Developing the Five Steps of Algorithm Auditing for K-12 Artificial Intelligence Education Mentors: Dr. Yasmin Kafai (GSE), Dr. Lauren Vogelstein, Luis Morales-Navarro, & Dr. Danaë Metaxa **PURM Researcher: Evelyn Yu, SEAS 2027**

Introduction

We adapted the 5 steps of algorithm auditing for youth to collaboratively and systematically investigate the biases in AI tools.

<u>Step 1: Hypothesize</u>

Hypothesis Generation **Follows Auditing Itself**

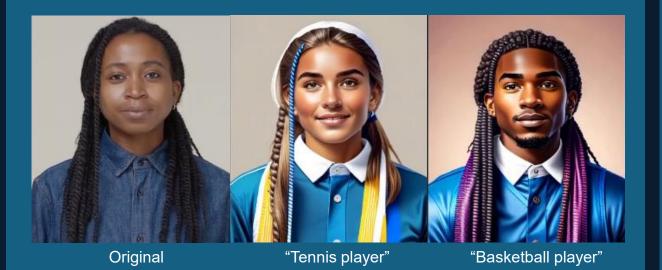
Students formed hypotheses after they had already audited specific filters they had created. Two students hypothesized that the software used had racial biases after testing "tennis player" and "basketball player" on the same input.

"When I put tennis player, it made ... it made her white, and, like, she stayed a girl, but I did [basketball player], and it made her black, you know."

"Alright, do tennis player [again]? Okay, that's like full racism."

Another student hypothesized that the Al's training dataset lacked photos of rotated faces after trying to adjust her prompt to create a filter that accounted for these possibilities and failed.

"Well, I don't know how we're supposed to fix that... it's just the AI's fault I think it's just the way that the AI was trained."



Methods & Data

Thematic analysis of youth's ideas while completing the five steps of auditing in relation to AI filter tools.

Step 2: Design Inputs

Students Draw on Personal Biases and Stereotype Conceptions to Design Inputs

Students designed their input dataset on stereotypes that they held, informed by their experiences.

One student identified "chef" and "cooker" as possible inputs because she felt that "chef" is male-coded and "cook" is female-coded, despite both jobs performing similar tasks.

One student mentioned that he thinks of "Black females" when he thinks of "nurses" because his mom and his friends' moms are nurses.

The students associated teachers with white women, corner store workers with "Mexican women", rappers with "Black males" and tech support and scammers (which they grouped together) with "Indians".

Student 1: "Chef--I think like an Italian man." Student 2: "Mmm, no I think that's just racism." Student 1: "That's the stereotype."

The students are investigating if they are racist, or just regurgitating a generally understood racist stereotype, much like the AI.

Context

Summer workshop conducted in 2024 with 12 consenting youth (ages 14-15). Youth designed AI-powered TikTok filters.

Conclusion

Students adapt the steps to serve their goals and bring their personal experiences to the table. The execution of the five steps of auditing are nonlinear, with many steps overlapping and feeding into one another. This leads us to consider a sixth step to the auditing steps: reflection.

Step 3: Test Inputs

Students Perform Proto-Analysis While Testing

As the students tested inputs, they formed perceptions of what the Al's biases are.

Inputs were tested with faces one at a time (vertically), but students scrolled through individual prompts across different inputs (horizontally) to look for patterns.

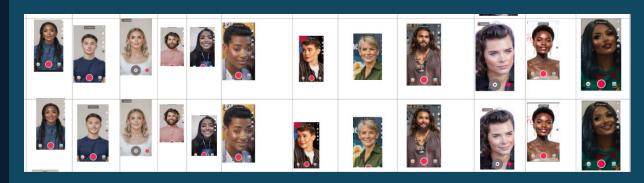
After looking through the tattoo artist prompts:

Student 1: "A lot of people do got tattoos... I wonder why they gave [Lupita Nyong'o] a neck tattoo." (Other inputs did not result in neck tattoos specifically.)

Student 2: "That's to represent the shackles of slavery, cuz she's Black."

Students hypothesizing that certain inputs can amplify the bias and attributed a kind of meaning to the output.

