

Comparing Impressions of Smart Home AI Over Time

Group 3 - Jaxon Ruiz, Soujit Ing, Shirley Delgadillo, Celeste Tan, Tien Le

Goal

Our goal for this research project was to ascertain if the role of artificial intelligence (AI) and associated technologies in our daily lives have grown or changed over time.

Research Questions

Our research questions are as follows:

- AI and technologies that utilize AI—especially smart home technologies—have become more accessible and thus more mainstream. As a result, have popular opinions regarding these technologies changed over time? If so, how?
- Have expectations as to what AI can accomplish or be used for changed due to this increased exposure?
- Do any changes to these popular opinions signify a growing or changing role of AI technology in our everyday lives?

Protocol

Data set specifications

We collected a total of 400 consumer online reviews for the Amazon Alexa App and Amazon Echo, collecting 200 reviews for each product. As Amazon Alexa was one of the first AI systems to be produced specifically for integration into the home and daily routines, we believed that looking at these two products would serve our research goal well. In order to gauge any changes in opinion due to the increased accessibility of these technologies, we collected reviews within a year of the initial release of the first iterations of these products (2015-2016)¹, and the current year (2022-2023), creating time-based data sets of 100 reviews each.

¹ Exception - six of the Amazon Alexa App 2015-2016 1-star reviews are from early 2017, as there were not enough from the time period in order to fill the data set

Of these time-based data sets, we collected two subsets based on the number of “stars” or rating attached to each review—a “high” set, consisting of four- and five-star reviews, and a “low” set, consisting of one-, two-, and three-star reviews. The resulting sets, which are an equal 50 reviews each, had a specific number of reviews per star rating, with the high sets containing 25 four- and five-star reviews each, while the low sets contain 16 three-star reviews and 17 one-star and two-star reviews each. We collected the data in this manner in order to eliminate bias that may result from uneven rating distribution, which would have naturally skewed our analysis to contain more positive or negative opinions. Additionally, this would allow us to fairly compare the high set to the low set in our analysis, potentially highlighting interesting differences between the expectations or use cases that are common within those sets.

