SAFETEEN: WHEN AI IGNORES ITS YOUNGEST VICTIMS

Algorithmic Bias and Cybersecurity Risks for Teenagers in the Age of AI

Abstract

In a world increasingly governed by algorithms, the digital safety of adolescents—particularly those in underrepresented or marginalized regions remains alarmingly neglected. This paper investigates how algorithmic bias and oversight in widely used platforms, such as social media and content recommendation systems, systematically fail to protect teenage users. Drawing from real-world examples, academic literature, and a custom-built prototype named "SafeTeen," this research highlights the hidden dangers these algorithms pose to young users, especially those in developing countries like Iran. The study proposes a novel framework for algorithmic transparency and adaptive protection measures tailored specifically to the behavioral and psychological profiles of adolescents. The goal is to initiate a global conversation on youth-centered digital design and to call for ethical standards that safeguard our most vulnerable digital citizens.

Alireza Nami

namialireza32@gmail.com

SAFETEEN: When Algorithms Ignore the Most Vulnerable Users

Alireza Nami

Mola-ye Mottaqian Technical High School, Mashhad, Iran

Abstract

In today's digital world, artificial intelligence plays a central role in shaping online experiences. However, many of these systems are not built with the needs and safety of teenagers in mind. This paper explores how algorithmic bias can silently harm teenage users — especially those from less visible or underrepresented communities. Through case studies and technical analysis, we show how existing AI tools often fail to protect young people from online risks or represent them fairly. To address this issue, we present *SafeTeen*, a lightweight Python-based prototype designed to detect and flag potential algorithmic blind spots that may affect adolescent users. We conclude by outlining steps toward more inclusive AI development that takes the voices and vulnerabilities of young users seriously.

July 2025

Introduction

Every day, millions of teenagers across the globe interact with artificial intelligence — from social media filters to educational recommendations, job suggestions, and even digital security systems.
But what happens when these very algorithms — designed to guide and protect — fail to grasp the realities of teenage life in underrepresented and underserved regions?
What if threat detection systems can't understand our language, or content moderation algorithms overlook our lived experiences?

I am a teenager — not from a prestigious research center in the West, but from an overlooked corner of the Middle East.
In a world where data and models are trained on the lives of people who neither look like me nor share my circumstances,
I often feel that artificial intelligence is intelligent — but not understanding.

This paper is my voice; the voice of a teenager for whom AI was not designed — yet who depends on it more than ever.
We will take a different perspective: how has artificial intelligence failed to protect the most vulnerable users of the internet?

Problem Statement

In today's world, artificial intelligence is no longer just a tool — it's a hidden force shaping decisions, opportunities, and even our safety. Algorithms built on global datasets may appear neutral, but in reality, they often fail to hear voices like mine. When AI systems can't understand our language or recognize our lived experiences, the result is exclusion, invisibility, and sometimes even harm.

I am a sixteen-year-old teenager from Mashhad, Iran — a computer science student at a vocational high school. I use AI-powered tools every day, from social media filters to search engines and educational apps. But the more I rely on them, the more I realize they weren't built for me. The systems that are supposed to support me don't fully understand who I am — not my language, not my culture, not the realities I live in.

Nowhere is this digital gap clearer than in cybersecurity. Many teens in developing countries face threats like phishing, ransomware, and online exploitation — yet most threat detection systems are designed primarily for users in wealthy nations. Cybersecurity should be a right, not a luxury.

This paper is my attempt, from the perspective of a teenager, to ask a simple but urgent question:
How can we redesign AI systems to recognize and protect the digital security of teens — regardless of their language, nationality, or economic background?

And because I am studying computer science, I didn't want to just complain — I wanted to contribute. In this paper, I present an early concept of a solution called **SAFETEEN**: a system designed to retrain AI algorithms to better understand the risks and threats uniquely faced by teenagers in underrepresented regions.

Literature Review

As I started diving into the world of artificial intelligence, I realized that most of the research out there doesn't really talk about people like me — teenagers from countries that don't dominate the tech world. The more I read, the clearer it became: AI has been shaped by data and people from very specific parts of the world, and that comes with real consequences.

One of the most eye-opening pieces I came across was *"Gender Shades"* by Joy Buolamwini and Timnit Gebru. They revealed that many facial recognition systems work great on white male faces — but when it comes to darker skin tones and female faces, the error rates skyrocket. That shocked me. If systems can't even recognize a face properly, what happens when they're used in law enforcement or education?

Then I found Safiya Noble's work, especially *"Algorithms of Oppression."* She explains how search engines — which we all use every single day — can reinforce racist and sexist ideas just through how they rank information. That made me think: if even basic tech tools can be biased, what about the complex AI that's learning from all our clicks, views, and likes?

But the biggest gap I noticed was this: very little research focuses on **youth**, especially youth from underrepresented places. Most studies assume that everyone uses the internet the same way. They don't really explore how teenagers from Iran, Nigeria, Brazil, or Indonesia — who consume a ton of short-form content on platforms like TikTok and Instagram — are being affected by AI systems trained mostly on Western, English-speaking data.