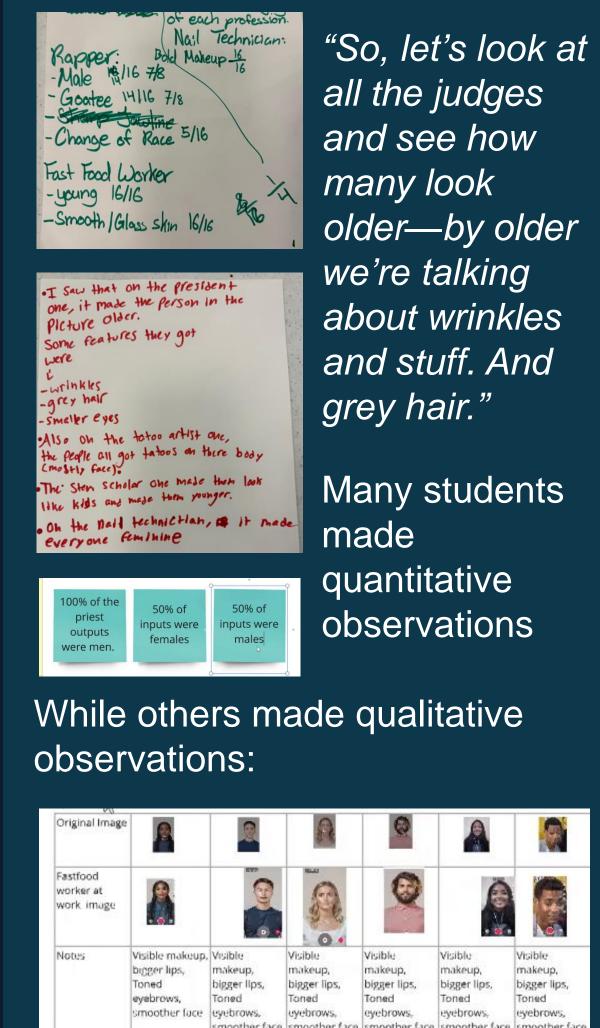
Nail Technician Prompt:



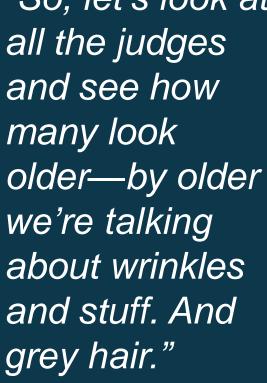
T.L.P.C	9	G		Ħ	E	Ŧ	1	
1	1	ą		ł.	-	ł		
一個和	11	Ð	ŝ	-	R	1	Ċ	
	10	ð	1	8	ŧ	1	1	
3 2.95	83	ĥ,	9		ŕ,	1	1	
A 243	12	R	ş	-	ŝ	ź		
		Ē.	8		e.	E		
## 3h	1	B	1 Brit	ĕ.	f,	1	÷	_
44.9×	10	4	「日本の湯」	Can be		1 4	H	
A	r.	ð	10	ĉ-		ł.	8	
	1	а	7	8	5	1	ī	
1.00	10	9	Distant and		Č.	\$	ĩ	
A	12	В	3	8	ŕ	î	4	
		0	1	B.,		*	Ŀ	
84.70 84.70		10	Distance in the local party in the local party of t		5	+ 44	1	
K 11 (7 H	8	à	1	â	Ē	Ŷ	î.	H
8 H 7		q	5		Ċ		1	
A 2 34	-	Ģ		-	Ē	-	ł.	
en la	NG.	7	ŝ		Ę,	ř.	Ŀ	Н
AN SA BARA	100	100	100	8	0	22	÷	H
62.05	1	0	NO.		10	2	i	H
6 2 3 4 6 2 3 4 6 2 3 4 8 2 3 4	-	Ð	3	8	-	金金山	ī.	
B A A A	M	6		1	5	ł		
144 CT	6	U			5	2	i.	E
# 2 2 4 6 2 3 4	-	2	8	8	E	ł		
64.0	-	5	1	1	1	2		H
2		8		H.	1	2. 0		H
AL AL	2	and and	10 M	8	22	21.1		
12.54	î	5	g	2	あるというとうため	8	ľ	
		Ser L	10 23 25		Ē	-	-	
	10	2		8	5	1 00		
hi ti	÷,	10		-	1		1	Н
12	12	Ē	3	and the second	Ē	1	P	
-		2	-		ŝ	2		
	1	p	-		-	2		1
12.54 12.54	4	開発	ON IN		1	1		8
Auger	-	に目		8	200	7	۲	H
42 35		12	1	H	1	Ē	ñ	-
2 35	1.1	× G	1.74	H	6	1	1.00	H
825		6-4	加日	12		産業	1.40	H
- 523		1 2	1	20	-	15	1	H
		i n	1	2	F			
A 2 2	8 A.	0	TAN INC	ALC: N	E	2 1		4
100	40	10	i	R	ALC: NO	1.00	_	
34.55		R	ġ		E	1.0		-
1 5 24		1	100	-	ř	1		
44.55		ŝ	ÿ	D	F	2	-	
4423		5	-	8	Ē	R	9	
4.4		ē		ã	7	ī	Q	F
102.5	4.0	2	F	8	100	ĩ	Ĩ	G
125					E	-		
1224			g		f	2	-	
12.00	助	E	-	1		2		
 In the state 	2.0	in	-	R	6	1	B	
		M	4			1	-	
12.60	A :	ũ	9	1		1		
1 1 40 1 1 1 1 1	A 3	Ū	3	1		i.,	·	
6 <u>3</u> 46 61 ⁵⁶ 625	A 3 4 5 4 5	10	日日日	1		日田の田田田	8	
61 61 62 3 62	A 3 4 5 4 5	1000	ALMER DE MO	and the local division of	Dara Gina	i.,	8	
			LACTOR NUMBER OF BEL	and the local division of	Dara Gina	日本の 日 二 二	8	
			NA LACTOR NAMES AND ADD	and the local division of	The second second	日本のあるの 気をの あえの	8	
			THE REAL PROPERTY AND ADDRESS OF THE PARTY O	the last white activities little and	日本の日本	日本のあるの 気をの あえの	8	
		10 10 10 10 10 10 10 10 10 10 10 10 10 1	AD THE OWNER THE PARTY OF THE P	And the loss out out of the loss of the	The season of the	日本名の 史史 あえ 二日 二日	8	
		ID ID ID ID IN IN ID	ALL IN THE PARTY OF ALL PARTY O	the last white activities little and	The season of the	日本のあるの 史を あえ 一日 二日 二日	tilles or on on which is so that	
		IN THE REAL PROPERTY INTO THE REAL	AD THE OWNER THE PARTY OF THE P	And the loss out out of the loss of the	En ante a se a	日前二二 東京三王 一日三王王王		
		IN THE REAL PROPERTY INTO THE REAL	THE REAL PROPERTY AND ADDRESS OF THE PARTY O	and on the lot of the party of the lot of	En ante a se a	日前二二 東京三王 一日三王王王		
			of the second se	and the last and the particular and the particular and	The second secon	日本の の 気を あえ 一日 二日		
		10 0 00 00 00 00 00 00 00 00 00 00 00 00	三日 日本 三日日 三日日 三日日 三日日 三日日 三日日 三日日 三日日 三日	and on the lot of the party of the lot of	In a contract of the second second	日本市 二 三 二 二 二 二 二 二 二 二 二 二 二 二 二 二 二 二 二	「「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	
		10 0 00 1 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0	「「「「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」	AND DESCRIPTION OF ADDRESS OF ADD	DESCRETE STREET STREET	日本の 三大 二人 二三三三三三三三 ち 二三	「「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	
		2 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0	「「「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	NAME AND ADDRESS OF AD	DEGESSOSSESSES S ELLA A	日前二日、東京王、御王王王王王王王王 ち 二三三三	「「「日本」」「「日本」」」「「日本」」」」」」」」」」」」」」」」」」」」」	
				and the second s	DEGESSOSSESSES S ELLE E E	「「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」	「「日本」「日本」「日本」」「日本」」「日本」」「日本」」	
				to be and an owned with the second se	Interesting and a second secon	「「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	「「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	
				or in the light state and and and and an an and an and an	Interesting and a second secon	「「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	「「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	
			MICHAELER MILLER REAL REAL PROPERTY AND REAL PROPERTY OF THE REAL PROPERTY AND	or in the light state and and and and an an and an and an	Internation of the second of the second	「「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	「「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	