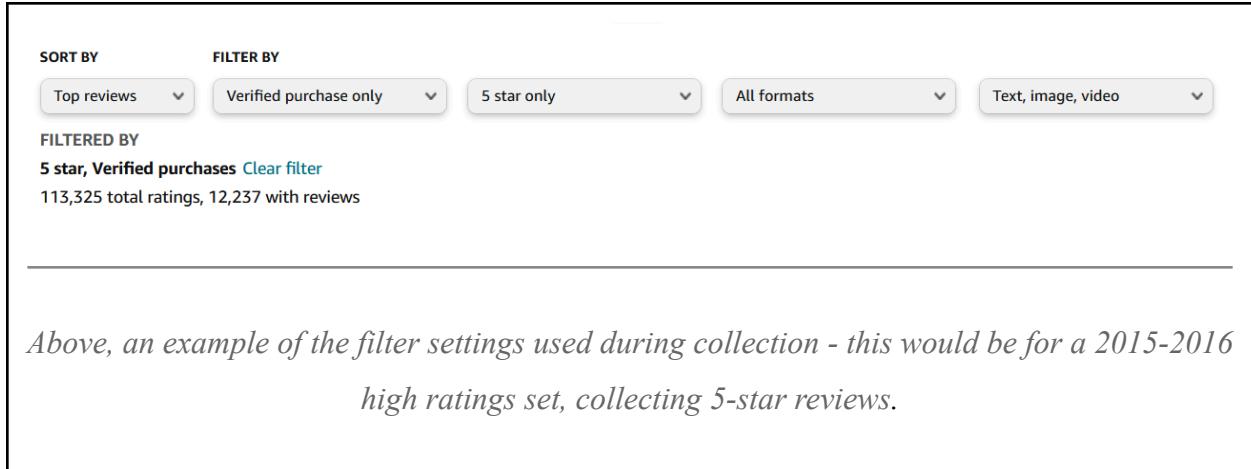
Method of collection

To create our data sets, we hand-harvested our reviews from the corresponding Amazon review pages. For the Amazon Alexa App review sets, we simply harvested them from [the Amazon Appstore page for the Alexa App](#), which has remained the same from release to now. For the Amazon Echo review sets, we harvested the reviews from two different pages; for the 2015-2016 data set, we harvested from [the Amazon page for the 1st generation Echo](#), which would naturally contain the most reviews from the initial conceptual debut of Alexa. However, as this product is no longer available for sale and the specific hardware and model that users interact with now has changed, we harvested the 2022-2023 set from [the Amazon page for the 4th generation Echo](#). As both generations were developed from the same core Echo concept—a speaker that serves as an interface and hub for interacting with Alexa and its services—we felt it was fair to collect the data in this way.

In order to collect our data, we utilized the filters Amazon offers. The filter settings we used for the data sets are as follows:

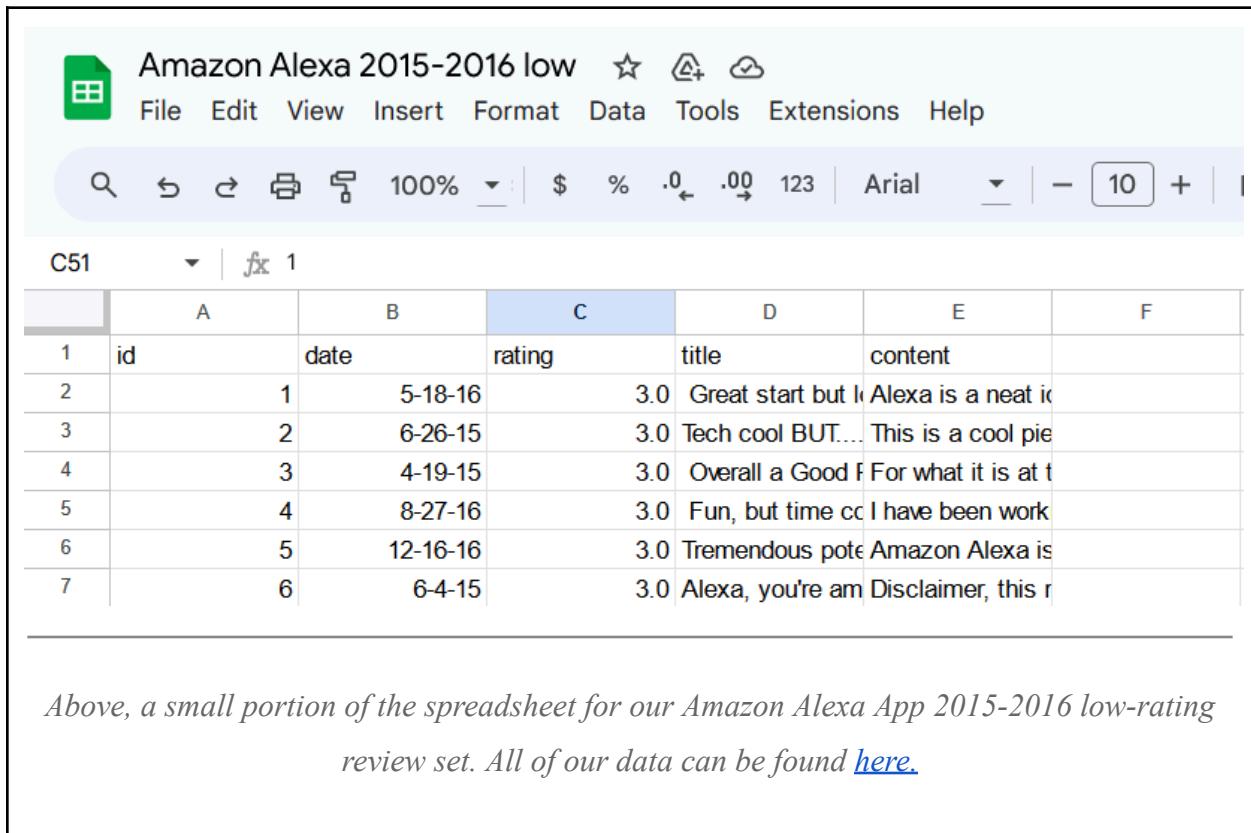
- Filter by “Verified purchase only” and the appropriate star rating (ex. “5 star only”, “2 star only”)
- 2015-2016 sets:
 - Sort by “Top reviews”

- 2022-2023 sets:
 - Sort by “Most recent”
 - Else, defaults (“All formats”, “Text, image, video”)



Above, an example of the filter settings used during collection - this would be for a 2015-2016 high ratings set, collecting 5-star reviews.