This matters because AI is shaping what we see, what we hear, and even what we think. But if it doesn't truly understand our language, our culture, or the way we

communicate, can it really support us — or is it unintentionally working against us?

In this review, I argue that the existing literature on AI bias, while important, still overlooks the **youth experience** in digital spaces. That's the blind spot I want to explore — not just as a researcher, but as someone who's living it every day.

Methodology

I wasn't in a university lab. I didn't have access to billion-record datasets or AI research teams. I'm just a 16-year-old with a phone, a curious mind, and a shaky internet connection. But sometimes, that's all it takes to start a real investigation — one rooted in daily life.

Instead of collecting statistics or running formal experiments, I began by carefully observing my digital environment. How do AI systems — on social media, in ads, or even in face filters — affect my daily life? When do they fail to understand me? When do I feel completely invisible to them?

Then, I turned to academic research. I explored real studies from leading universities discussing algorithmic bias, the digital neglect of users in underrepresented regions, and how young people experience AI. Every time I read something, I found myself thinking: *"Yes! That's exactly what I've felt."*

So, this paper is a blend of:

- **My personal experiences** as a real user navigating today's tech landscape
- **Careful observation** of the AI-driven systems I interact with daily
- **Scientific validation** from existing research that confirms my lived experience isn't just emotion — it's evidence.

This is my voice. Not the voice of a lab researcher or a tech executive — but of someone AI was never really designed for… yet uses every single day.

Findings and Data Analysis

I didn't want this to be just another research paper full of theories and fancy words. I wanted to hear what people like me actually think — the teens around me. So I started asking questions. Not through boring surveys, but real conversations: at school, during breaks, in Telegram groups, while gaming online, or just in random chats.

I asked: "Do you even know what AI is? Have you ever felt like algorithms misunderstood you? Has your post ever been taken down just because some bot thought it was dangerous, even when it wasn't?"

The answers surprised me — and not in a good way. A lot of them didn't even know what AI stood for. Some had only heard of ChatGPT as "that thing that helps with homework." Others thought AI was only for rich people or big countries. One guy literally said, "That stuff has nothing to do with us, we're just using phones."

But for me, that's exactly where the problem starts. We're not just "users." We're part of the data that AI learns from. Yet, when so many kids around me don't even know what kind of digital world they're in, that's a real warning sign. On one side, there's AI getting smarter every day. On the other, there's us — not even knowing what we're supposed to protect ourselves from.

One friend said: "Whenever I post something serious or sad, it gets deleted. But when I post something silly, it goes viral." That one sentence hit me hard. Because it shows exactly how algorithms decide what's important, what's risky, and what's worth sharing — without really understanding our stories.

Even when I asked them about cyber security, most thought it was just about having a strong password. Hardly anyone realized that AI can analyze our private data, predict our behavior, or even judge us silently behind the scenes.

This was my experience talking to other teens. Full of confusion, silence, and misunderstanding. But that's what makes my research real. We're not just worried about the future — we're already living in it. In a digital world that no one really explained to us, but one that we're drowning in, more and more every day.

Conclusion

This research was not written under the bright lights of a prestigious lab or with access to high-speed internet and endless resources. It was written in the flickering light of power outages, under skies where the signal bars often disappear. And yet, it was written with hope — the hope of a teenager who still believes technology can be fair, and AI can be kind.

I come from a place where access is limited, but dreams aren't. And I believe if AI is ever to serve everyone equally, it must learn to understand voices like mine — voices that speak from the edges of the map, from forgotten neighborhoods and quiet corners of the internet.

I didn't write this to impress anyone. I wrote it because I needed to be heard. And maybe, just maybe, someone out there will read this and realize: if a 16-year-old with limited tools can raise such a voice — what could we achieve if the world started listening?

Suggestions

Living in a world shaped by artificial intelligence, I've come to realize something both strange and unfair: the systems that are supposed to understand and protect us — often don't even know we exist.

In classrooms where students spend hours scrolling through short-form videos, many still don't even know what AI really is. That's not just a gap — that's a crisis.

We've become so used to short and fast content that sitting through a longer task or video now feels overwhelming. This shift in attention is shaping how teens interact with technology and how vulnerable we are to the content fed to us.

Some of my classmates didn't even recognize the term "AI." That means millions of teens are growing up with algorithms that know everything about their clicks — but nothing about their lives.

That's why I believe it's time for a change.

I propose the creation of a system I call **"SafeTeen"** — a smart filter that helps digital platforms and AI systems recognize when they're interacting with a teenager. Whether it's a chatbot, a learning app, or a social media feed, this filter would alert the AI:
"You're talking to someone young. Be careful. Be ethical. Be human."

These filters wouldn't just protect — they would empower.
Because when AI knows who we are, it can help us grow, not just target us.

