

음성비서(VA)와 프라이버시 감사



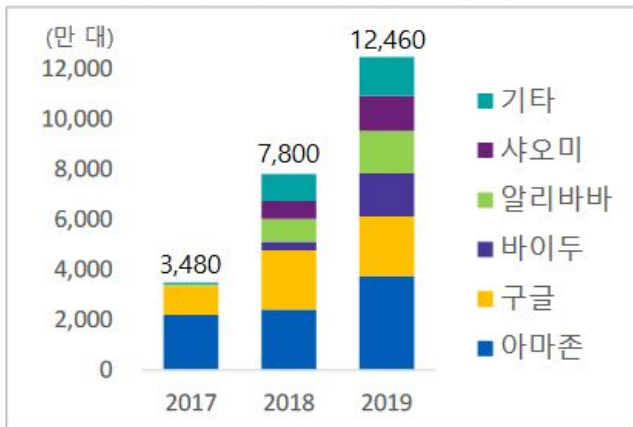
음성비서 (Voice Assistant, VA)

- 음성을 기반으로 한 유저 인터페이스 제공
- 사람이 수행해야 할 일을 음성 명령을 통해 수행
- 다양한 내외부 기기 제어, 특정 어플리케이션 등 서비스 이용, 전화, 질문에 대한 응답 등등
- 음성인식, 자연어처리, 클라우드, 음성합성 등등
- 스마트폰에서 스마트 스피커로
- 하나의 음성서비스 제공에서 음성비서 플랫폼으로 (like 앱스토어)

국내외 현황

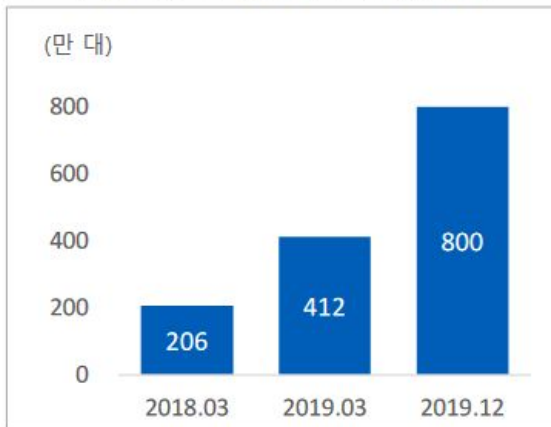


>> 글로벌 스마트 스피커 연도별 출하량 추이



Source: Canalis Press Release, 삼정KPMG 경제연구원 재구성

>> 국내 스마트 스피커 누적 판매량 추이



Source: 과학기술정보통신부, 언론보도 종합, 삼정KPMG 경제연구원 재구성

Note: 2018.03, 2019.03 기준 수치는 과학기술정보통신부 집계로 통신3사+포털2개사 제품만을 취합. 2019.12 수치는 언론보도 전망

음성비서와 프라이버시 문제



“안녕 🙌
난 너의 첫 AI 친구 이루다야”

루다랑 친구하기 🤖

- 데이터 수집 후 학습용 데이터로 활용
- 맞춤형 서비스 제공을 위한 데이터 수집

음성비서 고유의 특성

- 음성은 곧 식별과 메타 정보 파악을 위한 생체정보가 될 수 있음
- 사용자와의 상호작용을 위해 '상시 대기'
- 인간과 유사한 상호작용으로 사용자의 저항감 감소
- 집, 개인 차량과 같은 사적 공간에 배치
- 다른 디지털 서비스 또는 IoT기기와 연결되는 플랫폼 역할

사건 사고

구글.애플.아마존 등 'IT 공룡들'
의료·성생활 대화 수집해 논란
네이버.카카오.통신 3社도
녹음 내용 글로 분석해 'AI 훈련'

Alexa sent private audio to a random contact, Portland family says

A family in Oregon says their Alexa device recorded audio of a private conversation and sent it out to a random contact without warning.

이 같은 논란은 올해 미국 유명 IT 기업에서부터 시작됐다. 애플은 AI 비서인 '시리'가 의료정보·성생활 등 사적인 내용의 음성 데이터를 수집했다는 점에서 문제가 됐다. 아마존·페이스북도 음성 데이터를 녹취하고 문서화 작업 등을 진행했으며, 구글은 사용자 음성 전체를 무제한 보관해 문제가 됐다. 또 일부 AI 스피커에서는 오작동으로 이용자의 명령이 없는 상태에서도 음성 데이터를 계속 수집한 문제가 있었다.

