generative AI chatbot designed mainly for Mandarin speakers in mainland China. It uses advanced multimodal AI technologies that respond to text, images, video and code inputs.

This case study explores how Baidu is working to integrate child wellbeing principles into the ERNIE Bot's design, focusing on three dimensions:

- **Safety and security:** ensuring that children feel safe when using the bot
- **Creativity:** supporting children's curiosity and imaginative learning
- **Autonomy:** enabling children to make choices and control their learning experiences.

The case study highlights how Baidu's has adopted child safety impact assessments, content filtering and privacy measures, as well as interactive educational tools for children. It also discusses the challenges that Baidu faces in balancing rich content with safety, and offers recommendations for improvement.¹⁵

Disclaimer:

UNICEF does not endorse any company, brand, product or service.

This report is a UNICEF publication. Acknowledgements of company representatives do not imply a company's approval or endorsement of the report, and any reference made to a specific company does not imply endorsement by UNICEF of the company's policies and practices.

OVERVIEW

Generative AI uses machine learning to find patterns in existing data and create new content. It can produce human-like text, images, audio and code, and support complex tasks.¹⁶

Generative AI can support children's learning, foster their creativity and improve their communication. However, children may also face increased risks, such as being exposed to harmful or inappropriate content, manipulation or abuse through personal interactions, or misuse of their data. Designing

15. This work draws on a separate project, Responsible Innovation in Technology for Children (RITEC), which is a collaboration between UNICEF and The LEGO Group, and is funded by The LEGO Foundation. It identifies eight principles (autonomy, competence, emotions, relationships, creativity, identities, safety and security and diversity, equity and inclusion) which can be used to assess child well-being and child rights. For more information see: <https://www.unicef.org/childrightsandbusiness/workstreams/responsible-technology/online-gaming/ritec-design-toolbox>.

16. UNICEF (2023), Innocenti Global Office of Research and Foresight, Generative AI: Risks and opportunities for children, <https://www.unicef.org/innocenti/generative-ai-risks-and-opportunities-children>.

AI products with child safety in mind is crucial to harnessing their benefits.

This case study focuses on three aspects of child wellbeing in relation to Baidu’s ERNIE (Enhanced Representation through Knowledge Integration) Bot:

- **Safety and security:** Baidu uses child rights impact assessments to identify risks such as exposure to violence, discrimination and false information (through AI “hallucination”). It manages these risks through content filtering and data privacy protocols, combining human oversight and AI-powered detection.
- **Creativity:** the ERNIE Bot enables children to learn through virtual experiments and creative activities that foster learning and innovation.

- **Autonomy:** the ERNIE Bot provides personalized guidance to support children’s independent learning.

Baidu faces challenges in balancing content richness with simplicity and safety. Due to generative AI’s probabilistic nature inappropriate content can sometimes be generated, so filtering and moderation need to be improved continually. Children’s changing needs also call for safety measures to be continuously evaluated and adapted. Baidu could further enhance its approach by engaging with child safety experts and other external stakeholders, and involving children directly in the design process.

CHILD WELLBEING DIMENSIONS IN ACTION

Dimension 1. Safety and security: children feel safe and are kept safe.

This case study examines key elements of Baidu’s processes for ensuring that children are safe and secure:

Safety and security processes	
Risk assessments	Identifying child safety impacts, and then integrating, tracking and communicating risk mitigation measures.
Age assurance	Processes to estimate or verify the age of a user so that digital services can be tailored to their needs and age-appropriate safeguards put in place.
Complaint and redress	A means through which children can report any child rights violations and request redress.
Child-friendly information provision	Children are spoken-to in their own language and in a manner that makes them want to pay attention, such as through clear, simple and short sentences, providing information only when it is necessary to make a decision, and using interactive tools.
Child participation in company decision-making	Children are consulted and their views incorporated into the design and deployment of digital services.

Risk assessments

Risk assessments should occur at key stages¹⁷ throughout the product lifecycle:

1. During **product ideation**, when companies are conceptualizing, researching and identifying the value proposition for a product, creating design plans and concepts for it.
2. During the **product development and testing** phase, when companies are building and iterating on the product, creating prototypes and conducting tests prior to launch.
3. During the **sales, marketing and deployment** phase, when companies push the product to market, enable its deployment, and monitor product use post-deployment for further iterations.

Baidu introduces “child safety by design” concepts at the **product ideation** stage. The product manager, child safety experts and compliance team collaborate on a children's rights impact assessment, to identify actual and potential risks that may affect child users. This includes different methods, including analysis of product features to assess potential risks, such as privacy breaches, addiction and seeing inappropriate content. Special attention is paid to children's cognitive levels and their differing needs.

According to these assessments, the anthropomorphic (human-like) communication style of the ERNIE Bot may cause children to form an emotional dependence on or excessive trust in AI, and children may inadvertently disclose personal information during interactions with the chatbot. Based on

the assessment results, the design team and product development team have jointly established design specifications, integrating risk mitigation measures into the product design. These measures include age-appropriate content filtering features, watermarking to indicate AI-generated content, and privacy protection functions. Simulated testing and expert reviews are used to make a preliminary assessment of the effectiveness of risk mitigation measures in the design plan.

Baidu reports that child safety by design continues during the **development and testing phases**, in which the development team and child safety experts carry out code reviews and privacy impact assessments to identify risks that may affect children in using the product, such as data leaks and product misuse. Based on the assessment results, the development team integrates features such as age verification and privacy protection into the product, to mitigate potential risks. The content review system is also developed and debugged during this phase, to ensure that it can identify and filter content that is inappropriate for children. Users can report violations by calling Baidu's reporting hotline, sending emails or clicking the online reporting button in the app. Baidu's customer service operates 24/7 to process user feedback.

Baidu reports that during the **sales, marketing and deployment stages**, the marketing team and child safety experts assess risks to children's rights arising out of marketing strategies. For example, they will assess whether ads have sexually suggestive visuals or language. They also assess whether the ads contain persuasive techniques

17. This is a simplified approximation of the stages of the product development lifecycle for a tech company. Each company may structure the lifecycle stages differently.

targeted at children, such as featuring animated characters or celebrities, or directing child users to ask their parents to make purchases. Additionally, the data privacy team assesses the risk of personal data breaches during product deployment. The effectiveness of mitigation measures is assessed through automated testing and simulations of how a child user might behave.

Other procedural aspects of child-safety-by-design

Digital safety by design for children is unlikely to be achieved by just one action, instead requiring several complementary and overlapping actions that, when combined, can help to ensure a safe and beneficial digital environment for children.

– Towards Digital Safety by Design for Children¹⁸

Risk assessments are a core aspect of child safety by design, playing a key role in the identification of risks, mitigation of those risks, and tracking and communicating mitigation effectiveness. However, as the OECD argues, companies must enact other processes to ensure that risks may be appropriately identified and addressed. This case study highlights four of these workflows:¹⁹

1) Age assurance mechanisms

- 2) Child-friendly information provision
- 3) Complaint and redress mechanisms
- 4) Child participation/ consultation in company decision-making

1. Baidu's age assurance mechanisms

Establishing age is important for digital safety by design, as services are only going to be able to protect children from unsafe content and behaviours if they know which of their users are actually children.

– Towards Digital Safety by Design for Children²⁰

While age assurance mechanisms are important elements of child safety, international standards require them to be designed in a way that adequately balances child users' rights, such as the rights to privacy or access to information. Baidu has implemented an age assurance mechanism applicable to all of its products, including the ERNIE Bot. Children cannot create accounts by themselves; their parent or guardian must create the account for them. Both parties must read the user agreement, which lists all the types of data that will be collected and processed (explained further below). Nevertheless, the impact of this approach has not been assessed.

18.OECD (2024), Towards digital safety by design for children (OECD Digital Economy Papers), OECD Publishing, <https://doi.org/10.1787/c167b650-en>, p. 30.

19.Note that in this case study these workflows are categorized slightly differently from the OECD, for the sake of simplicity. The OECD's taxonomy of workflows also includes technical tools to determine harm and child-centered design; aspects of these are covered in the "substance" segment of this risk assessment.

20.OECD (2024), Towards digital safety by design for children (OECD Digital Economy Papers), OECD Publishing, <https://doi.org/10.1787/c167b650-en>, p. 30.

The Enhanced Representation through Knowledge Integration (ERNIE) Bot by Baidu

During account creation, which is necessary to access the ERNIE Bot, Baidu collects obfuscated age-range data to determine whether the user is under 14 years old. For instance, Baidu may utilize optional self-declared information provided at registration, such as an age range or partial birth dates (month and year). The company reports that it does not collect or store data such as facial features or fingerprints. There is no data storage in the background of the application, and information about the user's age is discarded after it has been used for age assurance.

2. Baidu's child-friendly information provision

Child-friendly information provision helps children to understand the terms of service associated with the various digital products and services they engage with, any risks associated with these products and services, and how to report harms and flag unsafe or illegal content.

– Towards Digital Safety by Design for Children²¹

During the development phase, the ERNIE Bot's design team considered how information is presented in the ERNIE Bot's user interface to ensure that children can easily understand it. For instance, the vocal training AI adopts a clear and intuitive interface design, utilizing voice guidance and visual cues to help children easily understand the training content (see below).



© Courtesy of Baidu

The vocal training AI of the ERNIE Bot's user interface.

Additionally, Baidu's product team worked with their user experience team to carry out interaction design tests tailored for children. Multiple rounds of user testing were conducted to ensure that information presentation aligns with children's cognitive abilities, using interactive animations, icons and other methods to help children understand the product's operational flow.

21.Ibid., p. 38.

3. Baidu's complaint and redress mechanisms

Helping children understand how to report incidents and harms, and flag unsafe or illegal content is essential. If complaints mechanisms are hard for children to find, access or use, they will have little utility [...] child-friendly complaint mechanisms are essential.

– Towards Digital Safety by Design for Children²²

Baidu reports that it has implemented a reporting mechanism for the ERNIE Bot and provides continuous customer service to handle user feedback, including regarding harmful content. Reported complaint categories include harmful information, pornography and vulgar content, as well as abusive and offensive behaviour. Reports of sexual abuse will be escalated and prioritized. Users can report violations with a single click of an in-app reporting button. Baidu reports that its customer service team responds to complaints within three working days and provides an appeals channel to ensure that issues are thoroughly resolved. Users can appeal Baidu's initial feedback or responses, and dedicated personnel are assigned to follow-up on the complaint and review it further, to ensure that the problem is resolved.

4. Baidu's processes for child participation/consultation in company decision-making

Children have a right to be heard in matters that affect them, and when children (and young people) play a part in helping to develop programs and policies that will affect them, it is more likely that they will be better aligned with children's needs, interests and backgrounds.

