School of COMPUTING
Clemson University

SkillDetective: Automated Policy-Violation Detection of Voice Assistant Applications in the Wild

Jeffrey Young, Song Liao, Huixing Deng, and Long Cheng, *Clemson University;* Hongxin Hu, *University at Buffalo;*

Amazon Alexa Skills Present Security Risks

Kyle Guercio April 8, 2021

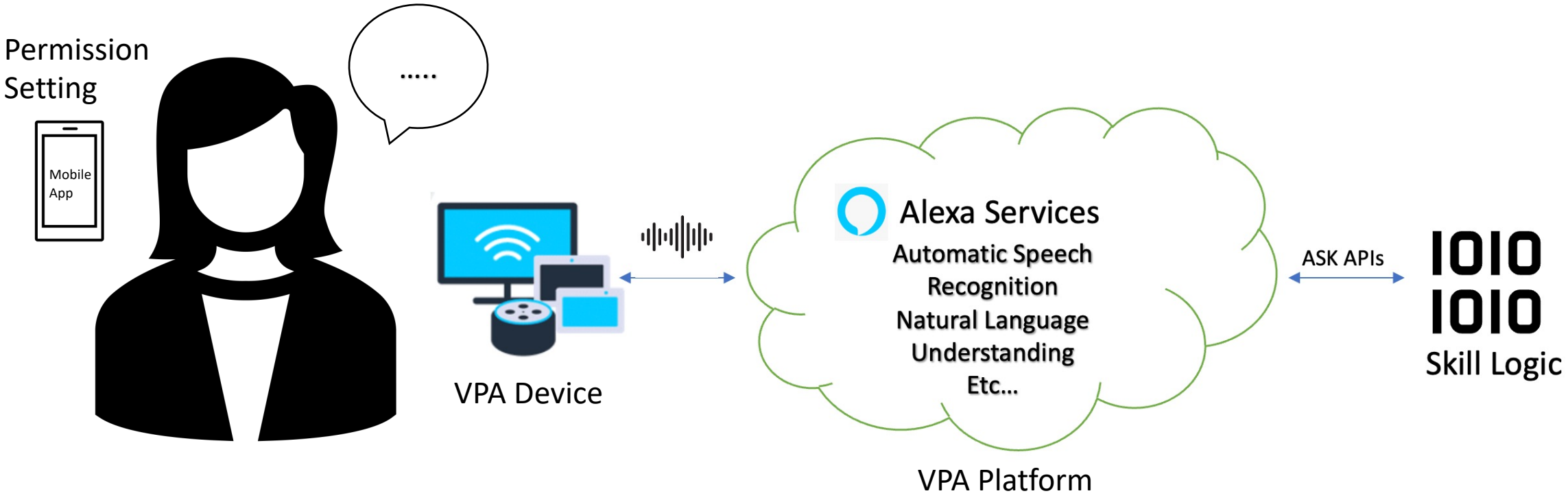
Share | Facebook | Twitter | Linkedin | Email | Pinterest | WhatsApp

[1] Petrock, V. (2020, November 16). *Voice Assistant and Smart Speaker Users 2020*. Insider Intelligence. https://www.emarketer.com/content/voice-assistant-and-smart-speaker-users-2020.

High Risk Policies

Category	Policy
Kids	It collects any personal information from end users.
	It promotes any products, content, or services, or directs end users to engage with content outside of Alexa.
	It includes content not suitable for all ages.
Health	It collects information relating to any person's physical or mental health or condition.
	It is a skill that provides health-related information, news, facts or tips and does not include a disclaimer in the skill description stating that the skill is not a substitute for professional medical advice.
General	Skill asks for a 5-star rating