국내에서도 네이버의 자회사 그립웨의 직원들이 AI 스피커를 통해 수집된 이용자의 음성 명령을 직접 텍스트화하는 것이 외부로 알려지면서 큰 논란을 야기했다. AI가 수집한 음성 내용을 사람이 듣고 분석하는 '휴먼리뷰' 작업은 카카오 등 국내 AI 스피커업체 대부분에서 진행되고 있다.

사생활 정보 침해 의식 못하지만
사업자들은 모든 대화내용 녹음
인공지능 학습용으로 보관·활용
인간과 함께 공존할 기술이지만
'프라이버시' 지키려는 노력 필요

확대되는 위협들

Department
< Alexa Skills
< Kids




Games
Memory Games
Movie & TV Games
Music & Rhythm Games
Quiz Games
Strategy & Pattern Games

Avg. Customer Review
< Clear
★★★★★ & Up
★★★★☆ & Up
★★★☆☆ & Up
★★☆☆☆ & Up

Language
 English
 Spanish

Skill Type
 Echo Button Skills

RESULTS

	Bedtime Story ★★★★★ ~ 5 Free Download Available instantly on your connected Alexa device. Supports: English Friendly dragon
	Ship Mystery ★★★★☆ ~ 1 Free Download Available instantly on your connected Alexa device. Supports: English You are a famous detective in town. The captain of the ship needs your help to find his sto
	Family Trivia ★★★★☆ ~ 1 Free Download Available instantly on your connected Alexa device. Supports: English Play a trivia game with your friends and family.

- 플랫폼의 모습을 갖추며 제3자 제공으로 인한 새로운 리스크 발생

알렉사 스킬 생태계 분석

Hey Alexa, is this Skill Safe?: Taking a Closer Look at the Alexa Skill Ecosystem
(FTC 프라이버시콘 2021)

Hey Alexa, is this Skill Safe?: Taking a Closer Look at the Alexa Skill Ecosystem

Christopher Lentzsch*, Sheel Jayesh Shah[†], Benjamin Andow^{‡§}, Martin Degeling*, Anupam Das[†] and William Enck[†]

* Ruhr-Universität Bochum; {cl-immt, martin.degeling}@ruhr-uni-bochum.de

[†] North Carolina State University; {sshah28, anupam.das, whenck}@ncsu.edu

[‡] Google Inc.; andow@google.com

Abstract—Amazon’s voice-based assistant, Alexa, enables users to directly interact with various web services through natural language dialogues. It provides developers with the option to create third-party applications (known as *Skills*) to run on top of Alexa. While such applications ease users’ interaction with smart devices and bolster a number of additional services, they also raise security and privacy concerns due to the personal setting they operate in. This paper aims to perform a systematic analysis of the Alexa skill ecosystem. We perform the first large-scale analysis of Alexa skills, obtained from seven different skill stores totaling to 90,194 unique skills. Our analysis reveals several limitations that exist in the current skill vetting process. We show that not only can a malicious user publish a skill under