– Towards Digital Safety by Design for Children²³

Baidu reports that it currently consults with parents to understand the needs and preferences of children. However, while feedback from parents is valuable, direct engagement with children is the ideal method to obtain the most accurate representation of their views. Baidu should consider direct engagement with children in the future, where practicable.

Child safety and security risks from Baidu's ERNIE Bot

This section illustrates some child safety and security risks that arise from the ERNIE Bot, and how the child safety by design processes described above have enabled Baidu to mitigate those risks. These are intended to be illustrative, rather than exhaustive, examples of safety and security risks.

Generative AI chatbots pose a broad spectrum of safety and security risks to children. The chatbots may generate harmful content, such as images or text depicting violence or gore, as well as sexual

22.Ibid.

23.Ibid., p. 40.

The Enhanced Representation through Knowledge Integration (ERNIE) Bot by Baidu

content. These chatbots may also be used to generate child sexual abuse material. Generative AI chatbots may also create harmful bias, as they absorb and reproduce the biases that are found in their training data: children who use these apps may unconsciously internalise the biases or stereotypes produced by these chatbots. Chatbots also create an additional risk through dangerous instructions: they may instruct children on how to harm themselves or others, and have in some extreme cases directed users to commit suicide.

The company tries to mitigate these risks by leveraging the capabilities of large AI models, such as the ERNIE Bot, to review and filter harmful content related to child safety in its products. Although large

language models still have limitations, their natural language capabilities can be leveraged to detect and filter certain types of content and behavior that may be harmful to children. In the example below, Baidu's generative AI systems attempt to identify and filter out different code words that bad actors use to disguise their attempts to solicit children. However, handling these risks remains an important challenge for Baidu and other tech companies.

Safety and security: using generative AI models to detect child sexual abuse behavior

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。

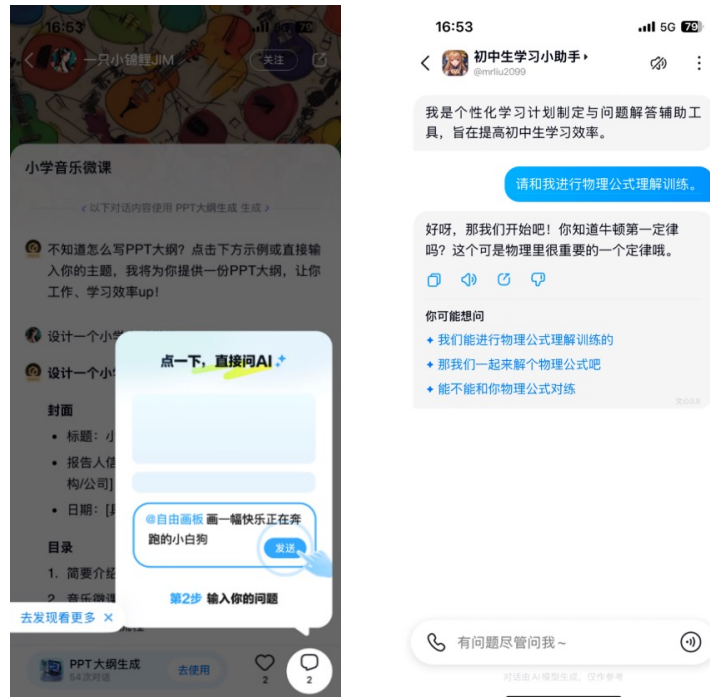
© Courtesy of Baidu

This image depicts how large language models can detect codewords used by sexual predators online.

Dimension 2. Creativity: child users engage curiously and use their imagination to build, invent and experiment.

Incorporating creativity into the design of the ERNIE Bot

Baidu reports that the ERNIE Bot includes interactive tools designed for children, such as virtual experiments and creative games. These features aim to support engagement in science and art-related topics. Users can access functions such as question banks, concept explanations and personalized study plans, which may align with diverse learning preferences.



© Courtesy of Baidu
The ERNIE Bot's user interface.

Dimension 3. Autonomy: child users freely choose how to engage, and experience feelings of agency, choice and freedom.

Incorporating autonomy into the design of the ERNIE Bot

Baidu reports that the ERNIE Bot supports children's autonomy, by enabling minors to set their own learning goals and adjust difficulty levels based on their individual needs, which may enhance their ability for self-directed learning. The ERNIE Bot application also offers a wide range of educational AI agents designed for primary and secondary school students, enabling minors to select learning or entertainment options that align with their cognitive levels.

Baidu reports that child user autonomy is also further supported through the ERNIE Bot's automated provision of guided assistance. The chatbot can be customized to provide feedback without spoon-feeding students with the answer, enabling students to enhance their problem-solving skills without getting stuck. This approach may contribute to increased student motivation and confidence, as well as their ability to learn autonomously, outside of dedicated classroom time.

Challenges and recommendations for protecting child wellbeing and safety

The process of creating this case study has identified several key challenges. Corresponding recommendations have been provided for the benefit of the company concerned, acknowledging that these issues are also common among other tech companies striving to protect child well-being and safety. It is important to note that there are no one-size-fits-all solutions.

Challenge: Difficulties maintaining a diverse set of features for the ERNIE Bot while ensuring simplicity of information provision for children, a core element of child safety by design.

Recommendations:

- Consider gaining a comprehensive understanding of children's cognitive levels and psychological characteristics through research and by consulting experts and children, as well as repeated testing and adjustments.
- Based on evidence, consider implementing tiered information design by tailoring content complexity to different age groups, introducing an adaptive mode switch (such as a dedicated "child mode" with simplified interactions and optional advanced mode requiring parental authorization), and adopting a task-based interaction approach to prioritize common child-focused use cases such as homework help and factual queries, reducing unnecessary complexity.

Challenge: While the content generated by large language models can provide knowledge and information, generative models' probabilistic algorithms may inadvertently generate inappropriate content. One example of this relates to bias. For instance, choosing young and

conventionally attractive women as AI agents may be received positively by users, but may also reinforce stereotypes about female assistants.

Recommendations:

- Ensure that training data is diverse and balanced across gender, age, ethnicity and other categories, design AI agent personas to be neutral or varied rather than stereotypical, and strengthen both real-time content filtering and manual moderation processes.
- Regular audits of AI outputs can help identify emerging biases or problematic content, while user feedback mechanisms for children and guardians can provide additional insights to guide ongoing model improvements.
- Undertake more evaluations of the approaches used and of the user experience.

Challenge: Another risk relates to anthropomorphism: younger children may believe that AI agents with personas are more trustworthy, which may make them less likely to verify information in their outputs. Children are particularly prone to viewing AI chatbots as "quasi-human confidants." But chatbots may fail to recognize dangerous situations or provide inappropriate advice when empathy and understanding are required: they may also fail to provide developmentally appropriate advice. This can lead to confusion, frustration or harm for children.

Recommendations:

- Use design-less human-like personas with factual, neutral language to reduce perceived human identity.
- Another effective method is to embed transparency cues that regularly remind children the AI is not human and may make mistakes.

- Include prompts encouraging verification, such as “check with a trusted adult,” or provide sources and offer parental guidance tools to help educate children on AI limitations and the importance of cross-checking information.

Challenge: Striking a balance between safety and innovation. The company needs to keep children safe from harmful content while also supporting open learning and exploration processes.

Recommendations:

- Recognize that child safety is not just a technical issue and requires responsibility and awareness from a broad range of stakeholders, including company staff and users. The company should prioritize collaboration between different teams, communication with parents and children, and adherence to policies and regulations.
- Consider incorporating predictive or forecasting strategies, such as futures methodologies, and review peer companies’ policies and practices.
- Continue to design policies to reflect the reality that child users may intentionally try to neglect platform rules out of curiosity, challenging or testing the system boundaries. This might include providing these child users with dedicated explanations, tailored to their level of understanding, about why the platform’s rules exist.
- Take into account the balance between safety and privacy in the age assurance process. For instance, the company may consider whether utilizing voice analytics, a form of biometric data, in addition to assessment of account behavior, is in line with data minimization and protection of privacy.

- Existing mitigation methods, such as providing child users and their guardians with notice of data processing and the de-identification and deletion of the data, will continue to be core tools in ensuring that this balance is struck.

It is important to recognize the unpredictability of AI-driven chatbots, which leads to many safety risks, and that in general a one-fit solution has not been identified. While Baidu has made some initial efforts to mitigate risks related to harmful content and behaviours, it may consider undertaking more evaluations of the approaches used, and of the user experience. Methods such as “system cards” (that document how the company has tested their products, and the results) could be explored.

From an enterprise perspective, the RITEC-8 principles are not only standards for fulfilling social responsibility but also the cornerstone of building a positive corporate image and enhancing user trust. In designing children's products, companies must thoroughly understand how to promote children's "competence" development and how to support their "emotional regulation." This approach not only helps products gain the trust of parents and educators but also enhances their market competitiveness. Moreover, these principles offer comprehensive guidelines for risk identification and mitigation, ensuring that products can effectively avoid potential legal and social risks caused by design flaws.

– Cai Tingting, Ma Wentao, Wu Qiang, Zhao Mengxi (Baidu)

Tencent Minor Protection Centre

The Tencent Minor Protection Centre is an integrated platform designed to strengthen child online safety and family education. At its heart is a parental control system which is built into Tencent apps, helping families to manage children's access to games, social media and digital content. The Centre's AI model not only powers parental tools and educational content, but also supports families with interactive guidance and case studies.

This case study explores how Tencent is working to address two key challenges for children growing up in the digital world: safety & security (including age assurance, content filtering and risk assessments) and relationships (through tools that encourage healthy family and social connections).

By combining technology, family education and stakeholder engagement, Tencent aims to balance child safety with digital inclusion, while also recognizing the ongoing need for feedback, impact analysis and adaptation as risks evolve.²⁴

Disclaimer:

UNICEF does not endorse any company, brand, product or service.

This report is a UNICEF publication. Acknowledgements of company representatives do not imply a company's approval or endorsement of the report, and any reference made to a specific company does not imply endorsement by UNICEF of the company's policies and practices.

OVERVIEW

Tencent, a global tech company headquartered in China, has a diverse portfolio of tech products and services, including gaming, social media, music streaming, payments and instant messaging. These products present a diverse range of opportunities and risks to children.

