What wake words do

James EK Parker, Melbourne Law School

https://orcid.org/0000-0002-9550-5655

ABSTRACT

This essay is about the wake word, as a cultural, economic, and legal technique. It begins, in Part 1, with the wake word's history, starting in the 1970s with the first work on automatic 'word spotting', tracing its early commercial applications in telecoms through to the launch of the Amazon Echo in 2014, and the political context surrounding it. Against the backdrop of the Snowden revelations, and mounting concerns about surveillance capitalism, the wake word was a technology of consent. It played a crucial role in the rapid social acceptance of smart speakers, the normalisation of voice-computing, and the authorisation of machine listening more generally. The consent promised by the wake word has always been a fiction, however, undercut by its actual operation at every turn. Part 2 describes how. We should think of the wake word, I argue, as a technique of expository power, appealing to liberal traditions of privacy and consent only to dramatically undermine them, in favour of an alternative vision of the data subject as a resource for diverse forms of value extraction. Part 3, finally, is about the wake word's future. I identify two countervailing tendencies: towards wake word proliferation on the one hand, and wakewordlessness on the other. At the same time as wake words are being normalised, and more and more companies realise their potential as a new form of sonic branding, we are also seeing the wake word's erosion in the name of more 'natural' interactions with our devices, and as part of a general trajectory towards more automated, pre-emptive, and 'environmental' form of governance and control.

KEYWORDS

Wake word, interface, voice assistant, smart speaker, big data, machine listening.

DISCLOSURE STATEMENT

The author reports that there are no competing interests to declare.

BIOGRAPHICAL NOTE

James Parker is an Associate Professor at Melbourne Law School, who works across legal scholarship, art criticism, curation, and production. He is an Associate Investigator with ADM+S, former visiting fellow at the Program for Science, Technology and Society at the Harvard Kennedy School for Government, and sits on the advisory board of Earshot, an NGO specialising in audio forensics. He would like to thank Zoe Reid for her research assistance, Kathy Reid for the conversations, and Joel Stern and Sean Dockray for the friendship and ongoing collaboration.

WORD COUNT

9,984 words

INTRODUCTION

The first reference to the 'wake word' I can find is in a patent from 2012, filed by a company called Rawles LLC (Salvador et al. 2016). As computing devices evolve, the patent explains, how we interact with them is changing too. Keyboards and mice have been joined by touch screens, motion, gesture, and increasingly voice. But 'natural language' interfaces present certain challenges, the patent says. If a device 'hears' the question 'what is the time?' how is it to know 'whether the user is talking to the device, or to another person in the room'? How to distinguish between what the patent calls 'general speech' and a 'command to be executed by the computing device'? Sometimes users will be able to register their intent by 'clicking on a window', or 'pressing an icon on a touch screen'. But not in the situation being imagined here. The patent depicts a figure, hands in pockets, addressing a 'voice controlled device' on a table. They are trying to buy tickets for a movie, and the device is connected to the cloud, which is shown hovering above. To make such a device a reality, the text continues, 'this disclosure provides techniques and systems for utilizing a natural language input… to trigger or enable a natural language command recognition functionality of a natural language controlled device.' The patent calls this trigger the 'wake word'.

It's probably not right to call this the invention of the wake word. From a technical perspective, wake word detection is an example of 'word spotting', which computer scientists have been working on since at least the 1970s. Characters in speculative fiction were invoking robots and computers by name long before that. But US Patent No. 9,275,637 B1 is an important document in the wake word's history nevertheless. Because Rawles LLC turns out to have been a shell company for Amazon, the device in the patent became the Amazon Echo, and the wake word they eventually settled on was 'Alexa'.

Alexa-the-wake-word and Alexa-the-digital-voice-assistant were launched at the same time as the Echo smart speaker in November 2014, before a full US release in July 2015 and

progressive international roll out after that. You can still find the original promotional film on YouTube (Amazon 2014). A package arrives at a doorstep somewhere in suburban America. 'What is it?' a girl wonders, as her father unboxes a black cylinder and places it on the living room table. 'It's called Amazon Echo', he explains. 'Is it on?' she asks. 'Oh, it's always on,' he replies, without irony. 'Can it hear me right now?' 'No', he says. 'It only hears you when you use the wake word we chose. Alexa.'

This essay is about the wake word. It's about what wake words are, what they do, and why they matter. My starting point is that the wake word has been one of the more important cultural, economic, and legal techniques (Vismann 2013) of surveillance capitalism (Zuboff 2018) over the last twenty years, despite going largely unremarked upon. Not only was it crucial to the initial take-up and rapid social acceptance of smart speakers following the Echo's launch, the wake word also played a vital role in the normalisation of voice operation and the authorisation of machine listening more generally (Parker and Dockray 2023). Alexa wasn't the first voice assistant, or even the first popular one. Siri had been available as a standalone app since 2010, before being incorporated into iOS5 in 2011. Google Now launched for Android in 2012. But neither was a core part of the user experience. Even when Apple announced the 'Hey Siri' command at its Worldwide Developers Conference in June 2014, it was only as an afterthought, over an hour into proceedings, as a helpful new way of invoking Siri while charging your phone. It was the Echo that introduced the public to truly 'voice-first' computing. It was the Echo that taught us to expect our devices to listen. And there would be no Echo without the wake word.

For two years after its launch, Amazon had no direct competition. It was only at the end of 2016, with 7 million Echo devices sold (Kinsella 2016), that Google finally unveiled a speaker of its own. In 2017, so did Microsoft and Alibaba, with Apple following early the year after. By September 2018, around a quarter of US adults – some 58 million – had at least

one smart speaker in their home (Voicebot 2018), making them 'one of the fastest-adopted new devices in history' (Deloitte Insights 2019). By this point, and on the back of the smart speaker's enormous success, voice assistants were being integrated into all manner of devices. It's not just our speakers and phones that are listening anymore. It's our laptops, cars, watches, tablets, televisions, earbuds, thermostats, fridges, and lightbulbs too. From a user base of zero in 2015, Alexa is now available on over 600 million devices worldwide (Amazon 2025a). Together, the likes of Google Assistant, Siri, Cortana, Bixby and, more recently, Meta AI and ChatGPT, can be accessed on billions more.

The essay is in three parts. Part 1 is about the wake word's history. I begin with Jeff Bezos' own version of events, about wanting to build the computer from *Star Trek*, before rewinding to the 1970s to begin again with a less self-serving account. Computer scientists were already experimenting with what they called 'trigger words' twenty years before Alexa, and other forms of targeted 'word spotting' were being developed in the telecoms industry well before that. Amazon didn't invent the wake word, it mainstreamed it. In doing so, it also provided a crucial metaphor for understanding the politics of the voice user interface, along with the forms of data extraction it introduced and extended. Metaphors matter. The Echo arrived barely eighteen months after the Snowden revelations, detailing mass surveillance by the US government and its international and corporate partners. In this political climate, convincing the public to buy and install listening devices through their homes depended on the smart speaker's construction as relatively passive or benign: always on but mostly sleeping; awoken only with the user's active consent.