The reviews from the appropriate time range were then collected, with the date, star rating, title, and text content of the review recorded into a spreadsheet.



Above, a small portion of the spreadsheet for our Amazon Alexa App 2015-2016 low-rating review set. All of our data can be found [here](#).

	A	B	C	D	E	F
1	id	date	rating	title	content	
2		1	5-18-16	3.0	Great start but I	Alexa is a neat i
3		2	6-26-15	3.0	Tech cool BUT...	This is a cool pie
4		3	4-19-15	3.0	Overall a Good F	For what it is at t
5		4	8-27-16	3.0	Fun, but time cc	I have been work
6		5	12-16-16	3.0	Tremendous pote	Amazon Alexa is
7		6	6-4-15	3.0	Alexa, you're am	Disclaimer, this r

Potential bias

As reviews are voluntarily posted and are naturally limited to the population of those who have bought the product, there is inherently some bias in our data set due to the very data itself. However, while our data may not necessarily represent the general population at either of the time ranges we examined, we believe our data set still serves as a useful cross-section for exploring a facet of popular opinion and addressing our questions regarding AI and daily life integration.

Due to our harvesting method, the order of exposure of each review was determined by Amazon, meaning for our low-ratings sets, the data set we created was influenced by factors Amazon decided makes a “Top Review.” While we had tried to combat that early on by exploring scrapers that would allow us to randomize our selection, it required way more temporal and/or monetary resources than we had to circumvent Amazon’s anti-scraping measures. If we had greater resources, we would have been able to more effectively address this bias.

Analysis and Results

We analyzed our data from a variety of angles, most notably common issues consumers encountered, intended use case, the personification of Alexa, and the reviewer’s general opinion expressed in each review. We tagged each review with a number of codes that we developed based on common themes we noticed while we processed the data.

Codes

Opinion

Represents certain notions and overall sentiment expressed by reviewers

- optimistic in potential: noted issues, but expressed optimism in potential
- disappointment: expressed some sort of failure to meet expectations
- useful: noted productivity boost with product
- very positive: spoke positively of and highly recommended the product

- affectionate: expressed personal affection towards the product

Positive features

Specific features reviewers enjoyed

- music integration: good integration of music-related features, both hardware (speaker quality) and software (intuitive and smooth interface with applications such as Pandora or Prime Music)
- convenience: features that make life generally more convenient
- integration with other devices: good integration with other devices, especially in a smart home context
- useful information: ability to provide information such as answers to spoken questions, the news, weather, ambient temperature, etc.
- organizational tools: features such as alarms, calendar management, and to-do lists that aid in organizing one's time
- accessibility: features that aid those with disabilities or are otherwise disadvantaged in daily life

Issues

Specific issues users encountered

- audio problems: audio output issues, such as distortion, stereo pairing issues
- difficult setup: substantial troubleshooting required for setup tasks such as connecting to Wi-Fi or other devices
- expensive price: high price of device itself or accompanying devices necessary for integration
- privacy: concerns about privacy
- requires or relies on app: product relies on app in a way that is detrimental to user experience, ex. by providing a technological knowledge barrier, by undermining the hands-off aspects
- poor accessibility: features were difficult for people with disabilities to navigate

- limited ‘knowledge’: lacks the ability to effectively answer spoken questions, respond to more specific requests for information (ex. local versus global news)
- limited features
- limited or poor interactivity with other products: interactivity with other products is poorly integrated or completely lacking; products can be hardware (such as in smart home contexts) or software (such as in integration with services such as Pandora or Prime Music)
- inconsistency: product performs inconsistently

