Trusted Tech
Africa Ltd

karot

Harm Modelling Report For Karot

Introduction

Karot is an innovative, AI-driven platform designed to equip young people in Africa with essential soft skills and financial literacy. Through practical education, interactive tools, and parental oversight, Karot empowers the next generation with the knowledge and confidence to build strong life skills and secure a successful future.

To enhance user safety and mitigate risks associated with digital learning environments, Trusted Tech partnered with the Karot team to conduct a harm modelling assessment. This evaluation aims to identify potential risks, address vulnerabilities in the product, and implement proactive safety measures.

Platform Overview

The Karot platform is accessible as both a web and mobile application, integrating AI functionality to deliver personalized learning experiences. Key features include:

- **Educational Interface:** Over 80 expert-led video lessons on soft skills and financial literacy.
- **AI-Driven Learning:** An interactive chatbot that simulates real-life scenarios, helping users apply skills in a controlled environment.
- **Financial Literacy Tools:** A built-in e-wallet allowing users to earn, save, and spend under parental supervision.
- **Parental Integration:** Dual-app system enabling parents to monitor progress and guide learning.
- **AI-Powered Tutors:** Automated video lessons to enhance personalized education experiences.

Trusted Tech
Africa Ltd

karot

Harm Description Statement for Karot

1. **Intended Use:**

- If Karot's AI-driven personalized recommendations are used to suggest financial literacy content, young users might misinterpret complex financial concepts without sufficient contextual guidance, leading to potential financial misunderstandings.

2. **Unintended Use:**

- If parents utilize the parental monitoring feature to excessively restrict peer-to-peer learning interactions, young users could face limited collaborative opportunities, reducing their exposure to diverse perspectives and problem-solving experiences.

3. **System Errors:**

- If the AI-driven learning system malfunctions, it could generate misaligned content recommendations, resulting in user frustration and gaps in skill development.
- If the parental monitoring system provides inaccurate insights regarding child activity and spending, parents may develop a false sense of security, leaving children vulnerable to inappropriate content or financial mismanagement.

4. **Misuse:**

- Malicious actors could exploit the peer-to-peer learning feature to spread misinformation or cause emotional distress, negatively impacting the mental well-being and knowledge base of young users.

Trusted Tech
Africa Ltd

karot

Defining the Product Purpose

The Problem Karot is Solving

Karot addresses a critical gap in Africa's educational and economic landscape: the lack of structured education in soft skills and financial literacy. These essential competencies are often absent from traditional curricula, leaving young people ill-prepared for adulthood.

Many schools and families lack the necessary resources and expertise to teach these life skills effectively. As a result, young individuals often enter adulthood without fundamental knowledge of money management, communication, teamwork, and career planning. This deficiency contributes to financial instability, limited career opportunities, and broader economic challenges.

New Capabilities Offered by Karot

- **AI-Driven Personalized Learning:** An intelligent chatbot that guides users through financial literacy and soft skills development via interactive conversations and real-life scenario simulations.
- **Comprehensive Financial Literacy Training:** A built-in e-wallet where children learn responsible money management through chore-based earning, saving, and controlled spending.

Karot integrates cutting-edge AI technology with an engaging educational platform, offering:

- **Parental Involvement:** A dedicated parent app providing oversight, progress tracking, and financial activity monitoring.
- **Secure Digital Learning Environment:** Age-appropriate content with robust moderation, reporting mechanisms, and user controls.