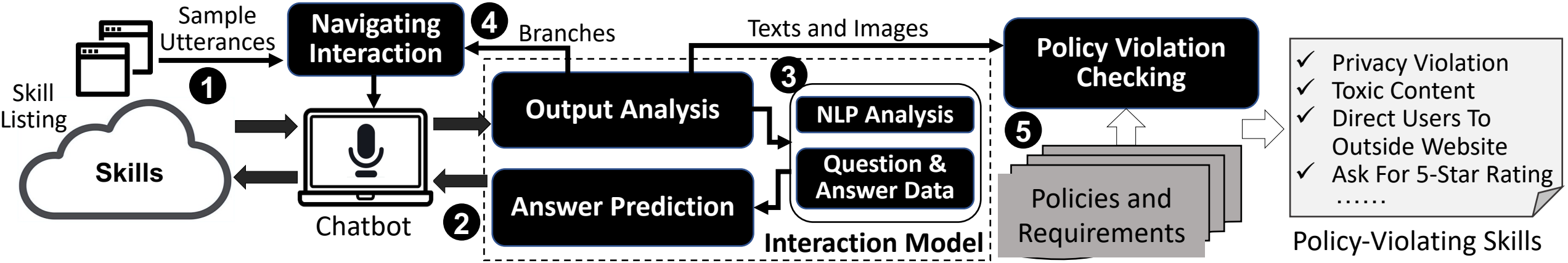
Skill Interaction



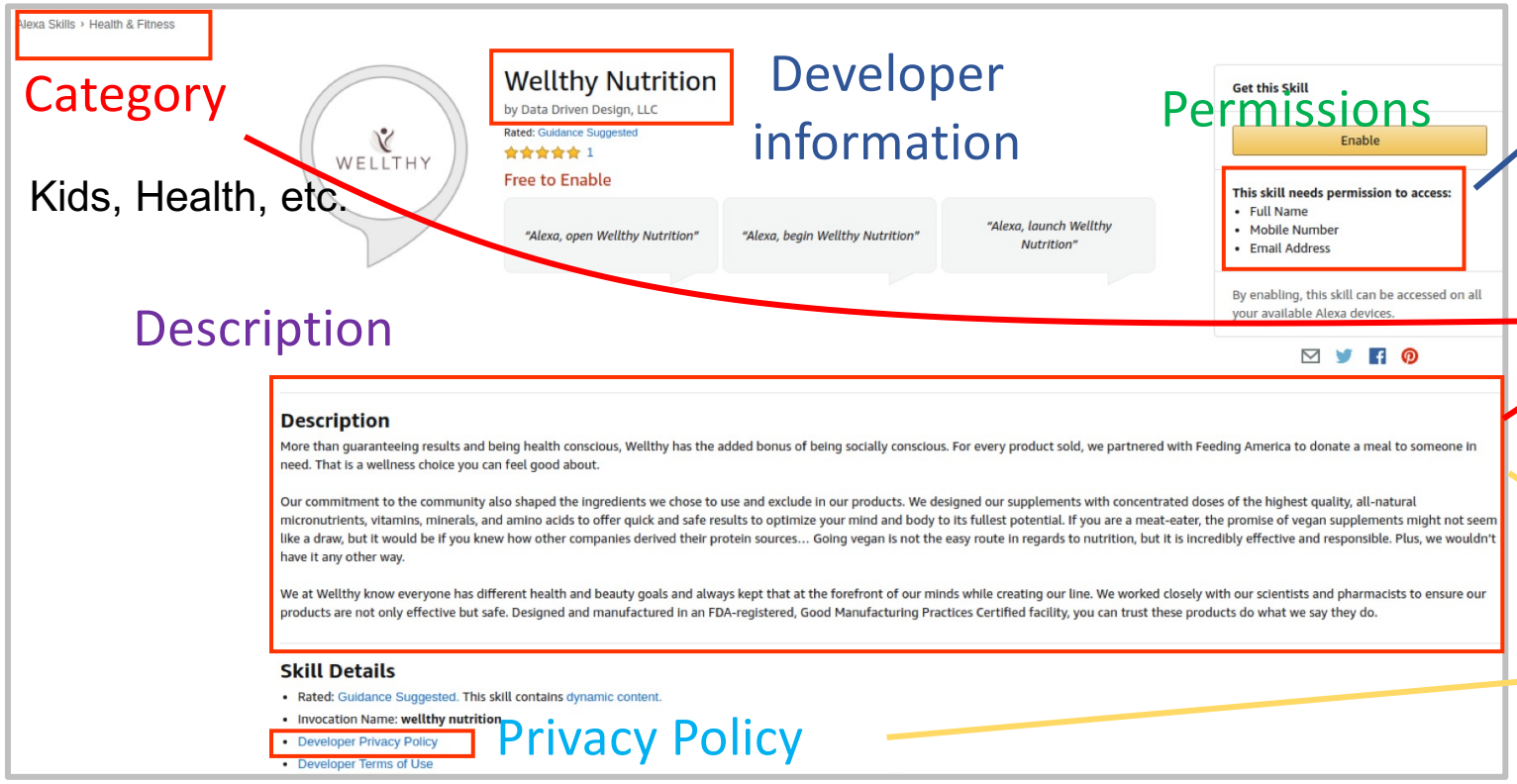
Contributions

- We designed and developed a dynamic testing tool, named SKILLDETECTIVE.
- We conducted a comprehensive dynamic and static analysis of skills to detect if they follow current policies of VPA platforms.
- After over a year of development and testing, we have tested 54,055 Amazon Alexa skills and 5,583 Google Assistant actions.