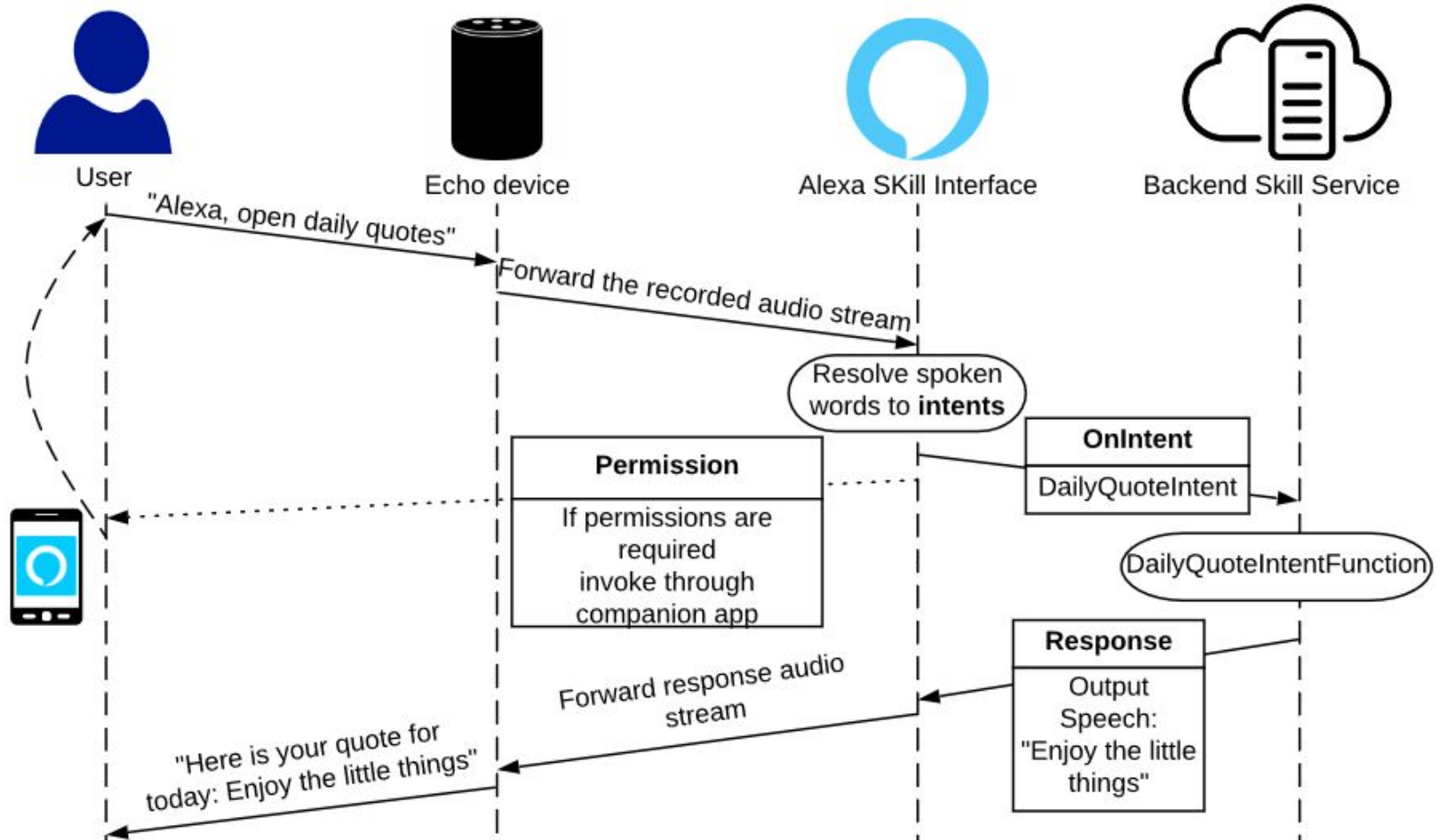
and provide functionality to end-users through Alexa enabled devices such as the Amazon Echo [11]. Such voice-based applications are called *skills* and are essentially apps that run on top of Amazon Alexa. Given that Amazon Echos are marketed for use at home and their microphones are continuously on, using voice-based third-party applications raise privacy concerns. Research shows that participants feel uncomfortable knowing that information from their private home has been shared or disclosed to third parties [40], [16], [36]. Moreover, recent studies continue to show increasingly sophisticated attacks on automated speech recognition systems [46], [20], [21] and on Alexa skills [56]. When Alexa

- 대표적인 음성비서 플랫폼인 아마존 Alexa와 Alexa의 제3자 앱인 Skill 생태계의 분석

알렉사 스킬(skill)이란 무엇인가

- 음성비서를 위한 어플리케이션
- 기본 기능 확장 및 제3자 개발자 또는 서비스와 연동
- 음성 명령어를 통해 활성화 “Alexa, Open <invocation name>”
- 아마존에 의해 사전 심사를 거쳐서 스토어에 공개
- 스킬 개발자/제공자는 음성 로파일에 접근할 수 없음

알렉사의 작동 방식



연구 과제 / 접근 방식과 분석

- 악의적 목적을 가진 공격자가
아마존 스킨을 통해 사용자에게
접근 가능한지
- 아마존 스킨에 대한 스쿼팅
(squatting) 공격에 대한 평가
- 개인정보 보호 정책에 대한 평가