To mitigate some of these risks, Tencent has established the Minor Protection Centre, and

built a service platform for parents. This parental control platform is integrated into apps within Tencent's ecosystem, particularly for managing access to games, social media and other digital content. The system includes tools such as the Tencent Public Welfare Home Education AI Model, which offers educational content, question-and-answer functionalities and case studies. Parental controls provides some safeguards tailored to developmental stages. At the same time, child safety

24. This work draws on a separate project, Responsible Innovation in Technology for Children (RITEC), which is a collaboration between UNICEF and The LEGO Group and funded by The LEGO Foundation. It identifies eight principles (autonomy, competence, emotions, relationships, creativity, identities, safety and security and diversity, equity and inclusion) which can be used to assess child well-being and child rights. For more information see: <https://www.unicef.org/childrightsandbusiness/workstreams/responsible-technology/online-gaming/ritec-design-toolbox>

by design is necessary to ensure that platforms are safe but do not unduly restrict openness, ensuring that children benefit from digital services.

This case study focuses on two key dimensions of child wellbeing – safety & security and relationships. Tencent reports that the Minor Protection Centre prioritizes child safety by design through features such as age assurance, content filtering and gaming time restrictions. The company indicates that its safety and security measures include risk assessments during development and direct engagement with children. Furthermore, the Minor Protection Centre provides AI-driven educational tools, under the Family Guardian Program, that support parental personalized learning and the problem-solving of family issues.

Tencent faces challenges in balancing its commercial and social responsibilities, particularly in mitigating consumption risks to children. Parental control platforms can assist child safety and security, including inbuilt age assurance mechanisms, content filtering and restrictions on excessive gaming time. These may mitigate some child safety and security risks, but more analysis of the impact of these approaches needs to be carried out.

Furthermore, it will be important to balance risk mitigation with the openness of products and services, so that children’s rights in the digital world are adequately respected. Future growth opportunities include expanding external stakeholder engagement, such as consulting more directly with children and child safety experts, and formalizing due diligence processes across the product lifecycle.

CHILD WELLBEING DIMENSIONS IN ACTION

Dimension 1. Safety and security: children feel safe and are kept safe.

This case study examines key elements of Tencent’s processes for ensuring child safety and security:

Safety and security processes	
Risk assessments	Identifying child safety impacts, and then integrating, tracking and communicating risk mitigation measures.
Age assurance	Processes to estimate or verify the age of a user so that digital services can be tailored to their needs and age-appropriate safeguards put in place.
Complaint and redress	A means through which children can report any child rights violations and request redress.
Child-friendly information provision	Children are spoken-to in their own language and in a manner that makes them want to pay attention, such as through clear, simple and short sentences, providing information only when it is necessary to make a decision, and using interactive tools.
Child participation in company decision-making	Children are consulted and their views incorporated into the design and deployment of digital services.

Tencent's safety and security risk assessments throughout the product lifecycle

Child safety by design should occur at key stages²⁵ in the product lifecycle:

1. During **product ideation**, when companies are conceptualising, researching and identifying the value proposition for a product, creating design plans and concepts for the product.
2. During the **product development and testing** phase, when companies are building and iterating on the product, creating prototypes and conducting tests prior to launching the product.
3. During the **sales, marketing and deployment** phase, when companies push the product to market, enable its deployment, and monitor product use post-deployment for further iterations.

Tencent reports that its approach to child safety by design begins at the **product ideation phase**, where a preliminary impact assessment is carried out to identify potential impacts on child rights. Cross-functional team meetings are held to review the results of these assessments and to determine the risk mitigation measures required. Dedicated teams work together with different roles during this stage: the family education advisory team oversees the conception and design of the education model, the family education team manages the content operation, and the product team takes responsibility for development.

Tencent reports that during the **product development and testing phase**, more detailed impact assessments

are carried out to evaluate and prioritize risks based on their severity and likelihood, aiming for high-risk issues to be addressed first. Prototypes are created and subjected to testing, especially for child users, to meet safety and ethical standards. The family education advisory team carries out field research, using questionnaires to collect data directly from communities, schools and families. A cross-departmental team – including members from product development, legal compliance and social responsibility – collaborates in assessing the product's impact on children's rights.

Tencent reports that regular risk assessments are carried out using data tools to identify potential issues. Feedback from parents and child users is also collected and analyzed to refine the product further. Throughout this phase, Tencent communicates with government regulators to ensure compliance, and leverages insights from academic partnerships and non-profit collaborations to optimize the educational model. Iterative improvements are promoted, with a focus on enhancing the model's human-like features to enable engaging interactions with users.

Tencent reports that during the marketing, sales and deployment phase, it seeks feedback from users to identify and resolve any risks or challenges that arise post-launch. The company receives feedback from parents and child users for ongoing improvement. Tencent also reaches out to non-profit organizations and academic institutions to share insights and evaluate the product's impact in line with child safety principles.

25. This is a simplified approximation of the stages of the product development lifecycle for a tech company. Each company may structure the lifecycle stages differently.

Other procedural aspects of child safety by design

Digital safety by design for children is unlikely to be achieved by just one action, instead requiring several complementary and overlapping actions that, when combined, can help to ensure a safe and beneficial digital environment for children.

– Towards Digital Safety by Design for Children²⁶

Risk assessments are a core aspect of child safety by design, playing a key role in the identification of risks, mitigation of those risks, and tracking and communicating the effectiveness of mitigation. However, as the OECD argues above, companies must follow other processes to ensure that risks may be appropriately identified and addressed. This case study highlights four of these workflows:²⁷

- 1) Age assurance mechanisms
- 2) Child-friendly information provision
- 3) Complaint and redress mechanisms
- 4) Child participation/ consultation in company decision-making

1. Tencent's age assurance mechanisms

Establishing age is important for digital safety by design, as services are only going to be able to protect children from unsafe content and behaviours if they know which of their users are actually children.

– Towards Digital Safety by Design for Children.²⁸

Tencent reports that its age assurance system incorporates mandatory real-name registration and random facial recognition checks. The company indicates that all users must undergo real-name verification through China's public security identification system, which automatically identifies whether the user is a child. This identification reportedly triggers the Minor Protection Centre's processes for addressing unhealthy video game use.

AI facial recognition is also used to verify children who are suspected of using adult accounts. The company indicates that it does not store whole face images; instead, these are further processed and stored separately from the identity ID to prevent data leakage. Tencent reports that after a limited period this information is deleted or anonymized, and that data storage and transmission is encrypted.

2. Tiered control strategy

In terms of time management, minors are allowed to access games only on Friday, Saturday and

26.OECD (2024), *Towards digital safety by design for children* (OECD Digital Economy Papers), OECD Publishing, <https://doi.org/10.1787/c167b650-en>, p. 30.

27.Note that these workflows are categorized slightly differently from the OECD in this case study, for the sake of simplicity. The OECD's taxonomy of workflows also includes technical tools to determine harm and child-centered design; aspects of these are covered in the "substance" segment of this risk assessment.

28.OECD (2024), *Towards digital safety by design for children* (OECD Digital Economy Papers), OECD Publishing, <https://doi.org/10.1787/c167b650-en>, p. 30.

holidays from 8 PM to 9 PM²⁹ (a legal requirement in China). When children interact with digital products that allow payments via direct debit, restrictions are set on the total daily expenditure. Users under 8 are prohibited from recharging; users aged 8-12 can recharge a maximum of 50 yuan per transaction and a total of 200 yuan per month; users aged 13-18 can recharge a maximum of 100 yuan per transaction and a total of 400 yuan per month. Furthermore, if a user's total daily spending exceeds 500 yuan, the system will trigger an abnormal payment alert and send a text reminder to the parents.

3. Family coordination tools

The company indicates that the platform allows parents to link their children's game accounts, enabling them to monitor gameplay duration and spending records in real time. Parents can also set customized time restrictions (e.g., allowing gameplay only on weekends), facilitating better management of their children's internet usage.

4. Tencent's child-friendly information provision

Child-friendly information provision helps children to understand the terms of service associated with the various digital products and services they engage with, any risks associated with these products and services, and how to report harms and flag unsafe or illegal content.

– Towards Digital Safety by Design for Children³⁰

Tencent reports that it utilizes interactive media such as games, animations and drawings to make the delivery of information interesting and to attract children's attention (see images below).



© Courtesy of Tencent

Animated videos produced by Tencent to educate children about online safety.

Additionally, Tencent states that its platform offers animated videos on internet safety issues specifically designed for children. For instance, The Minor's Fraud Prevention Illustrated Guide incorporates common online scams into nursery rhymes, accompanied by animated scenes that engage the attention of child users. It is reported that these animated videos have surpassed one million views and are shown in schools.

29.This follows the approach outlined in the National Press and Publication Administration's circular on "Further Strengthening the Management and Effectively Preventing Minors from Becoming Addicted to Online Games."

30.OECD (2024), Towards digital safety by design for children (OECD Digital Economy Papers), OECD Publishing, <https://doi.org/10.1787/c167b650-en>, p. 38.

5. Tencent's complaint and redress mechanisms

Helping children understand how to report incidents and harms, and flag unsafe or illegal content is essential. If complaints mechanisms are hard for children to find, access or use, they will have little utility [...] child-friendly complaint mechanisms are essential.

– Towards Digital Safety by Design for Children³¹

Helping children understand how to report incidents and harms, and flag unsafe or illegal content is essential. If complaints mechanisms are hard for children to find, access or use, they will have little utility [...] child-friendly complaint mechanisms are essential.

– Towards Digital Safety by Design for Children

Tencent's platform includes numerous reporting channels for child users to provide feedback on technical issues or to report harmful information and illegal activities. Currently, the company indicates that there are four complaint channels:

1. Direct feedback and complaints through the platform
2. Sending issues to the reporting email address
3. Reporting via Tencent's dedicated customer service channel
4. Mailing concerns to Tencent's Data Privacy Protection Department.

Tencent reports that, upon receiving a concern, it commits to reviewing and providing feedback within 15 days.

The company notes that all four complaint channels available to child users require assistance and consent from their guardians. Of these four, some may be more challenging to use by children, such as the reporting of concerns to Tencent via email. These are likely to require significant assistance or direction from the users' guardians. Others, such as the in-app one-click reporting function through the platform, may be easier for children to utilize with minimal support.



© Courtesy of Tencent
Interface of Tencent's in-platform direct feedback and complaint channel.

31. Ibid.

6. Tencent's child participation/ consultation in company decision-making

Children have a right to be heard in matters that affect them, and when children (and young people) play a part in helping to develop programs and policies that will affect them, it is more likely that they will be better aligned with children's needs, interests and backgrounds.