This story turns out to have been a useful fiction, undercut by the wake word's actual operation at every turn. What is the wake word? An interface, or meta-interface: which is to say, with Alexander Galloway (2012), not so much a thing as an effect. Likewise for Cornelia Vismann (2013, 83), the question is always 'what media do, what they produce, and what

kinds of actions they prompt'. What do wake words do? How do they work and fail to work? What do they authorise, promise, or make possible? What kind of subject do they produce? These are the kinds of question that interest me in Part 2. As a form of vocal discipline, the wake word assimilates and alienates non-normative speakers. But it is also a manufacturing technique: a way of initiating data relations, and a form of expository power that appeals to liberal traditions of privacy and consent only to dramatically undermine them, in favour of an alternative vision of the data subject as a resource for diverse forms of value extraction.

Part 3 is about the wake word's future. I identify two countervailing tendencies. On the hand, wake words will tend to proliferate and diversify, as voice assistants are made an increasingly normal part of computing and companies start to realise the wake word's potential as a form of sonic branding. On the other hand, there is a dialectical movement towards the wake word's erosion, in the name of more 'natural' interactions with our devices, and as it's gradually replaced with persistent ambient monitoring, as part of a general trajectory towards more automated, pre-emptive, and 'environmental' form of governance and control.

PART 1

Building the computer from Star Trek

Where did the wake word come from? The story Amazon CEO Jeff Bezos likes to tell is that he wanted to build the computer from *Star Trek* (Weise 2016) and had pitched the idea of a 'voice-first' interface to his technical adviser Greg Hart in early 2011 (Vlahos 2019). This device, he thought, would be activated with the word 'Amazon' (Brustein 2016). In Bezos' whiteboard sketch from the meeting, a photo of which was later published in *Wired*, there is a hastily scribbled cube with a speaker, microphone, on-device digital signal processing, a WiFi connection, and a 'kill' switch. Next to this it says 'INVENTION' (Stone 2021).

Star Trek is certainly part of the wake word's story. Technology often follows science fiction, and Bezos talks regularly about being a fan of the show, including how it also influenced his ambitions for space travel and mining (Vlahos 2019). When it comes to cultural depictions of voice computing, however, we can be more specific. *Star Trek* is notable in the sci-fi canon for including something like a wake word and the dialogic forms it entails. In *2001: A Space Odyssey* (1968), for instance, HAL engages in long, free-flowing conversations which it often initiates, like a human character would. In *Star Wars* (1977-) Luke and Leia talk to their droids more like friends than machines or servants. But in *Star Trek*, especially from *The Next Generation* (1987-1994) on, the USS Enterprise's operating system is invariably invoked with the word 'computer', before being issued a command. 'Computer, locate Lieutenant Commander Data.' 'Computer, estimate the time from this location to star base 416 at warp 9.' 'Computer, dim lights.' (Felker 2020a, 2020b)

When the Echo launched in 2014, *Star Trek* become a key part of its reception. 'Amazon built the Star Trek computer for your house' read one early headline at CNN (King 2015). In 2017, when 'computer' joined 'Alexa', 'Amazon', and 'Echo' on the list of official wake words, The Verge (Bohn 2017) was quick to thank the company for 'giving anybody who's ever watched *Star Trek*' what they wanted. This is a rather idealised way of understanding the Echo's invention, however. For Bezos, it is also self-serving. Not only does it make invention the kind of thing a CEO can do on a whiteboard after watching TV, it says nothing about the timing, or the technical, commercial, political, and legal conditions out of which this model of voice computing emerged, and by which it was also shaped.

Word spotting

From a technical perspective, wake word detection is an example of 'word spotting': the problem of locating a specific word in a continuous stream of sound and speech. This problem

had already been identified by researchers in the 1970s, partly because it was such a defined and manageable task. Word spotting, Christiansen and Rushforth (1977) explained, was 'much simpler than the general problem of speech understanding'. As a result, 'a satisfactory solution' might be possible 'using a simpler system.' Christiansen and Rushforth's own efforts, using 'dynamic programming' techniques for 'template matching', were not successful by contemporary standards and only worked for the person whose speech it had been trained on anyway. It was a different, statistically based method that would see the first breakthroughs, and word spotting's first commercial applications at the start of the 1990s. This method used the hidden Markov models first developed at IBM's Continuous Speech Recognition group across the 70s and applied it to word spotting for the purposes of automating call routing, first at AT&T Bell, then across the telephone industry (Li 2017).

In one key paper, Lawrence Rabiner and his colleagues at Bell describe a 'large scale trial' in which callers were invited to choose between five options – 'collect', 'calling-card', 'person-to-person', 'bill-to-third', and 'operator' – which they system would, in principle, recognise automatically. As it turned out, however, many callers would say things like 'I want to make a *collect* call please', rather than just 'collect'. This was a problem because 'most conventional isolated word recognition algorithms have not been designed to recognize vocabulary items embedded in various carrier sentences' (Wilpon et al. 1989, 254). So the paper describes an 'HMM-based recognition algorithm' to 'attack the problem of recognizing a small set of vocabulary words in fluent speech'.

This word spotting algorithm would go on to be implemented in Bell's flagship Voice Recognition Call Processing system, launched in 1992 and used to justify the axing of some 6,000 operators at the company. Similar systems and job cuts followed around the world. In technical writing on word spotting from later in the 90s (Rohlicek et al. 1989; Rose 1996), call routing is the paradigmatic use case, though it was quickly joined by other tasks like

'message browsing' for voicemail systems and, crucially for the history of the wake word, 'command spotting' in so-called 'open microphone systems.' In such systems, Jan Rohlicek explained (1995, 124), 'the recognizer is always listening, waiting for the user to say a known utterance of a specific "trigger word" such as "Computer, turn on the lights."' 'Extraneous speech not preceded by the word "Computer" must be ignored.' Clearly, Bezos wasn't the only one watching *Star Trek*.

Though it tends to be forgotten now, this kind of 'open mic' system was commercially available years before automatic dictation and transcription software; word spotting, remember, being the simpler task. The first dictation tool for home computing, Dragon Systems' DARPA funded Naturally Speaking™, was released in 1997. But Apple had launched Casper, a rudimentary voice assistant for desktops, already in 1992 (Engst 1992). In a demo on Good Morning America, the host leans towards a monitor on a desk and says 'Casper, good morning.' (Roemelle 2017) A synthesised voice replies 'Good morning America! And good morning, Joan', before one of the researchers behind the software takes over to demonstrate the technology's possibilities for office work and e-commerce. No-one asks about privacy. 'Casper, copy this text', Lee says. 'Casper, paste'. 'Casper, open checking.' 'Casper, pay Macey's $65.' Casper never really made good on this promise, and for years Siri ran without an equivalent oral trigger. It was only in 2014 that Siri would finally become handsfree, as Casper had already been twenty years earlier.