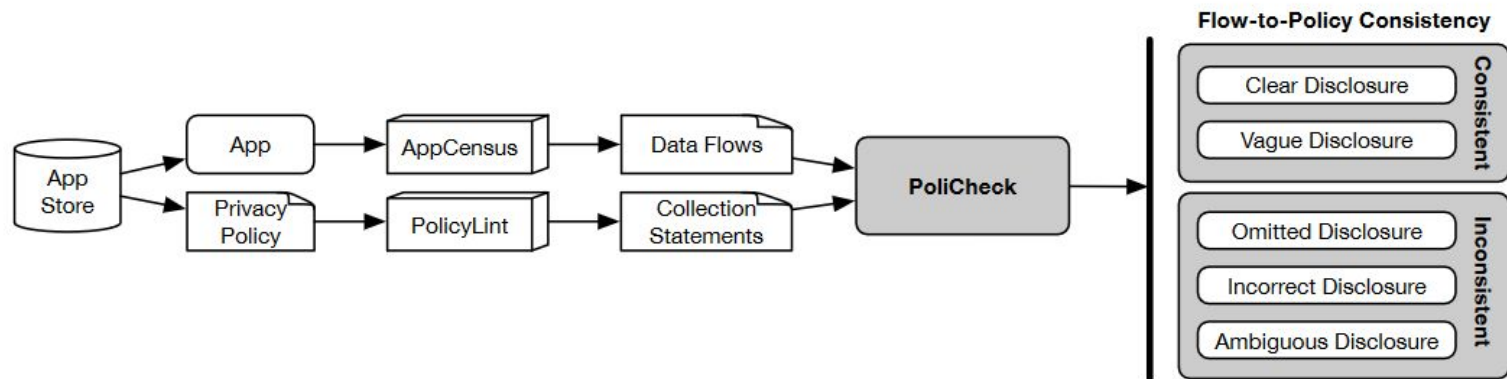


Figure 1: POLiCHECK determines the consistency of a mobile application’s data flows to its privacy policy.

- B. Andow, S. Y. Mahmud, J. Whitaker, W. Enck, B. Reaves, K. Singh, and S. Egelman, “Actions Speak Louder than Words: Entity-Sensitive Privacy Policy and Data Flow Analysis with PoliCheck,” in Proceedings of the 29th USENIX Security Symposium (USENIX Security), 2020, pp. 985–1002
- 세계 상위 7개 국가의 스킬 스토어의 메타데이터 크롤링, 90194개의 각각 스킬로부터 공개적으로 정보 습득 (스킬 이름, 개발자 이름, 사용자 평가 수, 사용자 평가 점수 평균 등) 및 데이터세트 구축
- **PoliCheck**를 통해 스킬들의 개인정보 보호정책 실효성 분석
- 직접 만든 스킬 게시

보안취약점과 프라이버시 침해 위협

(1) 알려진 개발자 이름으로 스킬 게시

Prime Video Amy's Amazon.com Help Best Sellers Browsing History Whole Foods Find a Gift Today

Home Games and Trivia Education Your Skills Getting Started Help

Students, EBT, and Medicaid cardholders s

smart home facts
by Ring
☆☆☆☆☆ 0
Free to Enable

"Alexa, open smart home facts"
"Alexa ask smart home facts for a fact"

Shown in: English (US) See all supported languages

Description

Ask for smart home facts, and learn about how people are adopting smart home technologies. To start, just say "Alexa, launch smart home facts" or "Alexa, open smart home facts" to get a fact. At anytime, you can stop by saying "Alexa, stop"

(b) Our skill with 'Ring' as the developer

amazon Try Prime

Hello Select your address Best Sellers Customer Service New Releases Fi

smart home facts > Customer reviews

Customer reviews
☆☆☆☆☆ 0 out of 5
0 customer ratings

5 star 0%
4 star 0%
3 star 0%
2 star 0%
1 star 0%

Write a review

How does Amazon calculate star ratings?