– Towards Digital Child Safety by Design³²

Tencent reports that it engages directly with children to understand their expectations and needs for digital services through interviews, surveys and group discussions. The company observes children's behaviors and reactions to digital services to identify their interests and challenges. Child participation in research is voluntary, with the consent of a parent or guardian. Data collected through this research is anonymized, subject to minimization principles, and collected in accordance with Chinese data protection law.

To contribute to awareness-raising around online safety, the company disseminates information on common scams and protective strategies in child-friendly formats, such as comics and interactive games. It also organizes roadshows and classes, primarily in urban areas, to promote online safety for young people. Each year, children are invited to visit the Tencent Minor Protection Centre, where a dedicated cybersecurity house provides a hands-on, interactive experience that combines play with learning about cyber protection.



© Courtesy of Tencent

A roadshow organized to promote online safety for young people.

Safety and security risks from Tencent products, and the Minor Protection Centre

This section illustrates some child safety and security risks from Tencent's products that are mitigated by the Minor Protection Centre, and how the child safety by design processes described above have enabled Tencent to mitigate those risks. These are intended to be illustrative, rather than exhaustive, examples of safety and security risks.

Unhealthy engagement with video games

One of the potential risks of gaming products is that that child users develop unhealthy engagement with video games. This unhealthy engagement, characterized by the displacement of essential activities such as sleep or physical exercise, can lead to adverse health outcomes. According to data from the Tencent Minor Protection Centre in 2024, boys aged 9-12 (representing 82% of the whole) are a high-risk group for video game-related issues, and parents generally lack effective educational methods for dealing with this. Within families predisposed to unhealthy engagement with

32.Ibid.

gaming, 39.3% exhibit permissive parenting, 20.36% are overprotective (with excessive safeguarding), and 14.98% are disengaged (lacking understanding of their child's behavior).

Tencent reports that it attempts to deal with the risk of unhealthy engagement with games through a set of pre-, in- and post-game measures. These measures reportedly include educational tools, one-on-one expert guidance, and parental control of minors' gaming hours through the platform.

Addressing content- and conduct-related risks

Tencent's suite of products may expose children to content-related risks that compromise their safety and security. In gaming communities, children may encounter age-inappropriate or explicit content, including violence or sexual themes, which can surface in chats and user-generated materials.

In addition, there are numerous conduct-related risks: one important example is that online games are widely used by abusers for grooming. The potential for cyberbullying and exposure to potential abusers also increases within these online ecosystems, where real-time interactions are not always fully moderated.

To mitigate these risks, Tencent has tried to develop different strategies, such as a safety policy, a

children's model, technical monitoring, human-computer collaborative review, a feedback reporting mechanism and other measures. These combine automated and manual moderation workflows. For instance, if a user posts, "I saw someone fighting with a knife in X district yesterday; it was very frightening," the system will utilize semantic analysis to identify sensitive terms such as "knife," "fighting" and "frightening," categorizing the post as depicting a violent scenario. Subsequently, a human reviewer will confirm that the post constitutes a violation of content guidelines, and will remove it.

Similarly, these automated or manual reviews may pick up on language commonly used by bad actors such as groomers, such as slang terms that target minors or references to inappropriate relationships, even if the language is subtle or coded. Tencent notes that its AI-powered content safety review system, which combines the multimodal capabilities of large language models with human review, accurately detects harmful content and behaviors, such as sexual content, with a recall rate of over 80%. The system is designed to flag these patterns for further investigation, aiming for potentially harmful content to be identified promptly to protect vulnerable users. Continuous review and adaptation of these strategies is required to mitigate safety risks which are a challenge for tech companies.

Dimension 2. Creativity: child users engage curiously and use their imagination to build, invent and experiment.

Supplementing relationships through a family education-oriented AI model

The Minor Protection Centre seeks to encourage parent-child relationships, with what it Tencent defines as the "first family education AI model"

in China's tech industry. The model aims to be a digital education partner for parents that provides personalized solutions to educational challenges.

The AI model is integrated into Tencent's platform in the form of a chatbot. The chatbot is built on

Tencent's Hunyuan large language model, which parents may use to access various tools and applications on the Tencent platform, such as educational videos, case studies and volunteer services. Tencent notes that although the chatbot is continuously improving, its goal is to facilitate precise content dialogue and offer targeted tool recommendations, as well as to provide educational consulting services, tailored to the needs of users within the conversation context.

The model is trained on real cases handled by the Minor Protection Centre. The family education model attempts to deal with current educational challenges and predicting and future educational trends. It may answer questions pertaining to family education, school education and community education. According to Tencent's benchmarks and testing, the model achieves a 94.6% accuracy rate in problem solving. Aside from question-and-answer, the model also contains a library of multimedia educational resources such as videos on various family education topics.

The Minor Protection Centre's also uses interactive support methods, such as volunteer Q&A sessions and educational tools. For the volunteer Q&A, Tencent selects qualified and experienced volunteers to address users' inquiries, allowing users to find suitable answers from volunteers with diverse backgrounds. For specific parenting scenarios where asynchronous textual and verbal guidance may be inefficient, parental users can enhance their skills through educational tools that provide additional information. These tools utilize interactive videos to review common misconceptions in parenting and to provide positive practices, leading users to a task completion area to implement specific methods in practice.

Challenges and recommendations for ensuring child wellbeing and safety

The process of creating these case studies has identified several key challenges. Corresponding recommendations have been provided for the companies' reference, acknowledging that these issues are also common among other tech companies striving to ensure child well-being and safety. It is important to note that there are no one-size-fits-all solutions.

Challenge: Tensions between profitability pressures and compliance requirements. For instance, since gaming and social platforms are primarily sources of revenue and aim to retain users, they aim to offer engaging experiences. Nonetheless, it is essential to prevent excessive or unhealthy use and to protect users, particularly children. Balancing these competing tensions in a dynamic environment can be challenging.

Recommendations:

- Consider responsible engagement features, including adding time limits, age-appropriate recommendations and break reminders, to encourage healthy usage habits without significantly reducing user engagement.
- Adopt a dual-objective framework to align profitability with compliance, by embedding child protection and child wellbeing into product KPIs, ensuring that success is measured by both revenue and safety outcomes.
- Expand and formalize processes that track and evaluate risk mitigation approaches during the product ideation and development stages. Include post-deployment workflows to integrate findings from risk assessments into the Growth Guardian Platform on a regular basis.

Challenge: While real-name authentication and systems to curb excessive or unhealthy use – such as limits on time and spending – demonstrate a commitment to child online protection, the challenge lies in implementing these safeguards in a way that does not compromise user experience or digital inclusion. Striking the right balance requires not only effective enforcement but also addressing gaps in parental supervision. Moreover, age-assurance and authentication mechanisms have drawn criticism for potentially infringing on children's privacy and their right to participate meaningfully in digital environments.

Recommendations:

- Consider using alternative methods, such as self-declaration combined with verification checks (e.g., short questionnaires, tasks, or comprehension checks) to assess whether a user is likely to be a child that does not involve collection and processing of biometric information, to avoid privacy-invasive practices such as facial recognition.
- Establish cross-functional oversight teams combining compliance, product and ethics experts to continuously review platform practices and ensure that commercial objectives do not compromise child rights.
- Enhance provision of information to parents through alerts and guidance, to monitor and manage their children's gaming activity without being overly intrusive.

The RITEC-8 [framework] is very useful in thinking about the impact of children's well-being. It guarantees children's well-being from multiple dimensions, avoiding the limitations of a single dimension. In terms of safety, it prompts consideration of the physical and data security of products; in terms of content, it guarantees quality and suitability, and avoids bad information. This principle acts as a complete filter that systematically filters out factors that may endanger children's well-being.

– Tencent

The YBC App

The YBC App is a coding education platform created by YBC, to equip young people with skills for the age of artificial intelligence. The app connects children to instructors and offers customised curricula alongside various learning tools. YBC states that it designs its curricula based on the developmental needs of different age groups, and fosters a community in which children can share and comment on coding projects.

This case study explores how the YBC App attempts to address child wellbeing through two key dimensions of the RITEC-8 framework: safety and security, and relationships.³³ It describes the app's focus on providing a safe learning environment and promoting positive interactions among younger users.

YBC's experience illustrates the opportunities and challenges of building digital education platforms that are both engaging and protective of children's rights.

Disclaimer:

UNICEF does not endorse any company, brand, product or service.

This report is a UNICEF publication. Acknowledgements of company representatives do not imply a company's approval or endorsement of the report, and any reference made to a specific company does not imply endorsement by UNICEF of the company's policies and practices.

OVERVIEW

YBC has created the YBC App, a coding education platform designed for young people, focusing on equipping them for the age of artificial intelligence.

The YBC App aims to connect children to instructors, and provides them with customised curricula and other learning tools.³⁴ YBC reports that it assesses

the developmental characteristics of child users from different age groups when designing its curricula. The app also has a community element, in which child users can view and comment on each other's coding projects.

Educational platforms can facilitate education for child users, foster their creativity and teach them to communicate more efficiently. However,

33.This work draws on a separate project, Responsible Innovation in Technology for Children (RITEC), which is a collaboration between UNICEF and The LEGO Group, and is funded by The LEGO Foundation. It identifies eight principles (autonomy, competence, emotions, relationships, creativity, identities, safety and security and diversity, equity and inclusion) which can be used to assess child well-being and child rights. For more information see: <https://www.unicef.org/childrightsandbusiness/workstreams/responsible-technology/online-gaming/ritec-design-toolbox>

34.Digital education platforms may offer complementary support to children's learning, but they are not a substitute for formal education systems.

children may also face heightened risks through these platforms: there may be harmful content and behavioral risks associated with online interactions, competition/ gamification features may lead to mental health issues, or children may have their data processed without their informed consent. The availability of transactions may also expose children to contract risks, such as excessive spending.

Ensuring that the product is designed safely – following the principle of child safety by design – is crucial to maximizing its benefits while minimizing the risks. This case study dives into the safety mitigations that YBC has enacted to address these risks.

The case study discusses two dimensions adapted from international standards: safety and security and relationships. YBC states that it employs

structured safety assessments during the product ideation, development and deployment phases. Measures include secure data handling practices, robust privacy protections and user feedback mechanisms to address evolving risks, such as those from AI-generated content. Additionally, YBC indicates that the app’s community-oriented features promote collaboration, communication and relationship-building among young users, through interactive coding projects, live seminars and offline events.

YBC faces challenges in adapting to the dynamic nature of child safety risks, such as emerging AI threats. Areas for growth include expanding risk assessment processes to cover a broader spectrum of child rights, and engaging more extensively with external stakeholders such as children, academics and non-profits during product development.

CHILD WELLBEING DIMENSIONS IN ACTION

Dimension 1. Safety and security: children feel safe and are kept safe.