Daily life integration

Measures of how products were integrated into reviewer’s daily lives

- noted reliance: mention reliance on product for some daily function, such as playing music while cooking, setting alarms, making hands-free purchases

Primary intended usage

Use cases that reviewers indicated they purchased the product for

- accessibility aid: as an aid those with disabilities
- companion: as a friend or human-like conversational partner
- entertainment: for entertainment purposes, such as playing music
- informational: for retrieving information such as news, weather, traffic, or other assorted queries
- overall personal assistant: as an assistant handling scheduling, minute-to-minute time management, calls, etc.
- smart home technology: as a hub or for otherwise interfacing with other smart home products

Personification

Reviewer uses language that personifies Alexa, such as using she/her vs. it/its

- Personified product

Language processing

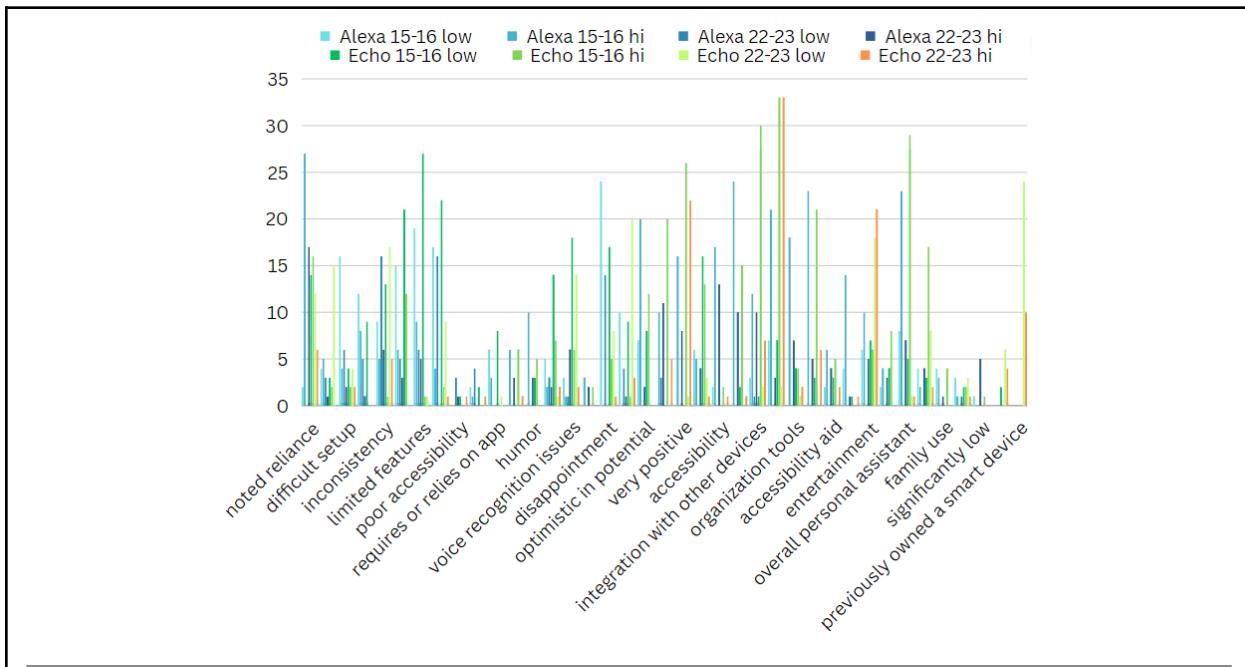
Traits that pertain to AI language processing and speech

- conversational: success at having a passable or enjoyable conversation with a person
- syntax mistakes: noted failures to understand syntax
- voice recognition issues: noted failures to hear or recognize voices
- humor: notes light humor in product

Reviewer tech experience

Represents relative technological experience of the reviewer

- notably high: high familiarity or expertise with technology, ex. college degree
- significantly low: tech experience questionable or expressed to be low
- family use: reviewer mentions family or communal use of device



Based on qualitative analysis and code tagging, we found the most common codes among different sets. As reviews were split into eight subsets split by time period, product, and rating (thus, roughly positive or critical sentiment), we could directly compare them across a couple areas of difference.