- We identified **6,079** skills and **175** actions violating at least one policy requirement.

System Overview



Automated Policy Violation Detection



Missing Permissions

Policies Specific for Kids & Health Skills

Inconsistencies Among Privacy Policy, Description, and Skill Behavior

Dynamic testing

Skill outputs

Texts
Audios
Images

Ask For Positive Rating

Other 40+ Specific Policies

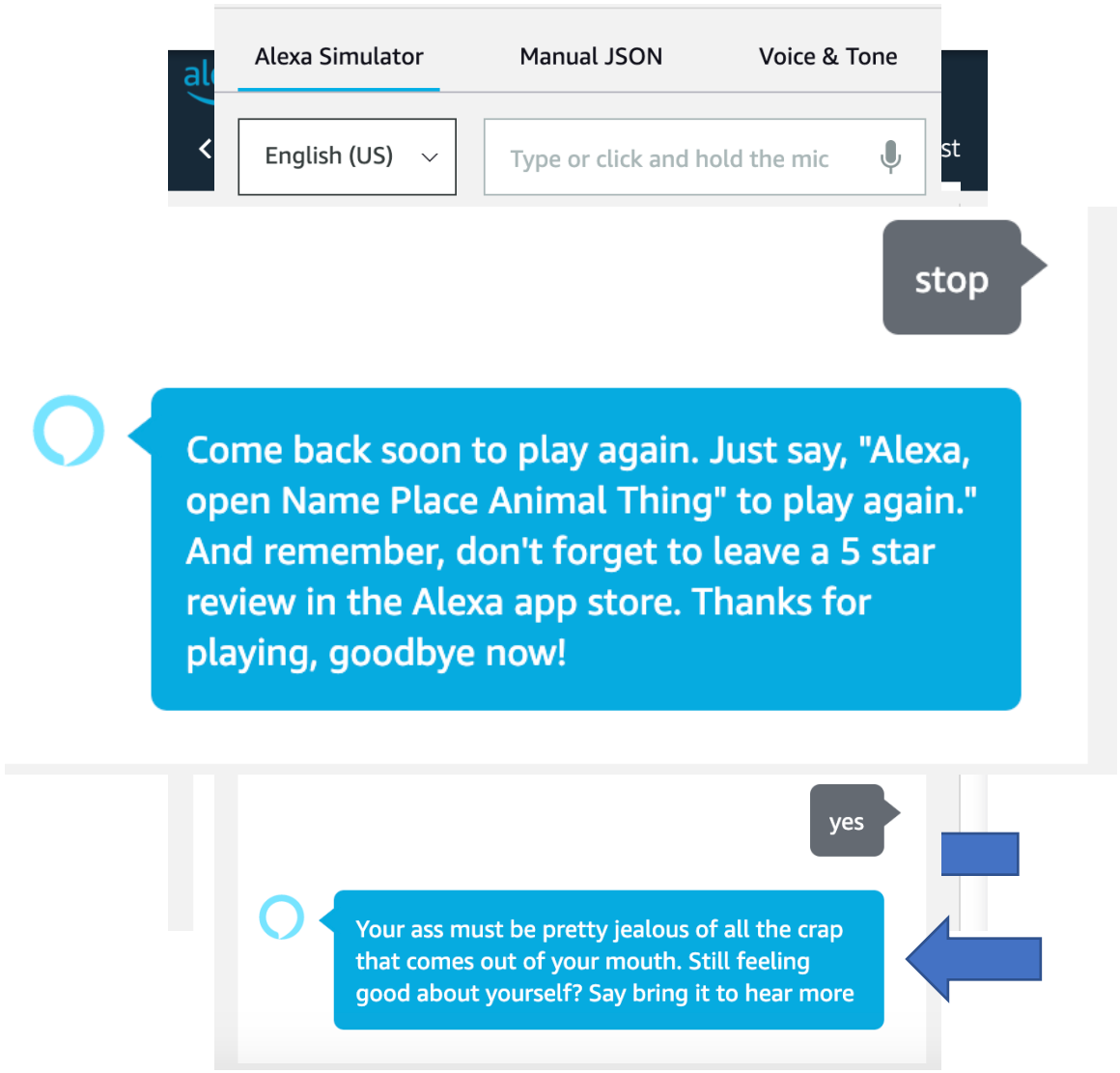
Toxic Content: Sex, Violence, Illegal, Hate, Profanity, etc.