This case study highlights key elements of YBC’s processes for ensuring child safety and security:

Safety and security processes	
Risk assessments	Identifying child safety impacts, and then integrating, tracking and communicating risk mitigation measures.
Age assurance	Processes to estimate or verify the age of a user so that digital services can be tailored to their needs and age-appropriate safeguards put in place.
Complaint and redress	A means through which children can report any child rights violations and request redress.
Child-friendly information provision	Children are spoken-to in their own language and in a manner that makes them want to pay attention, such as through clear, simple and short sentences, providing information only when it is necessary to make a decision, and using interactive tools.
Child participation in company decision-making	Children are consulted and their views incorporated into the design and deployment of digital services.

YBC's safety and security risk assessments throughout the product lifecycle

Child safety by design should occur at key stages³⁵ in the product lifecycle:

1. During **product ideation**, when companies are conceptualising, researching and identifying the value proposition for a product, creating design plans and concepts for the product.
2. During the **product development and testing** phase, when companies are building and iterating on the product, creating prototypes and conducting tests prior to launching the product.
3. During the **sales, marketing and deployment** phase, when companies push the product to market, enable its deployment and monitor product use post-deployment for further iterations.

YBC reports that its child safety by design begins at the **product ideation** stage. Safety by design assessments cover multiple workflows, including the technical features of the platform, curriculum development, instructor hiring and safety pertaining to content on the platform. For instance, YBC indicates that it has created a structured assessment process for instructors who are permitted to teach on the platform, ensuring that cleared personnel are safe and skilled in teaching children.

YBC indicates that the YBC assessment process integrates numerous safety and security-oriented risk mitigation features and workflows. For instance, the app features a secure, real-time content review mechanism and user-flagging workflow, to protect users from exposure to harmful content. To ensure

the privacy of child users, YBC has adopted the data protection principles of necessity and data minimisation when designing product features involving data collection and processing. To prevent accidental payments by children while using the YBC App, it employs features such as password protection, phone confirmation and a parent mode to ensure that the payment is executed by the parent, not the child.

At the **product development and testing** phases, YBC reports that it establishes various workflows for safety and security mitigation. Internal teams supervise the content safety of their curricula: these teams include members of the Curriculum Content Research Institute, the Teaching Evaluation Research Institute, the Innovative Product Research Institute, the AI Class Center, and the Experience Class Center for content development. Security teams also carry out multiple rounds of security-oriented testing of the product, including adversarial tests, to ensure data protection. To ensure that privacy concerns are respected, YBC requires parents to review and agree to its privacy policy before their children can use the app.

At the **sales, marketing and deployment phase**, YBC notes that it relies on its user feedback mechanisms to continue to improve the safety and security of the YBC App. Feedback on the app may be submitted through phone calls, emails and other channels. YBC regularly summarizes user feedback and conducts both quantitative and qualitative analyses to provide a better service experience for users. To address emerging risks, such as those arising from AI, YBC has created post-deployment measures such as regular review and updating of its privacy policy.

35. This is a simplified approximation of the stages of the product development lifecycle for a tech company. Each company may structure the lifecycle stages differently.

Other procedural aspects of child safety by design

Digital safety by design for children is unlikely to be achieved by just one action, instead requiring several complementary and overlapping actions that, when combined, can help to ensure a safe and beneficial digital environment for children.

– Towards Digital Safety by Design for Children³⁶

Risk assessments are a core aspect of child safety by design, playing a key role in the identification of risks, mitigation of those risks, and tracking and communicating the effectiveness of mitigation. However, as the OECD argues above, companies must follow other processes to ensure that risks may be appropriately identified and addressed. This case study highlights four of these workflows:³⁷

- 1) Age assurance mechanisms
- 2) Child-friendly information provision
- 3) Complaint and redress mechanisms
- 4) Child participation/ consultation in company decision-making

1. YBC's age assurance mechanisms

Establishing age is important for digital safety by design, as services are only going to be able to protect children from unsafe content and behaviours if they know which of their users are actually children.

– Towards Digital Safety by Design for Children³⁸

YBC states that the YBC App is not intended to contain any content that would be unsuitable for users under 18. YBC's tutors verify the age, developmental stage and interests of students so that age-appropriate content can be provided to them. The YBC's lectures are divided into eight levels to cater to different age groups.

2. YBC's child-friendly information provision

Child-friendly information provision helps children to understand the terms of service associated with the various digital products and services they engage with, any risks associated with these products and services, and how to report harms and flag unsafe or illegal content.

– Towards Digital Safety by Design for Children³⁹

36.OECD (2024), *Towards digital safety by design for children* (OECD Digital Economy Papers), OECD Publishing, <https://doi.org/10.1787/c167b650-en>, p. 30.

37.Note that these workflows are categorized slightly differently from the OECD in this case study, for the sake of simplicity. The OECD's taxonomy of workflows also includes technical tools to determine harm and child-centered design; aspects of these are covered in the "substance" segment of this risk assessment. For a more comprehensive assessment, please refer to the OECD's publication.

38.OECD (2024), *Towards digital safety by design for children* (OECD Digital Economy Papers), OECD Publishing, <https://doi.org/10.1787/c167b650-en>, p. 30.

39.Ibid., p. 38.

According to YBC, the YBC App's safety and instructional language is phrased in simple and understandable terms for children. In addition, information throughout the app is presented with cartoon characters, bright colour schemes and cute graphics to keep children engaged.

3. YBC's complaint and redress mechanisms

Helping children understand how to report incidents and harms, and flag unsafe or illegal content, is essential. If complaints mechanisms are hard for children to find, access or use, they will have little utility [...] child-friendly complaint mechanisms are essential.

– Towards Digital Safety by Design for Children⁴⁰

YBC reports that both the YBC official website and the app feature a "report with one click" channel. User reports can include issues such as plagiarism, inappropriate visual content, unfriendly behaviour, infringement of rights, violation of laws and regulations, and other concerns. Users will be notified of the results of report processing. Additionally, users and their parents or guardians can report and send feedback directly to tutors.

4. YBC's child participation/ consultation in company decision-making

Children have a right to be heard in matters that affect them, and when children (and young people) play a part in helping to develop programs and policies that will affect them, it is more likely that they will be better aligned with children's needs, interests and backgrounds.

– Towards Digital Safety by Design for Children⁴¹

YBC indicates that children's views are incorporated into the YBC App's development cycle through the feedback channel. Information from this feedback is used to regularly update the YBC App's content offerings and mechanisms. For example, to ensure that each lecture is suitable for children of a certain age, YBC will adjust the content if it receives reports from children or parents that a lecture may not be age-appropriate.

Child safety and security risks from the YBC App

This section illustrates some child safety and security risks from the YBC App, and how the child safety by design processes described above have enabled YBC to mitigate those risks. These are intended to be illustrative, rather than exhaustive, examples of safety and security risks.

Addressing consumer risks from the YBC App

YBC states that the YBC App is free from paid-for games, product sales and advertisements. Real

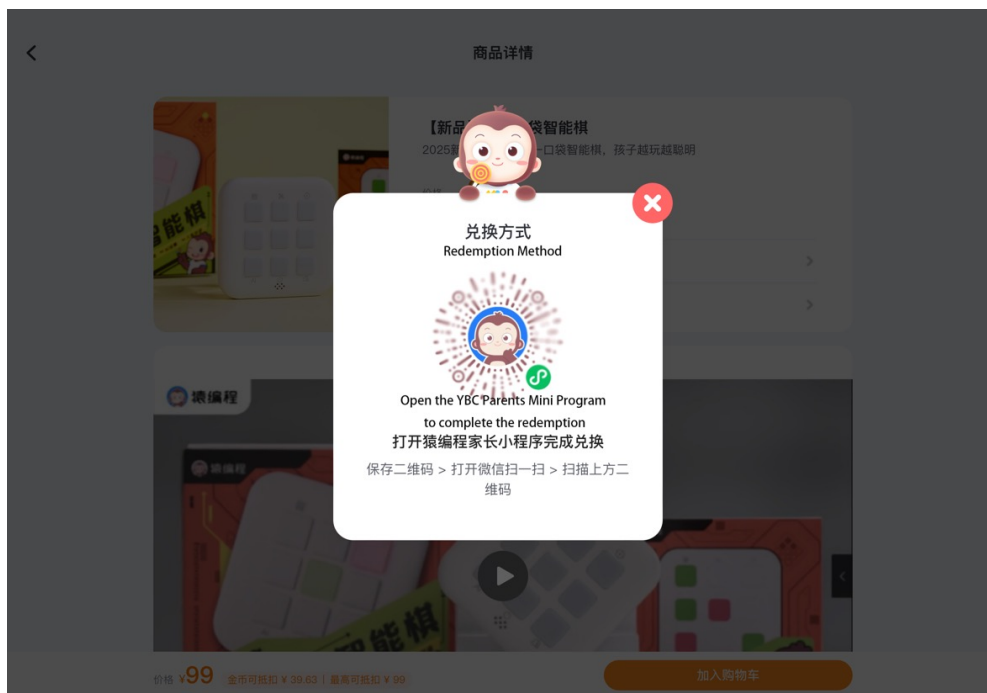
40.Ibid.

41.Ibid., p. 40.

money cannot be used anywhere in the YBC App. There is a companion app for parents to use on their phone, in which real money transactions can be made. Children can also earn points, such as by actively participating in class discussions and engaging in constructive activities, and these points can be exchanged in the app's store for a variety of items, such as stationery, toys and gifts, including backpacks, children's books and robots, which are particularly popular. The YBC App maintains a refund mechanism and provides customer support:

this significantly reduces consumer-based risks.

YBC indicates that the YBC App employs features such as password protection, phone confirmation and a parent app to prevent accidental payments made by children while using the app. These measures ensure that payments are executed by the parent, not the child. As illustrated in the figure below, redeeming earned points requires parent authorization, which can only be performed by parents.



© Courtesy of YBC

Redeeming earned points can only be carried out by parents in the parent's Mini Program.

Addressing contact risks arising from interactions on the YBC App platform

The YBC App has contact-related risks arising from teacher-student interaction through teleconferencing, a core element of the educational service. Additionally, the YBC community is an open communication platform where children

may communicate with each other to share and like each other's coding projects. However, like all online communities, there is a risk of malicious users, adults pretending to be children, and content and behavior-related abuse, including grooming or harassment and bullying.