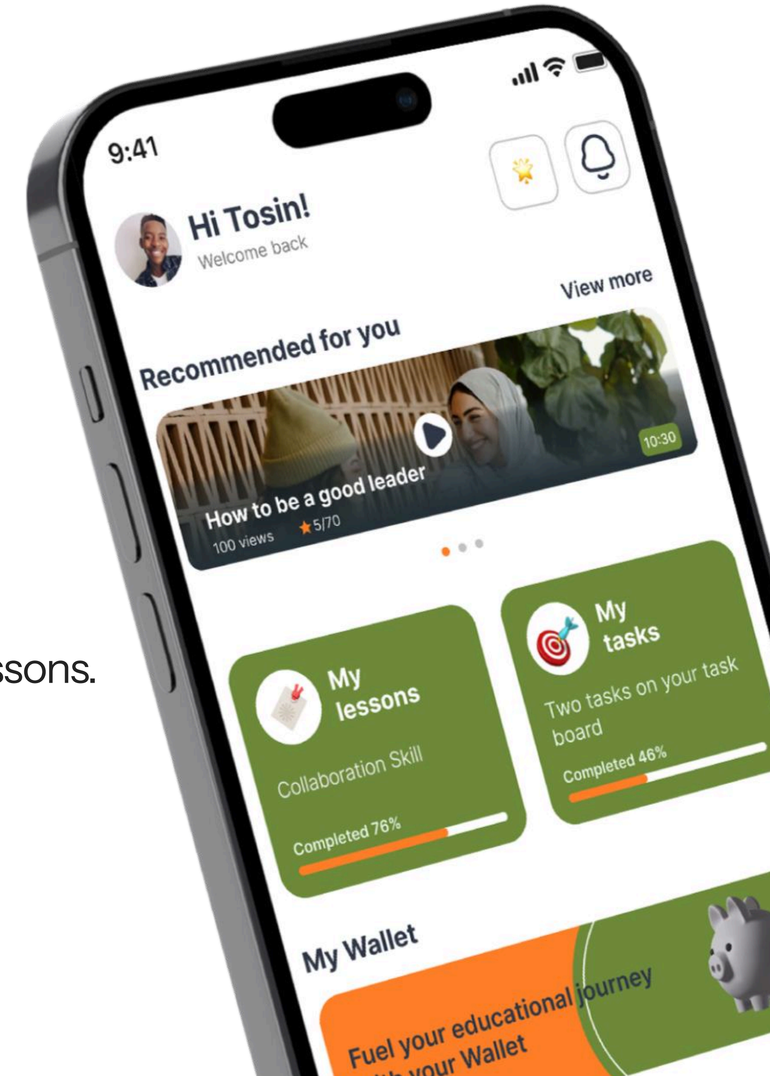
Use Cases and Impact Assessment

Learning & Skill Development for Young People (Primary Users)

- **User:** Students aged 9-18
- **Scenario:** A 15-year-old from an underserved community seeks to enhance financial literacy and soft skills but lacks access to traditional training.

How Karot Helps:

- The student registers on Karot via a smartphone.
- Engages with AI-driven chatbots for interactive financial literacy and entrepreneurship lessons.
- Accesses expert-led video lessons and completes topic-based quizzes.
- Uses the Karot e-wallet to apply financial management concepts in a controlled setting.



Trusted Tech Africa Ltd

karot

Individual Impact on Young People

- **Increased Digital & Financial Literacy:** Users develop responsible money management skills through hands-on budgeting tools and savings challenges.

- **Career Readiness & Economic Empowerment:**
  - Learners gain entrepreneurship knowledge and job preparation guidance.
  - Increased confidence in starting small businesses or securing employment opportunities.

Assessing and Mitigating Harm in Karot

Trusted Tech, in collaboration with Karot, conducted a harm modelling exercise using frameworks such as the World Economic Forum's Global Coalition for Digital Safety Toolkit and Microsoft's Responsible Innovation Best Practices Toolkit. This analysis evaluated potential risks and proposed interventions, focusing on:

1. **Online Forum Safety Measures:**
   - Reporting Mechanisms: Implementing tools for flagging inappropriate content and user behavior.
   - Content Moderation: AI-assisted filtering and human moderation for community safety.
   - User Controls: Adjustable privacy settings allowing users to manage their interactions.

2. **Age Verification for Child Sections:**
   - Know Your Customer (KYC) Measures: Preventing adults from masquerading as children to access the online forum.
   - AI-Based Identity Verification: Ensuring appropriate user segmentation and secure interactions.