Comparing High and Low Sets

Alexa App

Understandably, there were few positive codes applied to the critical reviews, which of course focused more on the device's downfalls, though positive reviews seemed more likely to acknowledge some faults in the product, albeit with far less severity. If weighting was applied over the intensity of some codes, there would likely be a heavier divide. Though both sets of reviews would often mention the music integration with Pandora and Amazon Prime Music, some inconsistencies in our code application meant that these themes were mainly coded in the positive reviews, since "music integration" was under positive features. From critical reviews, the most common code was for "limited features", reflecting the more barebones features at release. On the other hand, the positive reviews were overwhelmingly satisfied with the included question and music features. However, at the same time, many positive reviews would express optimism for future updates and features.

Amazon Echo

Much like the Alexa App set, the Echo high positive sets and the low critical sets differ in positive code application, with the low sets having 85% (155 vs 22) less "positive features" tags overall, and as well as mostly being applied with "opinion" codes such as "disappointment" and "frustration"; the reverse is obviously true for the high sets.

Among the critical sets, the most common codes were "limited or poor interactivity with other products", at 31 code applications. Strangely, this is a sharp contrast to the most common codes applied to the positive sets, which were "music integration" (66) and "integration with other devices" (37). As these codes all deal with the same underlying mechanisms—the hardware and software interactions between different devices and services, whether the Echo provides an

intuitive interface, and whether these interactions help fulfill the author’s use case—this overlap could be evidence that what primarily differentiates these two sets and thus what makes an AI system desirable for integration into daily life isn’t necessarily the capabilities of the machine itself, but how well they meet our expectations as to what AI is supposed to accomplish.

Comparing 2015-2016 and 2022-2023 Sets

Alexa App

While both sets seemed relatively similar, the largest difference was seen where the 2015-2016 reviews emphasized being interested in the potential, rating highly out of hope, while the 2022-2023 reviews seemed more content in the product’s current state. Notably, the 2022-2023 reviews seemed to generally lack as much substance as the 2015-2016, which mentioned the entertainment and humor of Alexa significantly more. This seems to indicate more energy and interest in the initially newer product, which became less significant as it became more popular. However, the lesser substance in the 2022-2023 reviews could be because of the review sorting algorithm, as higher quality old reviews had more time to be rated ‘more helpful’ on Amazon, potentially filtering out less informative reviews, which were more prevalent in the 2022-2023 set. At least in the reviews collected to represent positive opinion of Alexa in 2022-2023, many had far less to say, reflecting in less concrete opinions and ideas found when categorizing data.

Amazon Echo

The greatest differences between the Amazon Echo 2015-2016 sets and the 2022-2023 sets are found in the number of issues reported, the intended use of purchased products, and the number of reviewers who personified Alexa.

The number of issues reported went down 56%, from 130 applications of issue-related codes in both Echo 2015-2016 sets to 57 total applications in the 2022-2023 sets. The most-reported issue in the 2015-2016 set was “limited ‘knowledge’”, with 33 total applications, while in 2022-2023 the most-reported issue was “inconsistency”.

The application of codes related to intended use cases also changed significantly between the two time-based sets. In the 2015-2016 set, the top three use cases were “overall personal assistant” (34), “smart home technology” (20), and “entertainment” (13). In the 2022-2023 set, the top three use cases were “entertainment” (39), “smart home technology” (10), and a tie with “overall personal assistant”/“accessibility aid” (2). This change is manifested not only in the change of order, with “entertainment” taking first place, but also the distribution of these applications. While the 2015-2016 set is more evenly distributed, with “overall personal assistant” only having a 14- and 24-point lead over second and third place, the 2022-2023 set has a much tighter spread, with “entertainment” being applied nearly four times as much as second place and almost 20 times as much as third place.