Wake word recognition

In Rawles LLC's original patent, the techniques used for wake word recognition are never fully described. Instead, the patent offers hidden Markov models and Gaussian mixture models as possibilities, before moving on to describe a system for what it calls 'wake word evaluation' (Salvador et al. 2016). The patent describes a series of modules which would

automatically assess a 'candidate wake word' for its overall suitability in a given language. Candidate wake words should be evaluated according to their 'ease of pronunciation'; the frequency with each they appear in the target language; how similar they are to other words, especially words that are 'taboo'; as well as how easy they are to recognise computationally, and how often they give rise to a 'false alarm' in the 'wake word recognition module'. These are precisely the sorts of factors that led a team of 'dissident' engineers to confront Bezos about his preferred wake word ('Amazon') just weeks before the Echo's launch. The project was 'hurtling toward a potential disaster', they said, because the speakers would wake every time they heard an ad for the company on television and start 'buying random stuff from the Internet' (Brustein 2016). Hence their alternative 'Alexa': not an especially common name at the time, and with its hard 'x' before the final vowel.

The technical side of wake word recognition has evolved considerably since Amazon's first patent. One crucial set of developments has been in 'far field' recognition, which has dramatically improved the ability of ASR systems to operate at distance and in noisy environments. From a computing perspective, it's one thing to spot a word spoken directly into a telephone handset, another to pick it out from across a room with music playing in the background. Far field technologies address this problem by using an array of microphones – the original Echo had seven – to estimate the user's location, so that other signal processing techniques (dereverberation, source separation, acoustic beamforming) can be applied to their speech accordingly (Chhetri et al. 2018). Deep learning has led to major improvements, here and across the board. Word spotting can now be done using deep neural networks (Ku et al. 2024), hybrid systems combining hidden Markov models with neural nets (J. Wang et al. 2020), and transformers (Y. Wang et al. 2021), on both the front and back ends. In every case,

the quality of the dataset is just as important as the model being used, which is one reason Amazon insists on collecting so much data (Koenecke et al. 2020).[1]

As with automatic call routing before, accuracy and efficiency seem to be the main drivers of innovation. Whereas some new systems offer fewer 'detection errors' and 'false alarms' (Liu et al. 2020), others promise to be smaller, less resource intensive, with lower computational demands and faster response times as a result. The trend in wake word detection as elsewhere in the 'internet of things' is towards 'ultra-low-power tiny machine learning (TinyML) systems' (Banbury et al. 2021). Size matters because the whole idea with wake word detection systems is that they are 'always on', so that more demanding tasks in terms of computation and energy consumption don't have to be. This is not just a technical achievement but a political one too. It also has major implications for what tech companies have been able to claim about privacy.

Privacy

The Echo arrived into a climate of mounting concern over corporate data collection practices, and barely eighteen months after the Snowden revelations, detailing mass surveillance by the US government and its international partners. As the cache of documents leaked by Snowden showed, programs such as EViTAP, RHINEART, VoiceRT, and SPIRITFIRE had enabled the US' National Security Agency (NSA) and members of Five Eyes to use automatic speech recognition and transcription technologies, along with audio-fingerprinting techniques and targeted keyword searches, to analyse international telephone calls, media broadcasts, and intercepted audio recordings at breathtaking speed and scale (Fidler 2015; Froomkin 2015; Kofman 2018). All this piggybacking, moreover, on massive corporate platforms designed to extract and monetise as much user data as possible; exactly like the one Amazon was now

---

[1] My thanks to Kathy Reid for drawing my attention to this point.

busy building. In fact, the company was fresh from signing a US$600m deal with the CIA just months before Snowden's leaks (Metz 2013).

So, when the Echo launched, surveillance was the unavoidable backdrop and privacy a readymade political and legal frame. 'Goodbye privacy, hello 'Alexa'' read one Guardian headline (Carroll 2015). The Atlantic published a piece on 'Why digital assistants are a privacy nightmare' (Waddell 2016). 'Virtual assistants such as Amazon's Echo break US child privacy law' explained the Center for Digital Democracy (Harris 2016). Amazon could insist that its security measures were exemplary, and that it wasn't knowingly giving governments a back door. And it did (Amazon 2025b). But the fact remained that it was asking the public to buy and install 'always on' networked listening devices through their homes, and that this massively increased their exposure to state surveillance, surveillance capitalism (Zuboff 2018), hackers, court orders, and espionage. Even Bezos' 'kill switch' would be cold comfort to anyone following closely. In a 2015 interview, Snowden described a surveillance tool called 'nosey smurf'. 'Smartphone users can do "very little" to stop security services getting "total control" over their devices,' he said. 'For example, if it's in your pocket, [GCHQ] can turn the microphone on and listen to everything that's going on around you – even if your phone is switched off because they've got the other tools for turning it on' (Taylor 2015). If it worked for your phone, presumably it would work for the Echo too.

Companies like Paranoid (2025) saw a market opportunity here, offering third party hardware that could be attached to your smart speaker to 'physically block eavesdropping' whenever activated. A product called the 'home wave' would 'discreetly generate noise and interference in close proximity to the smart speaker's microphones to jam them'. But it never took off. Amazon undoubtedly benefited from an atmosphere of fatalism after Snowden: a sense that if you're going to be surveilled anyway you might as well have a cool new toy. It helped too that the company pursued such an aggressive pricing policy. The Echo was initially

sold as a loss-leader to generate hype and encourage sales (Turow 2021). But Amazon also went out of its way to make the Echo seem as non-invasive as possible. And much of its strategy centred on the wake word.

The fact that wake word detection happens on device, so that no sound or speech is uploaded to Amazon's servers until the wake word is detected, was crucial in this respect. This design feature is what enabled Amazon (2025c) to insist that, unlike the NSA, 'Alexa does not listen to your personal conversations'. How does Alexa know that a conversation is 'personal'? By the fact that the wake word hasn't been detected yet. 'Echo devices are designed by default to detect only the sound waves of your chosen wake word,' the company explains today (2025c). 'Everything else is ignored.' This is the design feature on which the metaphor of wakefulness hangs. This is what enabled Amazon to characterise the Echo as 'always on' but mostly sleeping: not *really* listening until data starts being uploaded to the cloud. Such a characterisation draws, of course, on the history of the 'sleep mode' in computing which, as Jonathan Crary points out (2013, 13), 'remakes the larger sense of sleep into a deferred or diminished condition of operationality and access': a 'state of low-power readiness'. In this way of thinking, 'nothing is ever fundamentally "off" and there is never an actual state of rest'. Sleep comes to designate its opposite: a new, highly focused but deeply fallible mode of permanent auditory attention.