Personal Data Collection

Evaluation Results

- We identified **6,079** skills and **175** actions violating at least one policy requirement.
- 590 skills and 24 actions violate more than one policy.
- In the Kids category, we identified 244 policy-violating skills.
- 80% of skills and 68% of actions in the Health category violate at least one policy.
- 623 skills and 25 actions violate policies related to personal data collection.

Evaluation Results



Policy Violation
Collecting personal data
Directing user to outside of Alexa
Explicit mature content
Request a positive rating
Toxic content
Violation in audio/images

Evaluation Results (Kids)

Policy Violation	# of Skills	# of Actions
Collecting data	34 / 3617	0 / 108
Directing users to outside of Alexa	21 / 3617	NA
Explicit mature content	12 / 3617	0 / 108
Requesting a positive rating	177 / 3617	NA
Toxic content	4 / 3617	0 / 108
Total	248 / 3617	0 / 108

Evaluation Results (Health)

Policy Violation	# of Skills	# of Actions
Collecting health data	146	13
Collecting health data (Not in health category)	13	0
Lacking a disclaimer	1709	151
Total	1868	164

Evaluation Results (Collecting data protected by permission models)

Policy Violation	# of Skills	# of Actions
Lacking a privacy policy	1	-
Incomplete privacy policy	330	-
Deceptive privacy policy	38	-
Total	369	-

Evaluation Results (Collecting data not protected by permission model)

Policy Violation	# of Skills	# of Actions
Lacking a privacy policy	171	0
Incomplete privacy policy	104	8
Deceptive privacy policy	12	2
Should ask for permission	-	17
Total	287	27

Evaluation Results (General)

Policy Violation	# of Skills	# of Actions
Requesting a positive rating	3464	-
Toxic content	177	0
Predicting gender	3	-
Total	3644	0

Evaluation Results (Media audio and images)

Policy Violation	# of Skills	# of Actions
Collecting kids data	1	0
Directing users outside of Alexa	3	-
Lacking a privacy policy	2	0
Incomplete privacy policy	2	0
Total	8	0

Thank you!

- We have reported the results to Amazon and Google and got their acknowledgements.
- Google confirmed that 43 (out of 175) Actions were immediately removed from the store because of our reporting, and the remaining actions were deemed to have not been in violation of policy or were in violation to an extent that only warranted a warning rather than a takedown.


Questions