YBC reports that it has implemented several

measures to mitigate these risks. First, all teachers on the YBC App undergo rigorous background checks, and any teacher with a history of harming minors or teenagers will be excluded. Second, teachers and tutors who come into contact with children are subject to ongoing assessments, including entry-level checks and routine inspections. Third, the

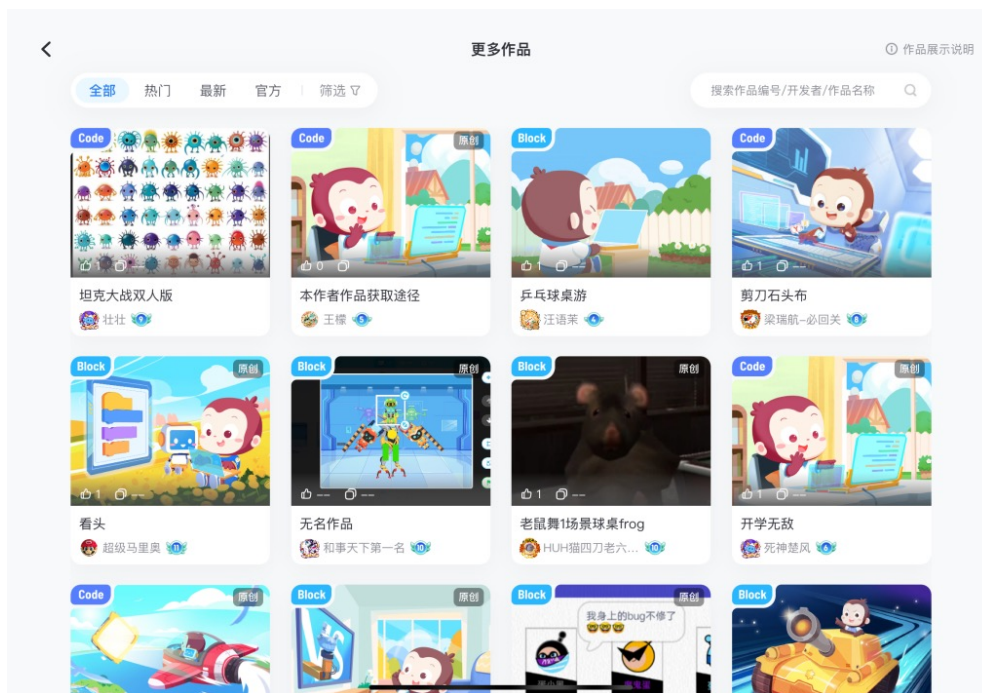
YBC App uses community management measures, such as real-name registration, to prevent adults from pretending to be children. To address the risk that bad actors are posing as child users, it uses keyword-filter technology and AI-powered image recognition to identify inappropriate behaviour or content and flag relevant accounts for enforcement.

Dimension 2. Creativity: child users engage curiously and use their imagination to build, invent and experiment.

Relationships: YBC App’s coding education community

YBC reports that the YBC App's programming platform provides children with a space to showcase

their creativity and hard work to a wide audience, including peers and parents. The platform has over 100,000 active student users, who have created more than 500,000 coding projects and written over 300 million lines of code.



© Courtesy of YBC
The YBC creativity community interface.

How Chinese Technology Companies Apply Principles of Responsible Innovation for Children and Child Safety by Design

YBC notes that Interaction forms the core of YBC's programming community. Students can like, collect, give points and share each other's

projects. Additionally, students can view or copy the source code of other students' work, with their consent, allowing them to learn from each other.



© Courtesy of YBC

The interactive interface of a coding project shared by a student.

YBC further indicates that the YBC App offers live coding lectures, in which students from across China participate in small, interactive classes. These sessions include class competitions that cultivate a sense of collective pride and camaraderie through

healthy rivalry. Through these experiences, students may connect with peers from diverse regions of China, reinforcing their connection to a larger learning community.



© Courtesy of YBC

A screenshot of comments from students during a lecture.

Challenges and recommendations for ensuring child wellbeing and safety

The process of creating these case studies has identified several key challenges. Corresponding recommendations have been provided for the companies' reference, acknowledging that these issues are also common among other tech companies striving to ensure child wellbeing and safety. It is important to note that there are no one-size-fits-all solutions.

Challenge: One of the key challenges in ensuring safety by design is the constantly shifting landscape of child safety risks.

Recommendations:

- Establish real-time monitoring and feedback mechanisms to track emerging threats, enabling the company to respond quickly and continuously improve safety measures while updating policies and product features to stay ahead of potential risks.
- Expand risk assessment workflows to cover more thematic areas. To support comprehensive risk mitigation in a dynamic context, the next step could involve institutionalizing processes that assess impacts across the full spectrum of child rights.
- Expand the scope of external stakeholders who are consulted throughout the product development lifecycle. In addition to the post-monitoring feedback channels it has established, YBC may also wish to directly engage with children, child safety experts, child safety academics or non-profits throughout product development, taking ethical considerations into account when consulting with children.

Challenge: Integrating safety findings often requires considerable resources, including time, expertise and financial investment. It can be challenging to appropriately allocate resources while still adhering to product development timelines.

Recommendations:

- Develop a risk-based framework to identify and address the most critical safety issues first, ensuring that time, expertise and financial resources are focused where they have the greatest impact.
- Integrate child safety considerations into every stage of the product development process, from design to testing, to prevent costly later fixes and ensure ongoing compliance with ethical and regulatory standards.

After carefully reading the RITEC framework, we found that many of the concepts advocated in RITEC-8 align closely with our company's existing principles. We will also further enhance our products and principles in accordance with the recommendations of RITEC-8. We plan to incorporate the RITEC-8 principles into the design of our future lectures and apps to ensure a safe, secure, and enriching digital environment for children.

– YBC

Walnut Programming 3D Graphical Editing Platform

Walnut Coding's 3D Graphical Editing Platform gives children a space to learn coding, make interactive projects and connect with others. The platform combines community features, such as forums, messaging, competitions and project-sharing.

This case study explores how Walnut Coding is working to address:

- **Safety & security:** through risk assessments, expert consultation, privacy by design and technical safeguards to protect children from online risks, harmful content and data misuse.
- **Diversity, equity & inclusion:** by providing diverse character tools, accessible features for children with disabilities, and affordable coding education for under-served communities.
- **Competency.**

Walnut Coding is striving to improve moderation, user feedback and risk management frameworks, aiming for a platform where all children can safely build skills, explore their creativity and interact with peers online.⁴²

Disclaimer:

UNICEF does not endorse any company, brand, product or service.

This report is a UNICEF publication. Acknowledgements of company representatives do not imply a company's approval or endorsement of the report, and any reference made to a specific company does not imply endorsement by UNICEF of the company's policies and practices.

OVERVIEW

Walnut Coding has created the Walnut Programming 3D Graphical Editing Platform, a 3D platform for children with community and educational features.⁴³

The platform features various mediums of engagement, including a public forum, direct messaging, a leaderboard and competition features, and a space for sharing coding projects. The platform also includes various creative tools that

42.This work draws on a separate project, Responsible Innovation in Technology for Children (RITEC), which is a collaboration between UNICEF and The LEGO Group and funded by The LEGO Foundation. It identifies eight principles (autonomy, competence, emotions, relationships, creativity, identities, safety and security and diversity, equity and inclusion) which can be used to assess child well-being and child rights. For more information see: <https://www.unicef.org/childrightsandbusiness/workstreams/responsible-technology/online-gaming/ritec-design-toolbox>

43.Digital education platforms may offer complementary support to children's learning, but they are not a substitute for formal education systems.

enable child users to develop and share interactive projects, such as creating stories, alongside access to coding tutorials and educational resources.

Educational platforms may facilitate education for child users, foster their creativity, and teach them to communicate more efficiently. However, children may also face heightened risks on such platforms: there may be harmful content and behavioural risks associated with online interactions, competition/gamification features may lead to mental health issues, or children may have their data processed without their informed consent. The availability of transactions may also expose them to contract risks, such as excessive spending.

Ensuring that the product is designed to be safe – following the principles of child safety by design – is crucial to maximizing its benefits while minimizing risks. This case study dives into the safety mitigations on the Walnut Programming 3D Graphical Editing Platform to address these risks.

The Walnut case study focuses on three dimensions of child wellbeing: safety and security, diversity, equity and inclusion, and competence.

Walnut Coding indicates that it carries out risk assessments of platform features such as messaging and leaderboards, consulting experts and stakeholders, and addressing identified risks through modifications such as enhanced content moderation and privacy by design principles. Technical safeguards, such as stricter filtering mechanisms and improved reporting systems, are implemented to prevent issues such as cyberbullying, online violence and data misuse. Walnut Coding attempts to integrate diversity, equity and inclusion through diverse character creation tools, accessible design for children with disabilities, adaptive coding courses and affordable programming education initiatives for under-served regions.

Balancing strict safety controls with user-friendly experiences remains an ongoing challenge for Walnut Coding. As digital risks evolve, Walnut Coding must remain agile, continuously updating its frameworks and methodologies. Deepening its stakeholder engagement by consulting a broader range of child rights experts and non-profit organizations will be important, in addition to employing more forward-looking risk assessment techniques to pre-empt future challenges.

CHILD WELLBEING DIMENSIONS IN ACTION

Dimension 1. Safety and security: children feel safe and are kept safe.

This case study highlights key elements of Walnut Coding's processes for ensuring child safety and security:

Safety and security processes

Risk assessments

Identifying child safety impacts, and then integrating, tracking and communicating risk mitigation measures.

Safety and security processes	
Age assurance	Processes to estimate or verify the age of a user so that digital services can be tailored to their needs and age-appropriate safeguards put in place.
Complaint and redress	A means through which children can report any child rights violations and request redress.
Child-friendly information provision	Children are spoken-to in their own language and in a manner that makes them want to pay attention, such as through clear, simple and short sentences, providing information only when it is necessary to make a decision, and using interactive tools.
Child participation in company decision-making	Children are consulted and their views incorporated into the design and deployment of digital services.

Walnut Coding’s safety and security risk assessments throughout the product lifecycle

Child safety by design should occur at key stages⁴⁴ in the product lifecycle:

1. During **product ideation**, when companies are conceptualising, researching and identifying the value proposition for a product, creating design plans and concepts for the product.
2. During the **product development and testing** phase, when companies are building and iterating on the product, creating prototypes and conducting tests prior to launching the product.
3. During the **sales, marketing and deployment** phase, when companies push the product to market, enable its deployment, and monitor product use post-deployment for further iterations.

Walnut Coding reports that its child safety by design begins with risk management at the

product ideation stage, handled by the company’s product development and legal and compliance teams. The company analyses actual and potential risks to child rights through impact assessments of the platform’s features, such as messaging functions, content sharing and leaderboards. These assessments involve reviewing existing research on child rights in digital environments and consulting with child rights experts parents, and educators to understand specific risks. New or emerging concerns raised during design or beta testing are escalated to a cross-functional review team. This team categorizes risks by their severity and urgency, ensuring timely mitigation.