Technically, the wake word is an acoustic event that triggers a series of protocols within an algorithmic system. Rhetorically, it invites users to imagine the smart speaker as relatively passive or benign. Politically and jurisprudentially, however, the wake word is all about consent. The wake word is meant to be a speech act in the liberal tradition, like a contract, promise, or command (Austin 1975). It's a legal technology, a way of agreeing to being listened to, of activating the terms and conditions, and communicating to Amazon that what you're about to say is no longer private. In this respect, the wake word draws on the

histories of liberal contractarianism (Cudd and Eftekhari 2021), phonocentrism (Derrida 1998), and what Judith Butler (1997) calls the 'conceit' of individual sovereignty at least as much as it does the history of computing. Recall the Echo's launch video. A young girl worries that the device is already listening, and her father reassures her that 'No. It only hears you when you use the wake word we chose. Alexa.' The wake word doesn't just initiate listening in this scene, it authorises it. 'Alexa is a part of your life only when you ask Alexa to be', Amazon (2025c) maintains today in a dedicated wake word explainer. 'Think of the wake word as a verbal cue that makes things happen. In the classic tale from The Arabian Nights, an invisible door into a mountain can only be opened by saying the phrase "Open Sesame." Use the phrase, the door opens. Don't use the phrase, the door stays shut.' There is magic here, but also agency and mastery: both of which turn out to be systematically, and strategically, overstated.

PART 2

Interface effects

The door is one of the classic metaphors in interface studies, along with the window and the frame. As Florian Cramer and Matthew Fuller (2008, 149) explain, the term interface 'appears to have been borrowed from chemistry, where it means "a surface forming a common boundary of two bodies, spaces, phases."' In this way of thinking, the interface is a threshold or boundary, literally between (*inter*) states (*face*). In computing, the 'place where flesh meets metal' (Galloway 2012, 31): the button, keyboard, screen, or mouse. Above all, according to this approach, the interface is meant to be transparent, or to feel that way. This is what people mean when they describe an interface as 'user-friendly' or 'natural'.

As Branden Hookway points out (2014, 10), however, the trouble with this 'standard, instrumental, definition' is that 'the human and machine are treated as constants rather than

variables themselves subject to the operations of the interface.' This is why, for Hookway (2014, 4), the interface is instead 'a form of relation'. The interface gives form to relations in the sense that it 'imposes itself as a condition necessary for the expression of human agency', and in doing so produces both a 'supplementation and augmentation of agency' (2014, 17). For Alexander Galloway (2012, 33), the interface is not, therefore, an object, technology, or thing – precisely not a door or a keyboard – but 'always an effect', or series of effects. Likewise for Wendy Chun (2011, 67–68), the interface is a 'functional analog to ideology'. 'Interfaces offer us an imaginary relationship to our hardware', she writes. 'They do not represent transistors but rather desktops and recycling bins.' And just as 'ideology creates subjects, interactive and seemingly real-time interfaces create users who believe they are the "source" of the computer's action'. This is why the study of the interface, like all media, is always also the study of power.

In Amazon's Arabian Nights analogy, the wake word isn't the door itself. It's the magic phrase that makes the door open: a kind of meta-interface. Once again, the metaphor doesn't quite work. After Ali Babar overhears the forty thieves use 'Open Sesame' to visit their stolen treasure and sneaks back to take some for himself, they hunt him down and try to kill him. The magic words, like all passwords, were meant to be secret. Amazon, by contrast, wants its entire customer base to know its wake word and exactly how to use it. In this respect, the wake word is more like the 'watch word' to which it is etymologically related: 'watch' deriving from the old English *waecce*, meaning the 'state of being awake', 'alert' or 'vigilant'. From the 14th century, watching was something soldiers did at night. They kept watch, while standing guard. A 'watch word' was what you said in response to the challenge of a sentry: a way of demonstrating your membership of a community; still a password but collectively held. We could say something similar about the Biblical shibboleth, to which the wake word is also related. In Jacques Derrida's famous telling (1992, 399):

> The Ephraimites had been defeated by the army of Jephthah; in order to keep the soldiers from escaping across the river (*shibboleth* also means river, of course, but that is not necessarily the reason it was chosen), each person was required to say *shibboleth*. Now the Ephraimites were known for their inability to pronounce correctly the *shi* of *shibboleth*, which became for them, in consequence, an "unpronounceable name"; they said *sibboleth*, and, at that invisible border between *shi* and *si*, betrayed themselves to the sentinel at the risk of death.

Unlike the watch word, the shibboleth draws attention to the theme of orality, and so the politics of the speaking voice. It is not enough to know the wake word, after all. You must be able to pronounce it too.

In an essay on accent bias in voice recognition systems, Halcyon Lawrence recounts an interaction with a friend in her native Trinidad. 'I watched in fascination', Lawrence writes (2019, 185), 'as she interacted with Siri, first in her native Trinidadian accent and then, when Siri did not understand her, repeating her question in an American accent. Siri sprang to life.' What does the wake word do in this scene? As a form of vocal discipline, it assimilates: an inverted shibboleth. For Lawrence, it is deeply alienating too, since 'there is no opportunity to negotiate with the device'. The discrimination is hard coded. Changing how you speak is your only option if you want to engage the interface. In both respects, we could think of the work word as a computational extension of longstanding imperial practices, along racial and other familiar lines (Eidsheim 2019). And we should not expect this dynamic to be relieved, moreover, by Big Tech's expansion into new linguistic markets, since some accents and non-normative ways of speaking will never be sufficiently profitable.

False wakes

In *How To Do Things With Words* (1975), J.L. Austin was interested in what he called 'speech acts': utterances in which to say something is also to do something: to promise, get married, or consent. Even though most of Austin's examples involve the spoken word, he was

'lamentably silent on the role of the voice' (Duncan 2004, 289). 'Tone of voice', 'cadence', and 'emphasis' are all alluded to but never taken up in detail. And in his well-known discussion of the 'felicity' of speech acts, Austin misses the fact that vocality is necessarily bound up in the achievement of linguistic effect. If an order, oath, or apology doesn't sound 'right', if it isn't clear or serious enough, if a wake word is delivered in the wrong accent or from too far away, it will 'misfire'. For the non-normative speaker, the wake word fails because a wake word detection algorithm is under-inclusive. It hasn't been designed with their voice in mind. But wake word detection systems are over-inclusive too. This is a form of performative infelicity that Austin never imagined: words that do things when they haven't been uttered at all.