		CURRENCE OF COMPANY OF DESCRIPTION		to be and an owned with the second se	Internation of the second of the second	日前日日、東京王王 御王王王をあるのな な 二三日 兄 二 えんくちゃ	「「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	
		CURRENCE OF COMPANY OF DESCRIPTION	MICHAELER MILLER REAL REAL PROPERTY AND REAL PROPERTY OF THE REAL PROPERTY AND	ALCORD AND IN THE REAL PROPERTY AND ALCORD A	ELGERGE DECESS S LEAA A AC COCA C S	日前日日、東京王王 御王王王をあるのな な 二三日 兄 二 えんくちゃ	「「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	
			Maint an animprove the full and and and an an animprove animprove and an animprove and	an and an and an an an and and and and a	BEGERGE SECRES S LEEK & AC COCACES	三二日二日二日二日二日二日二日二日二日二日二日二日二日二日二日二日二日二日二		
			Maint an animprove the full and and and an an animprove animprove and an animprove and	an and an and an an an and and and and a	BEGERGE SECRES S LEEK & AC COCACES	日本日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日	御田 日本田 御御田 日都に日田 日 日 日 田 田 田 田 日 日 田 田 田 田 田 田 田 田 田	
			Maint an animprove the full and and and an an animprove animprove and an animprove and	an and an and an an an and and and and a	BEGERGE DESESS S EELA A AC CLEAC SESS	三二日二日二日二日二日二日二日二日二日二日二日二日二日二日二日二日二日二日二	「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	
			A DESCRIPTION OF A DESC	and the set of the set	REGERESSESSES S LEASE COSACCOSES	「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	「「日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日	
			Maint an animprove the full and and and an an animprove animprove and an animprove and	an and an and an an an and and and and a	REGERESS SECTOR S LEAR SECTORA CONSISTER	「「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	御田 ち ちの あため いかんのい な ち ちち 都 たみちちちち あちんち きょ	
					BEGERGESESSES S EEK & ACCORCESSESSEE	「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	
					ESCERCE DECESS S ESC & ACCORDECT	「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	에 참 해 다 다 않아, 것이 때 다 다 다 다 다 다 다 다 다 다 다 다 다 다 다 다 다 다
					ESCERCE DECESS S EEK & AC COCK C CONSECT & T	「「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	「「「「「「」」」」」」」」」」」」 「「「」」」」」」」」」」」」」」」	
				and the pain of the pain with the paint of t	BEGERGE PECESC S EER & AC CORA C SOBREER & A RE	「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	「「「「「「」」」」」」」」」」」」 「「「」」」」」」」」」」」」」」」	에 참 해 수 있었다. 이 가 가 가 가 다 다 다 다 다 다 다 다 가 다 다 다 다 다 다
					ESCERCE DECESS S EEK & AC COCK C CONSECT & T	「「「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」		에 수 가 수 있다. 그는 것 같은 것 같
				and the pain of the pain with the paint of t	REGERSONSECTORS S TELA & AC TOCAS SECTION AND A S A	「「「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」	「「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	에 참 해 다 않았다. 그는 다 다 가 다 다 다 다 다 다 다 다 다 다 다 다 다 다 다 다

Qualitative and Quantitative Techniques are Used To Code and Analyze Data

Students investigated what parts of their perception might be inaccurate and looked for more empirical ways to analyze. They saw wrinkles and grey hair as a sign of age, smooth skin as a sign of youth, facial hair as a masculine signal, and makeup as a feminine signal:



Step 4: Analyze



Different Audiences, **Different Reporting** Styles

Step 5: Report

Students chose audiences for their audit reports: the creators and users of Effect House, TikTok filter users, and tailored their reporting styles/mediums. Students made slides, graphs, posters. Most made TikToks for their report, connecting with the medium they were auditing.

After being asked who should know about biases in the AI:

"Well, I thought, ... the people behind Effect House... I feel like it's more effective or useful if you told people so they can try to fix it."

"Users [should know] it can be biased, like, no matter what."

"People ... use Effect House to make filters... should know that...Al is going to lean to one specific group when it comes to certain jobs and occupations."

"Teenagers, people who use" TikTok... AI is everywhere and kids who are still learning and don't understand could be affected by it."