Walnut Coding reports that it integrates assessment findings by adjusting certain product features. For example, the company indicates that it has refined its content moderation workflows to enable automated review and flagging of content that may violate platform standards. It also states that it applies privacy by design principles aimed at minimizing and securing the collection of personal

44.This is a simplified approximation of the stages of the product development lifecycle for a tech company. Each company may structure the lifecycle stages differently.

data. Platform policies are reportedly updated to help address risks such as cyberbullying or unauthorized sharing of personal information. Technical features, including moderation protocols, are incorporated into messaging functions to help reduce the likelihood of inappropriate contact between users. These measures are subject to periodic review as part of the company's broader risk mitigation efforts.

Walnut Coding states that child safety by design continues through the **product development and testing** phase. It carries out automated security assessments that stress-test the system's algorithms for potential security vulnerabilities. If a risk is identified, such as around potential data leaks or inappropriate content sharing, Walnut Coding seeks to modify product features (such

as content moderation tools) to include stricter filtering mechanisms. Additionally, Walnut Coding may enhance user reporting systems to allow children and parents to easily flag any unsafe behavior or content.

Walnut Coding reports that it takes quantitative and qualitative measurements of the effectiveness of risk mitigation measures. For instance, the company analyzes quantitative data, such as the number of flagged content items, the speed and accuracy of content moderation, and the frequency of user reports related to safety concerns. These metrics are tested across different testing cycles; for example, a decline in reports of inappropriate content, or an increase in the speed of flagging harmful content, signals that its measures are effective.

Child safety by design: child and parent stakeholder engagement

Walnut Coding indicates that it carries out user testing and prototype evaluations with a focus on child safety. The company invites a group of children, parents and educators to interact with early versions of the platform, observing their usage patterns to identify any risks related to privacy, data security and exposure to inappropriate content.

Qualitative feedback from children, parents and educators who participate in testing phases is also collected. This helps the company to understand user perceptions of the safety features and whether they feel secure when using the platform. Walnut Coding also reviews user satisfaction surveys and organizes focus group discussions to gauge the success of mitigation measures from a user's perspective.

During the sales, marketing and deployment phase, Walnut Coding reports that it monitors the platform for risks such as inappropriate interactions in messaging, unauthorized content sharing or potential privacy breaches. It may then tighten content moderation filters or introduce additional reporting mechanisms for unsafe

behavior. Additionally, Walnut Coding updates its user guidelines and policies based on feedback and detected risks, ensuring that user behavior aligns with child safety principles. For instance, these may include stricter rules for in-app communication, or content uploads can be introduced as part of the mitigation strategy.

Post-deployment, Walnut Coding notes that it discloses product impacts related to child rights through multiple channels, to ensure transparency and engagement with relevant stakeholders. Walnut Coding is planning to publish annual reports to provide detailed insights into its product development processes, specifically focusing on how child rights are considered and safeguarded during testing and prototype creation. These reports include detailed sections on child protection policies, the effectiveness of the content moderation system and improvements based on user feedback. Walnut Coding also releases press releases on its website to inform the public and stakeholders about its child safety protocols.

Other procedural aspects of child safety by design

Digital safety by design for children is unlikely to be achieved by just one action, instead requiring several complementary and overlapping actions that, when combined, can help to ensure a safe and beneficial digital environment for children.

– Towards Digital Safety by Design for Children⁴⁵

Risk assessments are a core aspect of child safety by design, playing a key role in the identification of risks, mitigation of those risks, and tracking and communicating the effectiveness of mitigation. However, as the OECD argues above, companies

must follow other processes to ensure that risks may be appropriately identified and addressed. This case study highlights four of these workflows:⁴⁶

- 1) Age assurance mechanisms
- 2) Child-friendly information provision
- 3) Complaint and redress mechanisms
- 4) Child participation/ consultation in company decision-making

1. Walnut Coding's age assurance mechanisms

Establishing age is important for digital safety by design, as services are only going to be able to protect children from unsafe content and behaviours if they know which of their users are actually children.

– Towards Digital Safety by Design for Children⁴⁷

Walnut Coding indicates that it includes a multi-step age assurance process. During the account creation process, users are required to input their birthdate, with parental consent required for children. The data is used solely to determine the user's age. It is stored in encrypted databases and accessible only to authorised personnel. Adult use of the platform is restricted to verified tutors, and unverified adults cannot message tutors. A balance must be maintained, as age-assurance and authentication measures have raised concerns about potential impacts on children's privacy and their right to participate in digital environments.

45.OECD (2024), *Towards digital safety by design for children* (OECD Digital Economy Papers), OECD Publishing, <https://doi.org/10.1787/c167b650-en>, p. 30.

46.Note that these workflows are categorized slightly differently from the OECD in this case study for the sake of simplicity. The OECD's taxonomy of workflows also includes technical tools to determine harm and child-centered design; aspects of these are covered in the "substance" segment of this risk assessment.

47.OECD (2024), *Towards digital safety by design for children* (OECD Digital Economy Papers), OECD Publishing, <https://doi.org/10.1787/c167b650-en>, p. 30.

Walnut Coding further indicates that it uses automation to detect unusual content, triggering a verification process. For example, if a child user tries to access content flagged as age-inappropriate, Walnut Coding's system prompts the user for further verification or blocks access based on the initial age input. Further verification includes external services that rely on document verification or facial recognition (with parental consent).

2. Walnut Coding's child-friendly information provision

Child-friendly information provision helps children to understand the terms of service associated with the various digital products and services they engage with, any risks associated with these products and services, and how to report harms and flag unsafe or illegal content.

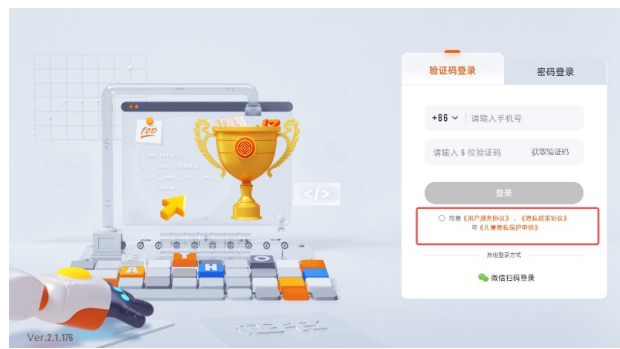
– Towards Digital Safety by Design for Children⁴⁸

According to Walnut Coding it uses age-appropriate language throughout the platform, simplifying complex concepts into short, clear sentences. For key features such as privacy settings, safety guidelines and coding instructions, Walnut Coding aims to present relevant information when necessary for decision-making. This model of information provision intends to avoid overwhelming users with too much content, and provides guidance when it is relevant. For instance, Walnut Coding's privacy instructions are:

To keep your account safe, Walnut Coding only shares your username and profile picture with others. Users' private information, like your real name and phone

number, is kept secure and will not be shared unless you or your parents allow it. You can change your privacy settings anytime by going to your account settings.

To complete account creation, you need to provide the following information: mobile phone number. If you do not provide the above personal information, we will be unable to create your account, complete your account profile, or generate a User ID for you.



© Courtesy of Walnut Coding

Login interface of Walnut Coding with highlighted access points to the User Service Agreement, Privacy Policy, and Children's Privacy Protection Statement.

In addition, Walnut Coding indicates that the 3D graphical programming platform includes features designed to help children navigate the platform, understand programming concepts, and ensure online safety. These include navigation guides, which are step-by-step tutorials on how to use different features, such as creating a project, saving progress and interacting with the platform community.

Walnut Coding's User Experience team works with educational experts and child psychologists to attempt to refine these communication strategies.

48.Ibid., p. 38.

3. Walnut Coding’s complaint and redress mechanisms

Helping children understand how to report incidents and harms, and flag unsafe or illegal content is essential. If complaints mechanisms are hard for children to find, access or use, they will have little utility [...] child-friendly complaint mechanisms are essential.

– Towards Digital Safety by Design for Children⁴⁹

Walnut Coding reports that it provides a complaints submission form that is available in the user interface, allowing children or their parents to report concerns. Parents may also lodge complaints on behalf of their children. Each submission is tracked through Walnut Coding’s incident management system (see image below), which prioritizes cases involving child safety or content violations.

用户ID	问题描述/问题/问题/问题	优先级	用户/责任人/责任人/责任人	处理状态	人工/自动/系统/系统/系统	创建时间
10000000000000000000	问题描述/问题/问题/问题	000000	1/2/3/4	5级/6级	已解决/已解决	2005-01-18 10:01:37:415
10000000000000000000	问题描述/问题/问题/问题	000000	14/2/1/0	5级/6级	已解决/已解决	2005-01-18 08:49:05:875
10000000000000000000	问题描述/问题/问题/问题	000000	11/2/1/9	5级/6级	已解决/已解决	2005-01-17 21:53:30:7
10000000000000000000	问题描述/问题/问题/问题	000000	1/2/1/3	5级/6级	已解决/已解决	2005-01-17 21:53:31:229
10000000000000000000	问题描述/问题/问题/问题	000000	2/2/1/3	5级/6级	已解决/已解决	2005-01-17 21:54:43:605
10000000000000000000	问题描述/问题/问题/问题	000000	10/2/1/11	5级/6级	已解决/已解决	2005-01-17 19:54:07:018
10000000000000000000	问题描述/问题/问题/问题	000000	10/2/1/5	5级/6级	已解决/已解决	2005-01-17 17:50:38:344

© Courtesy of Walnut Coding
Backend interface of Walnut Coding’s incident management system.

A support team reviews each case and escalates it to relevant departments, such as the Legal and Compliance team. Children or their parents/guardians may also communicate directly with support staff regarding urgent issues.

Walnut Coding further indicates that it provides other redress mechanisms, such as to block offending users, remove inappropriate content and provide feedback on the resolution of issues. In severe cases, such as cyberbullying or breaches of privacy, the platform enables users to request a full investigation, and the platform collaborates with authorities and child protection organizations if necessary.

4. Walnut Coding’s child participation/consultation in decision-making

Children have a right to be heard in matters that affect them, and when children (and young people) play a part in helping to develop programs and policies that will affect them, it is more likely that they will be better aligned with children’s needs, interests and backgrounds.