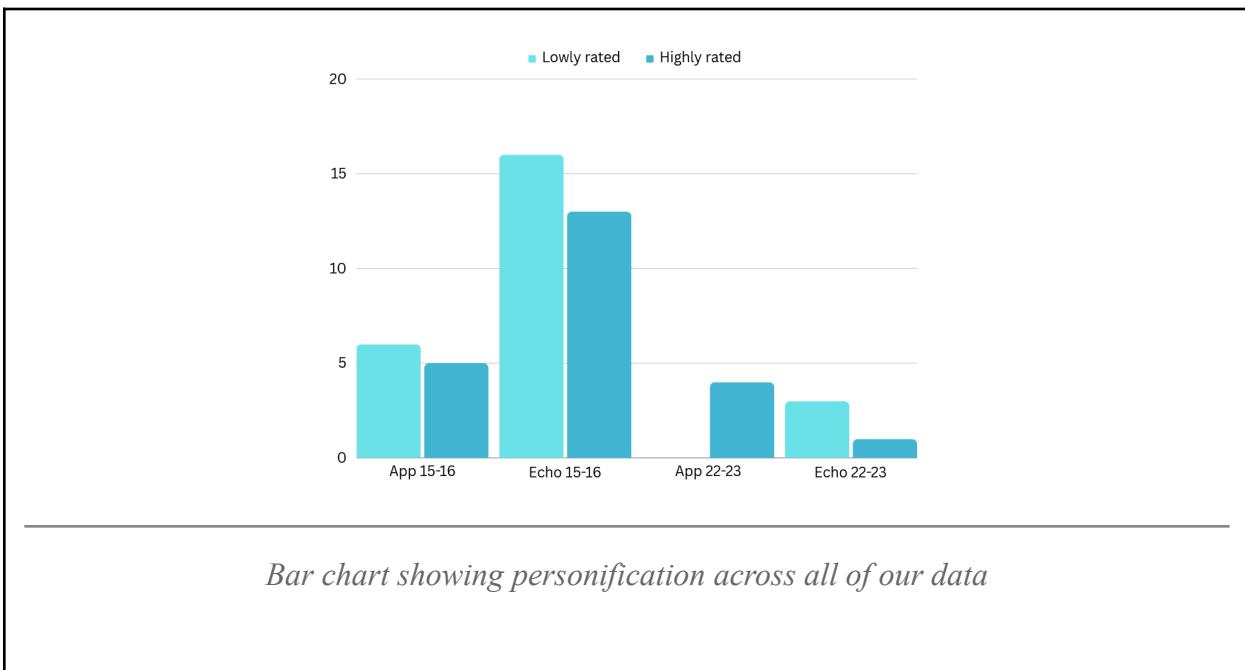
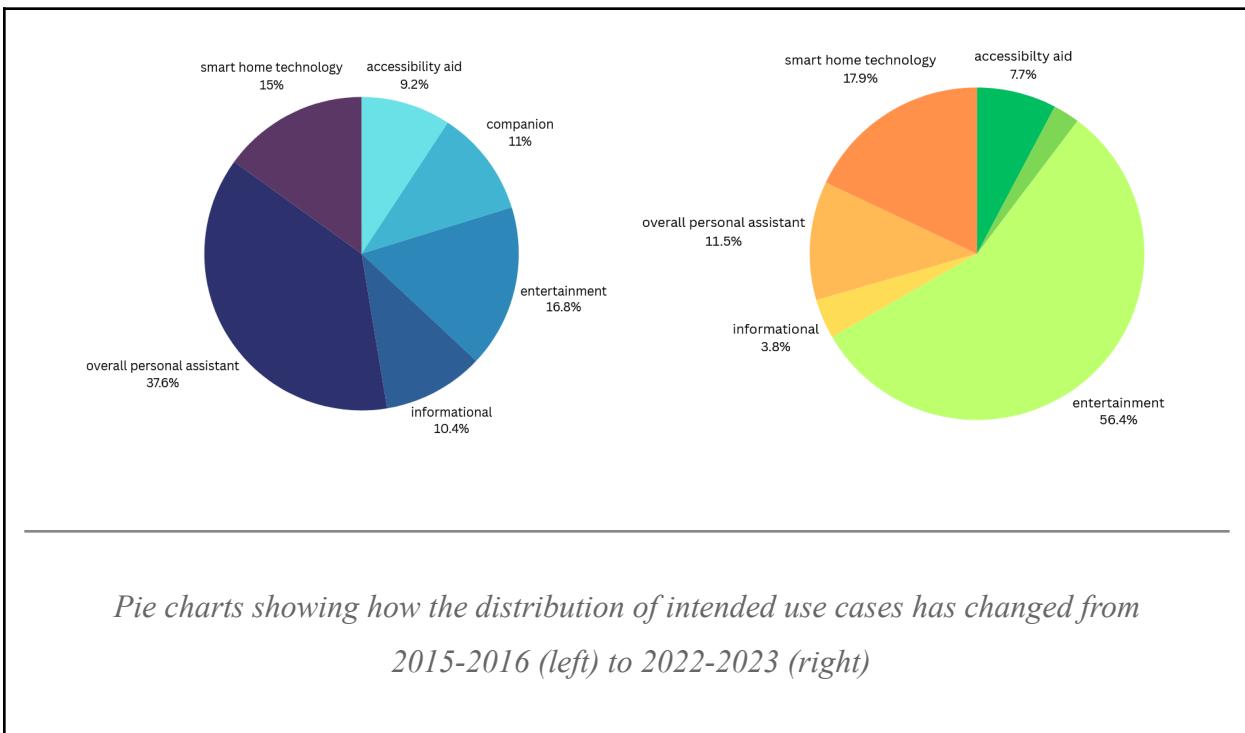
Lastly, the number of reviewers who personified went down dramatically, decreasing by 86%, from 29 code applications to 4 total across both sets.