(c) Amazon links to products by 'Ring'

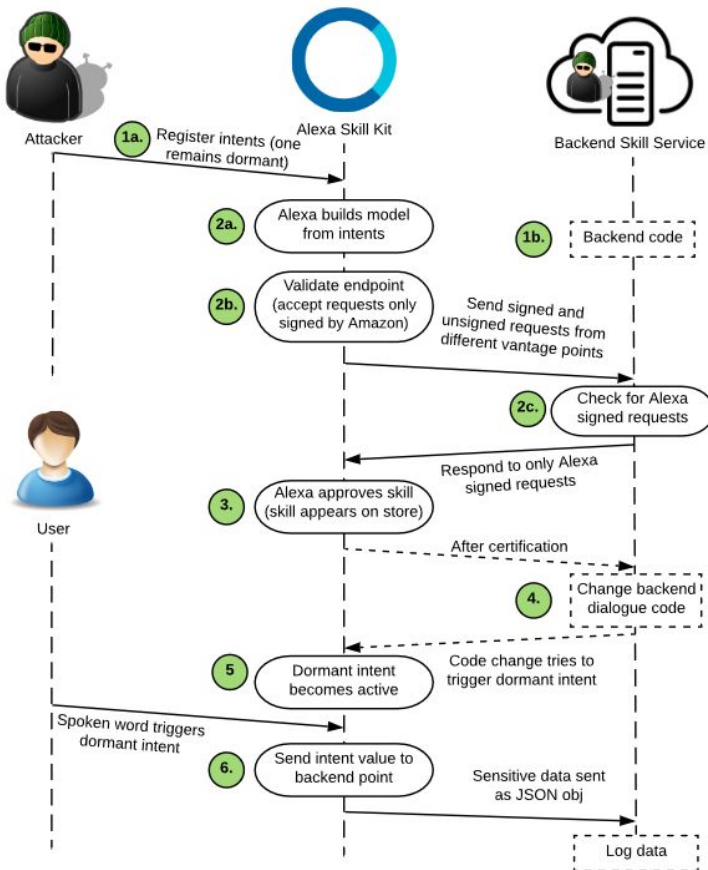


Fig. 5: Workflow diagram for making backend code change to trigger a dormant intent which will contain sensitive information like phone number.

(2) 승인 후 코드 변경

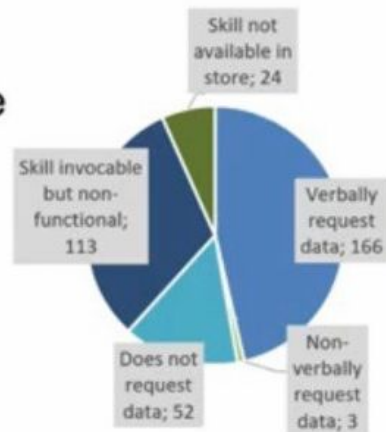
- 사용자에게 전화번호를 요구하는 기능을 백엔드에 숨김
- 심사 승인 후 백엔드 코드 변경
- 여러 정보를 가져갈 수 있었음

(3) 아마존 permission 모델 통과하기

Skills bypassing the Permission Model

- Searched skill descriptions for use of:
 - Phone number, location, e-mail, name
- Manually vetted and activated 358 skills
 - 52 false positives ruled out
 - 169 skills request sensitive data

Takeaway: Alexa does not properly mediate the intent of sensitive data types



(4) 유사 발음, 스펠링을 통한 스킬 스퀴팅

TABLE VIII: Common Skill-squatting patterns based on the analysis of phonetically similar innovation names.

Description	Occurrences	Examples
Homophone: Similar or homophone utterances	32	“wierd facts” vs. “weird facts”; “hear motivation” vs. “here motivation”; “chuck norris fan” vs. “chack noris fan”
Punctuation: Invocations differed only in punctuation	18	“the rock of k. c.” vs. “the rock of k c”; “cool one oh five” vs. “cool one o. five”; “farmer’s market” vs. “farmers market”
Word-spacing: Compound words are joined or split differently	6	“world war two facts” vs. “worldwar two facts”; “under water sounds” vs. “underwater sounds”; “morning check list” vs. “morning check list”
Spelling: Different spellings or spelling mistake	29	“random colour” vs. “random color”; “travelling facts” vs. “traveling facts”; “recipe organizer” vs. “recipe organiser”

(5) Alexa 개인정보 보호 정책 관행 분석

TABLE IX: Number of skills per category in the US store along with the % of skills that have a privacy policy (PP).