Trusted Tech Africa Ltd

karot

CATEGORY	TYPE OF HARM	Contributing Factors	Severity	Scale	Probability	Frequency	Potential
Risk of injury	Physical or infrastructure damage		Low	Medium	Low	Low	Low
	Emotional or psychological distress		Low	Medium	Medium	Low	Low
Denial of consequential services	Opportunity loss		Low	Low	Low	Low	Low
	Economic loss		Low	Low	Low	Low	Low
Infringement on human rights	Dignity loss		Low	Low	Low	Low	Low
	Liberty loss		Low	Low	Low	Low	Low
	Privacy loss		Medium	Low	Low	Low	Low
	Environmental impact		Low	Low	Low	Low	Low
Erosion of social & democratic structures	Manipulation		Medium	Medium	Medium	Low	Low
	Social detriment		Low	Low	Low	Low	Low

Trusted Tech Africa Ltd

tarot

Category	Type of Harm	Severity	Scale	Probability	Frequency	Potential
Content Risks	Child Sexual Abuse Material (CSAM)	Low	Low	Low	Low	Low
	Child Sexual Exploitation Material (CSEM)	Low	Low	Low	Low	Low
	Pro-terror Material	Low	Low	Low	Low	Low
	Content Supporting Extremist Organizations	Low	Low	Low	Low	Low
	Violent Graphic Content	Low	Low	Low	Low	Low
	Content Inciting Violence	Low	Low	Low	Low	Low
	Content Promoting Dangerous Physical Behavior	Low	Low	Low	Low	Low
	Material Promoting Suicide, Self-harm, Disordered Eating	Low	Low	Low	Low	Low
	Developmentally Inappropriate Content	Low	Low	Low	Low	Low
	Hate Speech	Medium	Medium	Medium	Low	Medium
	Disinformation and Misinformation	Medium	Medium	Medium	Low	Medium
	Deceptive Synthetic Media (Deepfakes)	Low	Low	Low	Low	Low
	Content Inciting Discrimination	Low	Low	Low	Low	Low

Contact Risks	Grooming for Sexual Abuse	Medium	Medium	Medium	Low	Medium
	Recruitment and Radicalization	Medium	Medium	Medium	Low	Medium
	Sexual Extortion (Sextortion)	Medium	Medium	Medium	Low	Medium
Conduct Risks	Technology-facilitated Abuse (TFA)					
	Technology-facilitated Gender-based Violence	Medium	Medium	Medium	Low	Medium
	Online Bullying and Harassment	Medium	Medium	High	Low	High
	Doxxing					
	Image-based Abuse	Low	Low	Low	Low	Low
	Impersonation	Medium	Medium	Low	Low	Medium
	Scams	Low	Low	Medium	Low	Medium
	Phishing	Low	Low	Medium	Low	Medium
	Catfishing	Low	Low	Medium	Low	Medium
Content/Contact/Conduct Risks	Child Sexual Exploitation and Abuse (CSEA)	Low	Low	Medium	Low	Low
	Manipulation of Public Opinion (Political/Ideological)	Low	Low	Low	Low	Low
	Algorithmic Discrimination	Low	Low	Low	Low	Low

Recommendations and Next Steps

1. **Developing Robust Policies:**
   - Strengthen Terms of Service and Privacy Policy to ensure clarity on data protection, content moderation, and user responsibilities.
   - Establish clear Parental Consent Protocols for monitoring features.
2. **Enhancing AI Safety Measures:**
   - Continuously refine AI-driven learning to minimize errors in personalized recommendations.
   - Improve AI moderation capabilities to detect and prevent harmful peer interactions.
3. **Strengthening Reporting and Moderation Systems:**
   - Introduce a dedicated Trust & Safety Team for monitoring harmful behavior.
   - Expand Automated Moderation Tools to detect misinformation and harmful content in peer-to-peer interactions.
4. **User Education and Digital Well-being Initiatives:**
   - Develop Guides for Parents & Students on safe and responsible technology use.
   - Implement Awareness Campaigns around misinformation, cyberbullying, and financial fraud prevention.
5. **Ongoing Risk Assessment & Ethical AI Implementation:**
   - Conduct Regular Safety Audits to identify emerging risks.
   - Ensure Karot aligns with global Responsible AI and Child Safety Standards.