While the Echo data sets face the same substance issues as the Alexa App sets, with 2022-2023 being shorter and far less in depth, we can still make some meaningful observations by comparing the distributions within each set. On its own, the shift in the issues code applications from “limited ‘knowledge’” to “inconsistency” can be attributed to software improvements developed over time to address these issues, when combined with the major change in use cases to be very “entertainment”-focused, the evidence can perhaps be taken to show a change in the expected role of the Echo over time. Rather than being fixed, this dip in reported issues with “limited ‘knowledge’” may instead be a sign that the Echo is no longer expected to be able to carry out information-related-tasks to consumers; this may be due to the more complicated syntax processing and soft human-like logics that these tasks require.

From Teammate to Tool

As discussed above, our analysis of these Alexa-base technology reviews revealed a significant difference over time in the personification of Alexa and the intended use case consumers had in mind for Alexa. In examining both the Echo sets and the Alexa sets together over time, a clear decrease in personification and a clear shift in use cases from the 2015-2016 time period to the present day. Intended use cases that feature complex “thinking” or more complicated syntactic processing and soft logics such as “overall personal assistant” (-86%),

“informational” (-83%) and “companion” (-89%) decreased dramatically. Instead, use cases that featured communication between machines or software services, as well as simpler, clearer syntax (ex. “Alexa, play...”) such as “entertainment” (+52%) and “smart home technology“ (code applications decreased, but the 2022-2023 sets showed a proportional increase).



Personification, which serves as an indicator of both how human reviewers believe she is as well as how human they expect her to be, decreased 88%, from 40 code applications to eight.

These changes can be interpreted to mean a change in the role Alexa is expected to take in people's daily lives, seemingly exhibiting that she is expected to be less of an equal—a teammate—and more of a tool that acts as an extension of us.

Applying this to AI design philosophies, this result supports HCAI proponent Ben Schneiderman's proposition that AI's use for its properties as a machine and an extension for human ability versus a human-like replacement, and even implies it is simply the natural form AI takes as humans use it, optimize it, and make it an ever increasing presence in our daily lives.

Conclusions

To conclude, we found through our research that there is evidence to suggest that popular opinion has changed over time since the conceptual debut of Alexa and artificially intelligent smart home assistants in general. Additionally, our analysis reveals some patterns that may indicate that the role we expect and wish AI to take in our daily lives has changed over time, naturally growing into one that casts them as supertools, rather than superhumans.