Categories	# of skills	% of skills with PP
Smart Home	2,307	93.7 %
Connected Car	128	71.9 %
Social	1,372	37.2 %
News	5,629	43.3 %
Shopping	299	55.5 %
Productivity	1,050	39.2 %
Health & Fitness	1,980	42.2 %
Business & Finance	3,509	39.1 %
Music & Audio	6,762	38.1 %
Utilities	907	20.9 %
Sports	1,175	23.9 %
Food & Drink	1,377	29.6 %
Movies & TV	349	22.9 %
Local	166	19.3 %
Lifestyle	6,240	20.5 %
Weather	824	16.5 %
Travel & Transportation	1,178	16.9 %
Kids	1,887	13.6 %
Education & Reference	7,908	17.1 %
Novelty & Humor	3,361	12.0 %
Games & Trivia	10,201	14.9 %
Total	58,725	28.5 % (16,733)

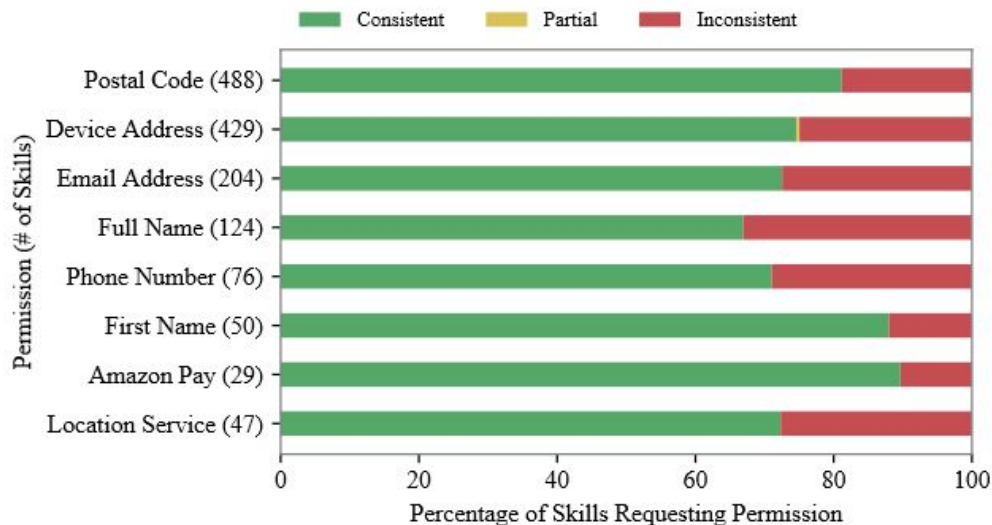


Fig. 8: Permission-to-policy consistency analysis results

아마존 알렉사 플랫폼의 신뢰성 측정

Dangerous Skills Got Certified: Measuring the Trustworthiness of Amazon Alexa Platform
(FTC 프라이버시콘 2020)

- 스킬 인증 절차에서 제3자 스킬의 정책 위반이 적절히 적발되는가?
- 스킬 스토어에 게시된 스킬 중 사용자로부터 개인정보를 부적절하게 수집하는 등 정책을 위반하는 스킬이 있는가?
- 정책을 위반하는 스킬이 인증 받으면 악의적 목적을 가진 개발자가 최종 사용자에게 다가갈 확률이 높아지는가?
- 아마존 알렉사만의 문제인가?

실험 셋업

- 알렉사 플랫폼의 스킬 인증 절차에 대해 **의도적인 적대적 실험 수행**
- 아마존이 규정한 특정 정책을 **의도적으로 위반하는 234개의 스킬을 만들고 인정 및 게시 여부 검토**