The wake word and what Amazon refers to as 'false' or 'accidental' wakes were invented together, just like the ship and the shipwreck (Virilio 2007). Alexa can be accidentally triggered by an ad, or a song, by conversations about Alexa, or elections, or with your partner Alex. In one widely reported example, an American couple's Echo took it upon itself to record a conversation they were having and send it to a colleague, who immediately called to tell them they'd been 'hacked'. Their device had been woken 'due to a word in background conversation sounding like 'Alexa'.' 'Then, the subsequent conversation was heard as a 'send message' request. At which point, Alexa said out loud 'To whom?' At which point, the background conversation was interpreted as a name in the customer's contact list. Alexa then asked out loud, '[contact name], right?' Alexa then interpreted background conversation as 'right'." (Wolfson 2018)

This may not have been great publicity for Amazon, but it's exactly the kind of problem technology companies like to have, since it's readily incorporated into a progress narrative. Today, Amazon is happy to admit that 'on some occasions, Alexa may wake up when no one has said the wake word.' 'That's getting better', it says (2025c). 'We are

constantly improving our wake word detection technology'. They will have to keep on going forever. All technologies need maintenance, and here background 'noise' and variabilities in speaker accent, diction, and tone will always 'interfere' with intelligibility in ways that that risk reproducing structures of racism, sexism, ableism, and so on. But the problem for the wake word is also language itself, and in particular what Derrida (1988; 1998) called the 'iterability' or 'citationality' of the utterance. The fact that the wake word can be repeated—and therefore exceeds or is irreducible to the speaker's intent—is what makes it viable *as* a wake word, what defines it as a *word* in fact, but also what makes the 'false' or 'accidental' wake unavoidable. Jeff Bezos can refer to Alexa in a board meeting, an actor can say 'Alexa' in an ad, without intending to wake a device or consenting to listening, but they will always run the risk of doing so anyway, because wake word recognition systems recognise wake words, not intent.

Amazon knows this. To mitigate the problem of false wakes from advertising, it requires audio featuring 'Alexa' to be submitted at least twelve days before airing, so that it can be 'fingerprinted' and registered across the assistant's back end (Amazon 2025d). An audio fingerprint is a unique hash or signature 'derived from perceptually relevant aspects of a recording' (Cano et al. 2005, 233). This hash can be produced in many ways (Cano et al. 2005; Kamuni et al. 2024), but the basic idea is that once a corpus of recordings has been gathered and fingerprinted an unknown audio sample can be compared to it for the purposes of identification. The technique's most well-known application is probably by Shazam, for music ID. But Amazon uses it here to register individual recordings of 'Alexa' as unique sonic events so that they can then be excluded from its wake word recognition systems as a result.

This is a clever workaround, but it doesn't avoid the fundamental challenge of citationality, which *cannot* in fact be avoided, since the speech act depends on the utterance being repeatable. The wake word works *because* it is iterable: precisely because it is *not*

unique. From this perspective, the 'false' wake looks less accidental and more like a founding risk. The false wake is inevitable. There is simply no way of guaranteeing the user's consent. Contrary to the official corporate position, we could say that what Amazon invented with the wake word, in the name of protecting privacy, was in fact a new form of privacy breach, which no amount of incremental technical improvement is ever going to fix.

What does the wake word authorise?

The problem is even more acute when you widen the lens from the fact to the quality of consent. What kind of listening you are authorising when you say the wake word has always been mysterious and is constantly changing. On one level, Amazon is up front about this. It's right there in the Terms and Conditions, where the company retains the right to 'change, suspend, or discontinue Alexa, or any part of it', and to 'amend any of this Agreement's terms' at any time and without notice (Amazon 2025e). But the onus is on users to keep informed of changes, even though the relevant provisions run to thousands of words of inter-linked agreements, notices, policies, and guidelines, and 'continued use of Alexa' will be taken as acceptance anyway (Obar and Oeldorf-Hirsch 2020; Neville 2020; Amazon 2025e).

Such changes are only ever in one direction: towards more listening, not less. In 2017, for instance, Amazon introduced 'voice ID'. 'When you create a voice ID', it explains, 'Alexa uses recordings of your voice to create an acoustic model of your voice characteristics and to update that model over time to improve Alexa's ability to recognize you' and so 'better personalize your experience' (Amazon 2025e). What precisely is meant by 'personalisation' here is unclear, and constantly evolving anyway, but it seems to include targeted advertising. 'Alexa Advertising ID' is a unique user profile 'made available … to third-party skill developers and content providers … for advertising and analytics purposes.' According to Amazon, 'each third-party developer and content provider receives a different Alexa

Advertising ID for you, and a different Alexa Advertising ID for each device you use to access the applicable third-party skill or content.' Advertisers not only know who is speaking, but where, as well as what kinds of requests they tend to make. 'For instance, if you ask Alexa to order paper towels or for recipe ideas, you might see or hear ads related to cleaning products or cooking utensils.' (Amazon 2025e)

So, Amazon isn't just doing vocal biometrics, it's also constantly updating and refining its models to identify and distinguish between speakers more accurately. And these models are tethered to advertising profiles so that Amazon can monitor and monetise the content of all your engagements with Alexa. These are the forms of listening Amazon is explicit about if – and it's a big if – you regularly read and understand the terms and conditions. But much of what Alexa does remains opaque, even if you do. 'Alexa is designed to get smarter every day,' Amazon says, to 'improve your experience and our services'.

> For example, we use your requests to Alexa to train our speech recognition and natural language understanding systems using machine learning. Training Alexa with real world requests from a diverse range of customers is necessary for Alexa to respond properly to the variation in our customers' speech patterns, dialects, accents, and vocabulary and the acoustic environments where customers use Alexa. (Amazon 2025e)

How though? What exactly does Amazon know about users' speech patterns, dialects, accents, vocabulary, and acoustic environments? How is this information being operationalised? How is it being transformed into capital or profit? Is the list even exhaustive? What will be added tomorrow?

Now we can see the wake word's political-economic function more clearly. It was always partly an exercise in misdirection: a trojan horse designed to inveigle listening devices into our homes via a fantasy of consent. The question was never just whether Alexa is listening to our personal conversations, but always also what is 'personal' about our conversations with Alexa. Yes, the wake word is a way for users to access Alexa, but it's also

a way for Amazon to access us: a technique of what Bernard Harcourt (2015) has called

'expository power', and part of an increasingly dominant mode of capitalist production,

defined by the manufacture, ownership, and automated analysis of information (Zuboff 2018;

Srnicek 2017; Wark 2019; Cohen 2019).

   This information isn't just out there, waiting to be collected and monetised. People

need persuading. Privacy concerns must be overcome, consent seemingly given. Exposure

needs to become a kind of default, and users taught not just to expect their devices to listen,

but to want them to. This is why, for Jathan Sadowski (2019, 2), 'data mining is a misleading

name.' 'A more apt term would be data manufacturing.' The wake word is a manufacturing

technique: a way of initiating 'data relations' (Couldry and Mejias 2018), and an

indispensable part of the 'process of datafication' of the home and beyond (Sterne and

Sawhney 2022, 295). In aggregate, this data is incredibly valuable. It can be used for targeted

advertising, and to help make Alexa 'smarter' in the hope that this will drive more users,

advertisers, and third-party applications, and so more data, to the Alexa ecosystem in turn.