– Towards Digital Safety by Design for Children⁵⁰

Walnut Coding reports having several child engagement workflows. The company organizes focus groups and user testing sessions that include children from different age groups. It has also established a child advisory board composed of young participants who meet periodically with product development teams. This board provides a channel for children to share input on features, safety measures and educational content. Walnut Coding collaborates with educators and child psychologists to develop consultation methods that are age-appropriate and accessible. The company has indicated that it has plans to publish feedback reports in the future, to show how these engagements have informed platform updates.

49.Ibid.

50.Ibid., p. 40.

Child safety and security risks associated with Walnut Programming's 3D Graphical Editing Platform

This section illustrates some child safety and security risks associated with Walnut Programming's 3D Graphical Editing Platform and how the child safety by design processes described above have enabled Walnut Coding to mitigate those risks. These are intended to be illustrative, rather than exhaustive, examples of safety and security risks.

Addressing consumer-based risks on the 3D Graphical Editing Platform

Children may use their parents' accounts or make independent purchases of products within the app. Without controls or risk mitigation measures this marketplace could pose financial risks to child users, such as unintended purchases or overspending through micro-transactions. Without proper safeguards, children might unknowingly subscribe to these services or may not understand the concept of recurring payments, resulting in regular charges to their parents' credit cards.

To mitigate these risks, Walnut Coding states that it requires parents to explicitly approve any purchases in the in-app marketplace; children may not directly transact. Instead, the course homeroom teachers at Walnut Coding communicate directly with the children's parents, seeking parental consent. After testing the children's programming skills, they recommend programming courses of varying levels.

In addition, Walnut Coding indicates that its staff provide users with detailed explanations of the content and services that can be accessed after payment. The platform uses secure third-party payment gateways to protect financial information and prevent unauthorized access. Walnut Coding will promptly send information to registered users

regarding any financial activity or subscriptions related to their child's account. Regular audits and monitoring are carried out, to identify and address potential issues related to consumer transactions and financial management.

Identifying and mitigating privacy risks

The 3D Graphical Programming Platform collects children's personal information, such as their names, ages and email addresses, to create accounts and deliver personalized experiences. For instance, the age assurance processes described above collect data about child users' birthdays. Without appropriate controls or risk mitigation measures, this data collection could expose sensitive information to unauthorized access. For example, inadequate security measures may allow malicious actors to obtain personal information, leading to identity theft, unauthorized contact or other privacy violations. Similarly, the platform may track children's location data to enable location-based services or content recommendations. Without robust privacy safeguards this practice could result in surveillance risks, allowing children's actions and locations to be monitored without their knowledge or consent. Such breaches could potentially expose children to harmful behavior, including tracking by malicious individuals.

To mitigate these privacy risks, according to Walnut Coding the platform implements end-to-end encryption to secure all data exchanges and communications. Personal information, such as names, ages and email addresses, is collected only with explicit parental consent, and stored in secure, encrypted databases. Walnut Coding enforces strict access controls to ensure that only authorized personnel can handle sensitive data. Additionally, the platform provides a privacy policy to inform both parents and children about how their data

is collected, used and protected. Location data collection is minimized and anonymized where possible, and users are given control over enabling or disabling location-based services.

Walnut Coding further indicates that the platform takes steps to secure communication data, such as from chats or messages, through advanced encryption, protecting it from unauthorized access or interception. Parental controls allow guardians to review and manage their child's privacy settings, ensuring that they have oversight

of what information is shared and with whom. To further enhance privacy protection, the platform carries out regular security audits and vulnerability assessments to identify and address potential risks, providing a safer and more transparent environment for child users.

However, as mentioned above, it is important to strike a balance, as age-assurance and authentication measures have prompted concerns about their potential impact on children's privacy and their ability to participate in digital environments.

Dimension 2. Diversity, equity, and inclusion: diverse children and childhoods are represented, and the access needs of as many different children as possible are served.

Walnut Coding's cultural, ethnic and economic diversity

Walnut Coding reports that it carries out market research to ensure that its cartoon character designs, both original and licensed, sufficiently appeal to children. Walnut Coding plans to introduce a customization feature that allows children to design their own favourite animated characters, subject to content moderation processes.

Walnut Coding notes that it tries to offer equity-focused coding courses designed to be accessible for all children, regardless of their background or prior experience. The courses include adaptive learning paths and support resources. User feedback on the coding platform seems to suggest that this feature enhances educational equity and inclusive participation. However, more attention should be placed on children with disabilities.

Furthermore, according to Walnut Coding, its platform attempts to consider educational gaps caused by regional economic disparities, by



© Courtesy of Walnut Coding
Characters from Walnut Coding's storyline and level design.

pricing the product at approximately 30% of the market rate, providing children in China's third- to sixth-tier cities with opportunities to access programming education. Sales data over the years shows an increasing number of customers from these lower-tier cities. Additionally, Walnut Coding has established a strategic partnership with Beautiful China to donate programming products to schools in remote mountainous areas and provide

accompanying educational and technical support (see images below).

Walnut Coding reports that this initiative currently covers over 20 schools in the provinces of Guangdong, Fujian and Yunnan, allowing Walnut Coding's products to serve more children and provide programming education to those in economically underdeveloped regions.



© Courtesy of Walnut Coding

Walnut Coding's donated learning materials and onsite teaching session in a remote mountain school.

Dimension 3. Competence: child users have increased perceptions of their effectiveness, ability and skills, and experience a sense of mastery.

Core features that enhance competence for Walnut Coding's users

According to the company, Walnut Programming's 3D Graphical Programming Platform is intended to support children's coding development through interactive, hands-on activities. Its drag-and-drop interface simplifies programming concepts into manageable components, allowing children to assemble code blocks to create projects. This incremental approach is designed to help build foundational knowledge and confidence, reducing the potential for children to feel overwhelmed

by traditional programming languages. Guided tutorials offer structured learning pathways, beginning with basic tasks and gradually increasing in complexity to support progressive skill development.

Walnut Coding reports that collaboration features are part of the platform's approach to supporting skill development. Through joint coding projects, children may have opportunities to practice technical skills while also engaging in teamwork, communication and basic project coordination. This collaborative setup is intended to encourage idea-

sharing, task distribution and collective problem-solving, which can contribute to the development of interpersonal skills relevant to coding and other contexts.

The platform is reported to include tools to track and reinforce progress, further boosting children's sense of competence. A progress tracking dashboard provides visual feedback on completed tasks, challenges and achievements, helping children to recognize their growth and identify areas for improvement. Project-based learning activities may encourage children to apply coding concepts to real-world projects, such as games or animations, where they can experiment, refine their skills, and gain a deeper understanding of programming. Together, these features may contribute to creating a supportive environment where children can steadily build their technical and creative competence in coding.

Challenges and recommendations for ensuring child wellbeing and safety

The process of creating these case studies has identified several key challenges. Corresponding recommendations have been provided for the companies' reference, acknowledging that these issues are also common among other tech companies striving to ensure child wellbeing and safety. It is important to note that there are no one-size-fits-all solutions.

Challenge: Balancing the implementation of robust safety measures with maintaining a user-friendly experience can be challenging. While features such as content moderation and parental controls are crucial, it is important to strike a balance that does not hinder children's ability to engage meaningfully with the platform.

Recommendations:

- Implement content moderation and parental controls that dynamically adjust based on children's age, usage patterns and behavior, ensuring robust protection without overly limiting engagement or exploration.
- Actively involve children and parents in testing, feedback and co-design processes, to ensure that safety measures are user-friendly and non-intrusive, and that they support meaningful interaction with the platform. When consulting with children, ethical considerations need to be in place.
- Enhance the accessibility of products for children with disabilities. This might include simplified interactions through larger buttons and compatibility with assistive technologies such as screen readers.
- Diversify the range of external stakeholders who are consulted, including child rights experts or non-profit groups as part of ongoing child safety by design implementation, taking ethical measures into consideration.

Challenge: Systematically integrate risk assessment seamlessly into the product development cycle, without slowing down innovation, is a complex task.

Recommendations:

- Regularly monitor emerging digital risks and trends, updating safety protocols accordingly, and providing educational resources to children and parents to promote responsible, safe and informed use of the platform.
- Balance the use of non-invasive, AI-powdered age assurance mechanisms against privacy considerations in line with international standards.

- Adopt technical measures to mitigate a wider range of child safety risks, including detecting harmful behavior in messages and comments, detecting fraudulent transactions on the platform, and detecting bias or discrimination in educational materials. The company should carry out child rights risk assessments on its products.

REFERENCES

1. Organisation for Economic Co-operation and Development. *OECD roundtable on digital safety by design for children*. OECD; 2023. Available from: 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. *Child rights by design: Guidance for innovators of digital products and services used by children*. Digital Futures Commission; 2023. Available from: https://digitalfuturescommission.org.uk/wp-content/uploads/2023/03/CRbD_report-FINAL-Online.pdf.
3. United Nations. *Guiding principles on business and human rights: Implementing the United Nations "Protect, Respect and Remedy" framework*. United Nations Human Rights Office of the High Commissioner; 2011. Available from: https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf.
4. United Nations Children's Fund (UNICEF), Office of Research – Innocenti. *Responsible innovation in technology for children (RITEC): Digital technology, play and child well-being (Phase Two Report)*. UNICEF; 2024. Available from: <https://www.unicef.org/innocenti/reports/responsible-innovation-technology-children>.
5. United Nations. *Convention on the Rights of the Child*. United Nations General Assembly, Resolution 44/25; 1989. Available from: <https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child>.
6. Committee on the Rights of the Child. *General comment No. 25 (2021) on children's rights in relation to the digital environment (CRC/C/GC/25)*. United Nations; 2021. Available from: <https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation>.
7. Organisation for Economic Co-operation and Development. *Children in the digital environment: Revised typology of risks* (OECD Digital Economy Papers No. 302). OECD Publishing; 2021. <https://doi.org/10.1787/9b8f222e-en>.
8. Livingstone S., Stoilova M. *The 4Cs: Classifying online risk to children* (CO:RE Short Report Series on Key Topics). Hamburg: Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI); CO:RE – Children Online: Research and Evidence; 2021. <https://doi.org/10.21241/ssoar.71817>.
9. Organisation for Economic Co-operation and Development. *Towards digital safety by design for children* (OECD Digital Economy Papers). OECD Publishing; 2024. <https://doi.org/10.1787/c167b650-en>.



携手为儿童

Contact

UNICEF China

12 Sanlitun Lu, Chaoyang District
100600 Beijing, People's Republic of China
Tel: (8610) 85312600
Fax: (8610) 65323107
Email: beijing@unicef.org

Copyright © UNICEF China, January 2026

Copyright © China Federation of Internet Societies (CFIS), January 2026

Copyright © Communication University of China, January 2026