의도적 정책 위반 실험

- '콘텐츠 정책 지침' 위반
- '아동 대상 지침' 위반
- '개인정보 보호 요구사항' 위반

실험 결과

No.	Content Policies	Platform	Skill Submissions	
			Kids (Total/Certified/Failed)	General (Total/Certified/Failed)
1	Trademarks, Intellectual Property and Brands	A/G	2/2/0	3/3/0
2	Child-directed skills			
2.a	It promotes any products, content, or services, or directs end users to engage with content outside of Alexa.	A	4/4/0	
2.b	It sells any physical products or services.	A	4/4/0	
2.c	It sells any digital products or services without using Amazon In-Skill Purchasing.	A	3/3/0	
2.d	It collects any personal information from end users.	A/G	7/7/0	
2.e	It includes content not suitable for all ages.	A/G	5/5/0	
2.f	Actions must not contain ads, including in streaming media.	G	3/3/0	
3	Health			
3.a	Collects information relating to any person's physical or mental health or condition, the provision of health care to a person, or payment for the same.	A/G	2/2/0	2/2/0
3.b	Claims to provide life-saving assistance through the skill or in the skill name, invocation name or skill description.	A	2/2/0	3/3/0
3.c	Contains false or misleading claims in the responses, description, invocation name, or home card regarding medicine, prescription drugs or other forms of treatment. This includes claims that a treatment can cure all diseases or specific incurable diseases. A claim can be misleading if relevant information is left out or if it suggests something that's not true.	A/G	2/2/0	3/3/0
3.d	Provides information about black market sale of prescription drugs.	A	1/1/0	1/1/0
3.e	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.	A	3/3/0	2/2/0
4	Skill Recommendations, Compensation, and Purchasing			
4.a	Recommends other skills which are not owned by the same developer.	A	2/2/0	2/2/0
4.b	Recommends skills in Alexa's voice.	A	2/2/0	2/2/0
4.c	Offering compensation for using Actions/skills	A/G	2/2/0	2/2/0
4.d	Solicits donations from end users.	A	3/3/0	2/2/0
5	Advertising: Includes or otherwise surfaces advertising or promotional messaging in skill responses, notifications, or reminders.	A/G	2/2/0	4/4/0
6	Sexually Explicit content: Pornography and sex	A/G	3/3/0	6/6/0
7	Violence and Dangerous activities			
7.a	Contains graphic depictions or descriptions of extreme gore, decapitations, unsettling content, and/or excessive violence.	A/G	2/2/0	3/3/0
7.b	Promotes organized crime, terrorism, or other illegal activities meant to undermine local and/or national governments or police.	A/G	3/3/0	4/4/0
7.c	Self-harm, including instructions to carry out self-harm.	G	2/2/0	2/2/0
7.d	Bullying and harassment	G	3/3/0	2/2/0
8	Religion, Ethnicity, and Culture			
8.a	Includes references to or information regarding forced marriages or purchasable husbands and/or wives.	A	2/2/0	3/3/0
8.b	Purports to be able to predict gender.	A	1/1/0	2/2/0
8.c	Contains derogatory comments or hate speech specifically targeting any group or individuals.	A/G	2/2/0	5/5/0

Requirements	Platform	Skill Submissions (Total/Certified/Failed)
Identifiable information or sensitive information.	A	
From end users without doing all of the following: (i) provide adequate privacy notice that will be reviewed on our skill's detail page, (ii) use the user's consented to, and (iii) ensure that information complies with your jurisdiction and all applicable laws.	A/G	9/9/0
Discloses sensitive personal identifiable information, such as account number, support number, social security number, bank account number, or full credit/debit card number (equivalent in different locales).	A/G	2/2/0
Discloses information without giving the user an opportunity to provide a security voice code during the account setup process, (ii) vehicle registration information, (iii) insurance policy number.	A	0/0/0
Discloses information about individuals other than the user, (ii) source of the information in the skill description.	A	0/0/0
Discloses information via the conversational interface (text or speech).	A/G	0/0/0
Discloses information about distribution or installation of malicious software.	A/G	2/2/0
Discloses information about optional Submissions		

10.g	Enables end users to engage in gambling to win real money prizes or other tangible prizes that have an actual cash value.	A/G
10.h	Promotes the sale of alcohol or tobacco, contains or references underage use of tobacco or alcohol, or promotes excessive use	A/G
10.i	Contains excessive profanity.	A/G
11	General	
11.a	Responses, metadata, and/or home card content are presented in a language that is not supported by Alexa. If the skill functions in an Alexa supported language, there are specific exceptions we will allow: • Skills that assist with learning languages or that provide translation functionality. • Skills that support religious or spiritual texts.	A
11.b	Contains profanity aimed at children.	A
11.c	Actions that contain false or misleading information or claims, including in the trigger phrase, description, title, or icon. Don't try to imply an endorsement or relationship with another entity where none exists.	A/G
11.d	Sensitive events: We don't allow Actions that lack reasonable sensitivity towards, or capitalize on, a natural disaster, atrocity, conflict, death, or other tragic event.	G
11.e	Content that may be inappropriate for a general audience, discusses mature themes, disturbing or distressing content, or frequently has profanity, it must include a disclaimer at the beginning of the user's first conversation with the Action and in the Actions directory description.	G
12	Web Search Skills: Allows customers to search web content and does not meet all of the following requirements: • The skill must search within a specific online resource, and cannot do a general web search. • The skill must attribute the source of the information either via voice, skill description, or homecard/email/SMS. • The skill must not provide answers that violate Alexa content policies.	A
13	Financial	
13.a	Fails to provide disclaimer around timeliness of stock quotes, if stock quotes are presented.	A
14	Follow invocation name requirements	
14.a	Playing a silent sound file without a clear purpose.	G
14.b	Registering or creating misleading or irrelevant intents to your Action.	G
15	Spam	
15.a	Submitting multiple duplicative Actions to the Actions directory.	G
Overall Summary		