The fact that all this data isn't derived from 'personal conversations', or that users 'agreed' to

give it up, is scant consolation, or excellent PR, considering that what users are pressed to

'agree' to is always growing, and that Amazon and its owners continue to accumulate

epistemic power along with political and industrial influence as a result.

   It's important to extend the analysis beyond short term capital accumulation here.

Alexa, like many voice assistants, has never been directly profitable. Despite selling millions

of units, Amazon 'lost' US$25 billion on Alexa between 2017 and 2021. Alexa-based

shopping never really took off. Neither did advertising. And Amazon has laid off thousands of

workers since 2022, 'with the Devices division reportedly hit especially hard' (Harding 2024).

But profitability isn't the only measure of value. 'We're focused on the value we create when

customers use our services, not just when they buy our devices', Amazon explains (Harding

2024). Alexa not only brings customers into the company's ecosystem, it helps keep them there, bolstering brand recognition in the process. Technical advances generate value as intellectual property and can be redeployed across the company's many other ventures. Alexa gave Amazon an early foothold in the market for AI agents, which in the age of LLMs, Big Tech continues to push hard. This is why Amazon has invested US$8 billion in Anthropic, the company behind Claude, and now Alexa+, the latest 'generative' version of its assistant, since 2023 (Reuters 2024). Yes, the aim is ultimately to make these agents profitable. But it's also to build a world reliant on them, along with all manner of other smart technologies, which the agents will then offer to control. Some of these other technologies and services may turn out to be valuable. But crucially for Amazon, this world will require massive compute: Amazon Web Services being by far its most profitable arm (Bishop 2025). From this perspective, the wake word looks like the thin end of a very long wedge: yet another method of the 'capitalization of human life without limit' (Couldry and Mejias 2018).

PART 3

Wake word proliferation

When the other big tech companies rushed to launch their own smart speakers in the years after Amazon, part of what they were competing for was data and the growing markets for voice assistants and the smart home. But they were also competing for cognitive and acoustic space: to have consumers say *their* wake word out loud and for anyone within earshot. The wake word is a form of sonic branding (Gustafsson 2019). It's just that, in addition to targeting the ear like a theme song or jingle, the wake word also targets the voice. In this respect, it has more in common with the rituals of church and state than Brian Eno's famous 'start-up sound' for Windows or Jim Reeke's 'bong' for Mac. Like reciting a prayer, swearing an oath, or singing the national anthem, the wake word is about invocation and orality (Dolar

2006; Parker 2015). It enrols the speaking body into a ritual of brand identification and reproduction as a condition of access to the interface. Once upon a time, you could turn on the radio or switch off the lights without a specific brand or product name ever entering your mind. Now, the aspiration is for as many interactions to pass via the wake word as possible. This is partly why in 2020 Amazon launched 'Alexa for residential' (2025f), 'an all-new service that makes it easy for property managers to set up and manage Alexa-powered smart home experiences throughout their buildings.' The scheme has since been re-named 'Alexa Smart Properties' to capture its growing interest in care homes, stadiums, and hospitality (Amazon 2025g). These are spaces in which users cannot easily opt out of the Alexa 'brand experience'. 'Build the new normal', Amazon declares (2025h). In the corporate fantasy, this new normal is a world in which every interaction with one's environment is annexed to and accessed via the brand as infrastructure.

In the second half of the 2010s, competition in this new field of sonic branding was soon playing out across a huge range of devices. Every company with the resources to produce a voice assistant wanted to make itself the point of access; to put *its* brand name in users' mouths. While the biggest companies used their voice assistants to enmesh consumers more completely in their ecosystems, and to sell hardware of their own, they also struck deals with many third-party manufacturers, who were willing to host these assistants, for fear of being left behind. Suddenly, you could access Google Assistant on your Sony TV, Cortana on your ThinkPad, and Siri in your car. In 2017, for instance, the wireless speaker manufacturer Sonos partnered with Amazon to host Alexa, to considerable fanfare, before adding Google Assistant compatibility in the years after that (Titlow 2017). In 2018, 66 Audio became the first headphone company to offer full Alexa integration (Kinsella 2018). As these companies quickly learned, however, this kind of association could sometimes prove counterproductive.

In one TV commercial from 2020, we see a woman getting into an SUV somewhere in suburban America. 'Alexa, take me downtown' she says, and a route appears on the car's display. 'Alexa, give me some salary negotiation tips', she continues, and it does. Now we see the woman handing a document to someone in a suit before, back in the car, 'Alexa, call Jeff,' and a phone rings. 'Hey, how did the interview go?' 'I start Monday!' 'Congratulations,' Jeff replies. Some text appears on screen. 'Make moves', it says, 'Alexa's on board'. Only then, in the ad's final seconds, do we realise that this was meant to be an ad for the car. 'Buick vehicles now come with the Alexa built-in,' the text reads, and the music fades out.

Perhaps Buick knew what they were doing. Perhaps the association with Amazon was worth reducing themselves to a footnote in their own campaign. Perhaps Amazon just paid them a lot of money. Either way, this ad perfectly stages the conundrum facing many companies today. Without a wake word of your own, your product risks being overshadowed by the brand identity of the larger company whose wake word is used to access it. This is why a whole secondary market has now developed for 'custom branded wake words' (Sensory 2025), intended to 'deepen user engagement, increase brand affinity, and inspire loyalty' (Soundhound 2025a). One of the market leaders here is SoundHound, an 'independent voice AI platform' that claims to 'connect people to brands through customized conversational experiences' and whose mission it says (2025b) is to 'add voice to everything'. SoundHound isn't a household name, but it's a substantial company. Having begun life as a music recognition platform in 2005, in November 2021 it went public for US$2 billion. Its customers include Mercedes, Hyundai, Honda, Kia, Samsung, Motorola, Netflix, Nvidia, and Qualcomm, along with many hotels, restaurants, and call centres. And it insists on the value of developing 'a dedicated wake word for your brand' – 'Hello Hyundai', 'Ok Honda'. 'By saying the custom wake word multiple times during the day,' SoundHound explains (2021), 'users won't forget which car, TV, or app they're using. Your brand will be the first

connection customers make when they use your voice assistant—which becomes especially impactful when creating an omnichannel experience across multiple devices.'

Clearly this kind of customization is good business, because in 2021 Amazon announced a competitor platform for Alexa. 'Alexa Custom Assistant' was initially only available for third-party devices, made by the likes of BMW ('Hello BMW') and Garmin ('OK Garmin'). But for the first time, in June 2023, Amazon announced a new wake word and 'custom voice assistant' for Amazon hardware. 'Hey Disney!' would bring 'the best of Disney, Pixar, and Star Wars characters' to Echo devices in Disney Resort hotel rooms and to users' homes via a monthly subscription to 'Amazon Kids+' (Amazon 2025i).