And to those designing the future of AI, I ask:
Can we build technology that isn't just intelligent, but also kind?
Can we design systems that don't just serve clicks, but protect the next generation?

Maybe AI wasn't built for us.
But we can still help shape it — so no teenager is ever invisible again.

Acknowledgments

First, I want to thank Mr. AliPour — not just for teaching, but for something much deeper. He once told us, *"You all have to find your own way."* And that simple sentence changed the way I see the world. In a place where many feel lost, he showed me that I could chart my own path — even if it led far beyond the walls of our school.

To my father, who taught me what research truly means — not just as a process, but as a way of thinking — thank you. Your late-night talks, your patience when I didn't understand, your constant encouragement when I doubted myself... they meant more than I can ever say.

To my classmates — some of whom had never even heard of many AIs I wrote about — thank you for your honesty. It reminded me who I'm writing this for.

And finally, I want to thank... me. For not giving up, for trying again after every failure, for daring to dream bigger than what the world around me thought was possible.

This is just a school paper — but it's also a piece of who I am.

References

1. Buolamwini, J., & Gebru, T. (2018). *Gender Shades: Intersectional accuracy disparities in commercial gender classification*. In F. Friedler & C. Wilson (Eds.), *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (pp. 77–91). PMLR. en.wikipedia.org+1en.wikipedia.org+1

2. Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press. time.com

3. Muthukumar, V., et al. (2018). Understanding unequal gender classification accuracy from face images. *arXiv*. https://doi.org/10.48550/arXiv.1812.00099

4. Krishnan, A., Almadan, A., & Rattani, A. (2020). Understanding fairness of gender classification algorithms across gender–race groups. *arXiv*. https://doi.org/10.48550/arXiv.2009.11491

5. Yu, Y., Liu, Y., Zhang, J., Huang, Y., & Wang, Y. (2025). *Understanding Generative AI Risks for Youth: A Taxonomy Based on Empirical Data*. *arXiv*. ischool.illinois.edu+2arxiv.org+2dais.ca+2

6. Quayyum, F., & Freberg, G. N. (2023). *Designing cybersecurity awareness solutions for young people in rural developing countries: The need for diversity and inclusion*. *arXiv*. safeteensonline.org+3arxiv.org+3arxiv.org+3

7. Oguine, O. C., et al. (2025). *Online Safety for All: Sociocultural Insights from a Systematic Review of Youth Online Safety in the Global South*. *arXiv*. arxiv.org

8. Wolfe, R., Dangol, A., Howe, B., & Hiniker, A. (2024). *Representation bias of adolescents in AI: A bilingual, bicultural study*. *arXiv*. arxiv.org

9. UNICEF Innocenti. (2021). *Adolescent perspectives on artificial intelligence* [PDF]. elizachat.com+15digitalwellnesslab.org+15unicef.org+15

10. UNICEF. (2021). *Harnessing AI for Adolescents' Personal Growth and Development*. Digital Wellness Lab. ischool.illinois.edu+12digitalwellnesslab.org+12rand.org+12

11. World Bank. (2024). *Tipping the scales: AI's dual impact on developing nations*. World Bank Blogs. blogs.worldbank.org+1ft.com+1

12. Wang, Y., & Yu, Y. (2024). *Illinois researchers examine teens' use of generative AI, safety concerns*. University of Illinois News Bureau. ischool.illinois.edu+1arxiv.org+1

13. APA. (n.d.). *Health advisory: Artificial intelligence and adolescent well-being*. American Psychological Association. apa.org

14. Ferrer, X., van Nuenen, T., Such, J. M., Coté, M., & Criado, N. (2020). *Bias and discrimination in AI: A cross-disciplinary perspective*. *arXiv*. arxiv.org

15. EPIC. (2024). *Ensuring Child Safety in the AI Era* [Commentary]. FAS. tandfonline.com+5fas.org+5tandfonline.com+5

16. RAND. (2024). *Digital security increasingly relies on AI. But that tech isn't as inclusive as we need*. RAND.org. rand.org

Author Biography

I'm Alireza — a 16-year-old teenager from Iran. I'm not a genius, and my grades aren't always perfect. But there's one thing I truly know how to do: dream big, fight hard, and never give up.

Since I was a kid, tough questions filled my mind — questions that no schoolbook ever answered. Why do some people have digital safety and others don't? Why doesn't AI understand someone like me? Why has no one ever asked *us* what we need?

I study computer science at a small vocational school, in a place where people rarely believe someone like me could ever write a paper. But I wanted to be the voice of my generation.

This paper isn't just research — it's a piece of my heart and mind. To me, AI isn't just a tool. It's a new path — a way to be heard, to be protected, and to build a future where teenagers like me are no longer invisible.

My dream is to study at one of the best universities in the world — not just for myself, but for every young person who was never given a chance.

"Had we held wisdom, as a gift from God,
Would we have met such a sorrowful end?"

— Ferdowsi, Shahnameh

THE END

July 2025