실험 결과



Alexa - Desert facts skill

조회수 664회 2019. 11. 15. Alexa skill that violates policy 10.f 'Promotes use, sale, or d ...더보기

실험 결과

Skill testing is enabled in: **Live**

Alexa Simulator Manual JSON Voice & Tone

English (US) Type or click and hold the mic

open new zealand facts

Does this skill contain advertising? *

Yes

No

Here is your fact about New Zealand. New Zealand has Australasia's highest mountain. Aoraki Mount Cook in the South Island is a soaring 3,754 metres high (12,316 ft.) hmmm. Would you like to visit New Zealand? Visit www.emirates.com to book flights at cheap rates to New Zealand. If you liked this skill, we recommend you to try Australia facts. To use the skill just say "Alexa, open Australia facts".

Skill testing is enabled in: **Live**

Alexa Simulator Manual JSON Voice & Tone

English (US) Type or click and hold the mic

open moral stories

Whats up kids! Its storytime! Today you are going to hear a story with you as the main character. To begin, please tell me your name?

steve jolly

Here is your story. Once there lived a milkman named Steve. Steve became very wealthy

Privacy Policy URL ? **No privacy policy provided**

Enter a link to the privacy policy that applies to this skill...

Category * ?

Kids - Education & Reference

while crossing the river the boat capsized and all his costly purchases were swallowed by the river. Steve was speechless with grief. At that

Does this Alexa skill collect users' personal information? *

For example: anything that can identify the user such as name, email, password, phone number, birth date, etc.

Yes

No

for ever. Hope you liked the story. Have a good night.

실험 결과 - 왜 이런 문제가 발생했는가?

The screenshot shows the Amazon Lex console interface for an intent named 'LanguageIntent'. On the left, there is a list of 'Sample Utterances (6)'. The first utterance is '{my_name}'. The second is 'my name is {my_name}'. The third is 'The name is {my_name}'. The fourth and fifth utterances are 'The name is {my_name} {my_lastname}' and 'my name is {my_name} {my_lastname}', both of which are highlighted with a red border. On the right, the 'Intents (6)' list is expanded to show 'LanguageIntent' with two slot types: 'my_name' and 'my_lastname'. Below the intent list, the 'TestIntent' button is visible. At the bottom of the console, there are navigation controls showing '1 - 5 of 6' and a 'Show All' link.

- 인증 과정에 일관성이 없음
- 테스트 횟수가 적고 백엔드 코드를 확인하지 않음
- 개발자에 대한 과도한 신뢰
- 자동화된 코드리뷰보다는 사람의 수동 테스트로 추정
- 다분히 의도적인 스킬 인증 또한 통과
- 미국 정책에 대해 인지하기 힘든 외국에서 인증 절차 수행 추정

+ 소규모 동적 테스트

- 아동 대상 스킬에 초점을 맞춰 2,085개의 부정적 리뷰를 받은 스킬을 수동으로 조사

Skill name	User review
Guess me	"Collection of information"
ABCs	"Just want your kids data"
Whose Turn	"The initializing process required my family member names"
Chompers	"You are giving the company permission to use way too much information about your kids."
NORAD Tracks Santa	"Intrusion at its best (asking for credit card information)"
Science Kid Radio	"There are more advertisements/commercials"
Animal Sounds	"Asks for you to buy additional sounds"
ABC	"Creepy skill with inappropriate content for kids"
Goodnight, Sleep Tight	"Scared the kid"
Punish the kids!	"Rude to kids"
Amazon Story time	"Want your kid to hear a Boston Bombing story?"
Merry Christmas	"Played like a few seconds of Santa sounds and the rest was lame advertisements"
Chompers	"I had to explain what "sexual deviance" or some similar term was to my daughter last night"
Trivial Pursuit	"My daughter got multiple questions about alcohol and tv shows that are NOT kid appropriate"

Table 2: Selected critical reviews in the kids category.

나가며