In the decade since wake words went mainstream, they've proliferated and diversified. Not only are wake words an increasingly common feature of everyday encounters with technology, there are also more of them than ever before, as companies realise the wake word's potential as a technique of data manufacturing and sonic branding. At the same time, though, and in tension with this clear tendency towards wake word expansion, we can observe a countervailing tendency towards wakewordlessness. Tech companies are starting to sell the wake word's removal back to consumers as a feature, product, or convenience.

Wakewordlessness

The wake word was always already eroded. Amazon's original patent, filed by Rawles LLC, notes that 'some implementations may buffer the audio input before and after the wake word to allow for variations. For example, the user may say, "Open my e-mail, Wake Word" and the system could recognize the buffered audio input recorded prior to the wake word as a command to open the user's e-mail.' In other implementations, 'the user may speak the wake word separate from the natural language command. For example… the user may speak the wake word, the voice controlled device may output an indication that it has been activated,

and the user may then speak the natural language command' (Salvador et al. 2016, 8). In the first example, buffering enables the wake word to authorise listening retrospectively. In the second, the system continues to authorise listening after a break in speech. What this amounts to in both cases – before the public had even heard of a wake word, let alone used one – is an acknowledgment that there is something 'unnatural' about it: that it may be hard to get people to speak this way.

This acknowledgment matters because the rationale for voice interfaces has always been their 'naturalism'. This rationale not only plays on the longstanding association between the voice, subjectivity, and nature (Derrida 1998; Cavarero 2005; Dolar 2006; Sterne 2011), it was central to the earliest efforts at speech recognition in the 1960s. 'The idea that speech constitutes the most "natural," "fundamental," and "spontaneous" means of communication for humans', Xiaochang Li (2017, 44) writes, 'at once drove the pursuit of its mechanization and defined the means by which that pursuit was undertaken.' For J.C.R. Licklider (1960, 10), in his famous paper on 'Man-Computer Symbiosis', 'there is a continuing interest in the idea of talking with computing machines. In large part, the interest stems from the realization that one can hardly take a military commander or a corporation president away from his work to teach him to type.' So, 'if computing machines are ever to be used directly by top-level decision makers, it may be worthwhile to provide communication via the most natural means, even at considerable cost.' Here is Amazon's senior vice president of devices half a century later. 'Voice is the most natural of interfaces,' he says (Gibbs 2016), 'the one we use from birth. It's simple for humans, but it's a very hard challenge for computers. When it's done right, it's delightful for customers.'

For Amazon, 'doing it right' has meant chipping away at the wake word as a barrier to 'natural' speech. In 2018, the company introduced 'follow up mode', an opt-in feature enabling users to 'have natural, back-and-forth interactions with Alexa, without needing to

repeat the wake word each time you want a response.' (Amazon 2025j) Initially, this worked by having Alexa continue listening for five seconds after the initial instruction (Statt 2018), but it was soon expanded into a more open-ended conversational state, in which users can issue follow-up commands and freely interject with questions, clarifications and non-sequitur (Amazon 2025j). 'Once enabled, you'll still need to use the wake word to start a conversation with Alexa but not to continue it.' 'After all, 'you don't usually repeat the names of your family and friends when you talk to them… It's not natural.' (Amazon 2025k)

Or consider 'quick phrases', launched in 2024 to enable users to 'get Google Assistant's help with everyday tasks without "Hey Google".' If 'follow up mode' undermines the wake word by extending its reach, quick phrases erode it by building it in. The feature makes certain common commands wakewordless: 'stop' or 'snooze' when an alarm goes off, 'answer' or 'decline' when you get a call, or 'turn on kitchen lights' whenever you like. In these situations, the wake word literally goes without saying. It's derived or inferred from the content of the statement in combination with the context of its utterance. Once again, the purpose is to enable users to 'speak naturally' to their devices (Google 2025). Once again, the irony is how carefully such 'natural' dialogue must be cultivated. Both examples come with dedicated explainer videos and web pages. Both invite users to read updated privacy agreements and information pages before opting in.

Ambient assisted living

Browsing one such page, I find the following. 'When Enhanced Follow Up Mode is on, your device uses a combination of on-device visual and acoustic processing to determine when you are … speaking to Alexa.' (Amazon 2025j) This is how it decides whether to continue the conversation, and so if a wake word is required. Here is another pathway to wakewordlessness then, bound up with but distinct from the desire for more 'natural' human-

machine dialogue: the wake word's increasing substitution with persistent ambient monitoring.

One clear example of this shift is 'Alexa Routines', a way of grouping together actions in the smart home so that users don't have to ask for them individually (Amazon 2025k). Routines can range from the simple – turning on the lights and playing some music – to something much more complicated. According to one writer at *The Verge* (Tuohy 2022), 'my Good Morning Routine gradually brightens the lights in my entire house, opens the shades, adjusts the thermostat, tells me the weather, traffic on the school run, and my calendar for the day, and turns the kettle on. It also starts my favorite radio station on every speaker in the house.' Originally, routines like this were either voice activated ('Alexa, Good Morning'), or set to start and stop at particular times. But Amazon has gradually expanded the range of 'events' that will trigger a routine to include sunrise and sunset, geo-location data from a smartphone or watch, and a growing range of sounds. Routines can be set to soothe your baby with lullabies when crying is detected; send a notification to your phone when an appliance beeps; remind you to drink water when an Echo detects coughing; or calm your dog by playing soothing music when they bark. So-called 'occupancy routines' now use ultrasonic motion detection – where the ultrasound is both emitted and monitored by the Echo device – to determine when someone is in a room and respond accordingly (Amazon 2025l). 'Alexa Emergency Assist' is a monthly subscription service that 'proactively listens for acoustic signals' so that users can be notified and an Amazon 'Urgent Response agent' called if the sound of a smoke alarm or breaking glass is detected (Amazon 2025m).

What we are seeing here isn't just the erosion of the wake word. It's the gradual substitution of one mode of governance for another: the replacement of a liberal paradigm, organised around the ideas of vocal agency, privacy and consent, which Amazon and other tech companies consistently invoked to encourage uptake of voice assistants, with a more

operational, pre-emptive, environmental logic (Andrejevic 2020). According to this logic, the subject is routinely by-passed or ignored. Their environment becomes a 'flexible, programmable context' (Andrejevic 2020, 18) and the system is designed to anticipate and fulfill the user's desire before they've had the chance to articulate it. From the perspective of such a system, the wake word is literally a waste of breath. Why go to the bother of saying 'Alexa', when you can just walk into a room? Why attend to your dog or baby, when your Echo can do it before you've had a chance to get up, or even really noticed?

The clear trajectory is towards what, in elder and disability care, they call 'ambient assisted living'. Alexa Emergency Assist already targets this market. But many other companies and research labs offer far more comprehensive forms of monitoring than anything yet available from the big providers. Monitoring of patient heart rate and sleep patterns, body temperature, medicine and diet adherence. Monitoring of patient behaviour, patient media consumption, patient location, activity and inactivity, along with signs of emergency like shouts, screams, and falls. And all for the purposes of automated intervention. Moreover, machine listening is claimed to be particularly good in this respect because wireless networked microphones are cheap, small, robust, non-intrusive, and easy to deploy, in addition to being multi-directional and multi-purpose (Navarro et al. 2018). In the name of patient health and security, all such monitoring will be permanent and wakewordless. 'Companion robots' like ElliQ abandon the wake word in favour of 'pro-active intelligence'. 'ElliQ doesn't wait to be called on', its maker explains (2025). 'Like a real roommate, ElliQ will initiate interaction, ask you questions, suggest activities, provide reminders, and more.'

The considerable material and ideological investment in ambient assistant living is taking place in a world of ageing populations and the ongoing privatisation of healthcare, so that the market in elder and disability care is sizeable on its own. But, as Mara Mills (2010, 39) warns, disability often serves as a 'pretext to other engineering concerns', and big tech

would like nothing more than for ambient assisted living to be generalised. We could think of residential care as a kind of laboratory for the smart home, and the 'assistive pretext' as a key driver of wakewordlessness and the authorisation of precisely the kinds of permanent and pervasive listening that the wake word always promised to withhold.

Conclusion

At the end of February 2025, Amazon announced Alexa+, its 'next generation voice assistant powered by generative AI', available for the first time via a paid subscription to Amazon prime. In the launch video, Alexa initiates the conversation. 'Oh, hey there' it says to a teenager lying in bed, hands clasped to her chest. 'So, we can just like talk now?' she wonders, and the device on her bedside table responds, 'I'm all ears, figuratively speaking'. The ad works through the assistant's new 'agentic' capabilities. A man gets help with a plumbing job. A mother uses Alexa+ to check her kids' schedule and order an Uber for their grandma. And no-one uses the wake word at all (Amazon 2025n). We are a long way from Amazon's promise ten years before – in a very different political climate, and with surveillance capitalism much less advanced – that the Echo would 'only hear you when you use the wake word we chose. Alexa.'

Digging into the other launch videos, it's clear that the wake word hasn't gone entirely. It's receded: become a less prominent because less politically salient part of the interface. 'Alexa, play the show', a woman says in one scene, before a long, elaborate conversation unfolds. No-one asks about privacy or consent. In their place, Amazon offers 'personalisation', 'deep knowledge', and more 'pro-active', 'natural', 'context aware' conversations. Not only does Alexa+ 'know what you've bought, what you've listened to, the videos you've watched, the address you ship things to, and how you like to pay—but you can also ask her to remember things that will make the experience more useful for you.' This

might include the fact that you 'love pizza', or that 'your partner is gluten-free', but also your 'documents, emails, photos and messages', your emotional state, facial expressions, tone of voice, morning and nighttime routines. 'You can now share just about anything with Alexa', the company says (Amazon 2025a).

Alexa+ is Amazon's contribution to the current wave of so-called 'generative' voice assistants, powered by LLMs rather than rule-based architectures. ChatGPT 'voice mode' is already available for smartphones and web browsers. Meta AI can be accessed on third party smartphones along with Ray-Ban Meta glasses and Meta Quest headsets. Google will roll out Gemini across mobile and other devices across the year. As Big Tech goes all in on these assistants, it is worth remembering the wake word, its role in making this world possible, and the political, legal, and imaginative work it will continue to do. Despite going mostly unremarked upon in academic and critical literatures, the wake word has been one of the more important techniques of surveillance capitalism. It appealed to sci-fi imaginaries years before the 'tech lash', and helped Amazon fend off privacy concerns in the wake of the Snowden revelations, enabling it to establish the smart speaker as a central feature of the smart home in the process. Huge volumes of speech and sound were captured in the process: first, rendered as data, then immediately enclosed. All this was ostensibly done with the user's consent: 'freely' given every time they utter the wake word. This consent is severely compromised, so far from the liberal ideal that it is better to think of the wake word as a technique of expository power, appealing to the liberal tradition precisely in order to undermine it in favour of alternate mode of governance altogether. Even as wake words proliferate as a newly intimate form of sonic branding, aimed at the voice as well as the ear, the wake word is also being eroded, and wakewordlessness sold back to users as a service. This is a world in which the subject can be by-passed, their needs anticipated, and environment modulated, all in the

name of a frictionless automation in which the wake word feels increasingly like an artefact of

a distant political moment.

References

Amazon, 2014. Introducing Amazon Echo. https://www.youtube.com/watch?v=CYtb8RRj5r4

Amazon, 2025a. Introducing Alexa+, the next generation of Alexa.
https://www.aboutamazon.com/news/devices/new-alexa-generative-artificial-intelligence

Amazon, 2025b. Privacy.
https://www.amazon.com/b/?node=23638098011&ref=MARS_NAV_desktop_privacy_Built_In_Controls

Amazon, 2025c. How Alexa Works: Wake word.
https://www.amazon.com/b/?node=23608571011#concept3

Amazon, 2025d. Audio Fingerprinting. https://developer.amazon.com/en-US/alexa/branding/alexa-guidelines/fingerprinting

Amazon, 2025e. Alexa Terms of Use.
https://www.amazon.com.au/gp/help/customer/display.html?nodeId=201809740

Amazon, 2025f. A new, easy way for properties to add Alexa to residential buildings.
https://www.aboutamazon.com/news/devices/a-new-easy-way-for-properties-to-add-alexa-to-residential-buildings

Amazon, 2025g. Alexa Case Studies. https://developer.amazon.com/en-GB/alexa/alexa-case-studies

Amazon, 2025h. Alexa Smart Properties. https://developer.amazon.com/en-GB/alexa/alexa-smart-properties

Amazon, 2025i. Hey Disney! https://www.amazon.com/b/?node=107538412011

Amazon, 2025j. Follow up mode. https://www.amazon.com/b/?node=23660877011

Amazon, 2025k. Conversation mode.
https://www.aboutamazon.com/news/devices/conversation-mode-helps-interactions-with-alexa-feel-more-natural

Amazon 2025l. What is Ultrasound Motion Detection?
https://www.amazon.com/gp/help/customer/display.html?nodeId=GSR22RYDWS3KBUYW

Amazon, 2025m. Alexa Emergency Assist. https://www.amazon.com/Alexa-Emergency-Assist-Monthly-auto-renewal/dp/B0BZSZBK3T

Amazon, 2025n. Meet the new Alexa. https://www.amazon.com/dp/B0DCCNHWV5

Andrejevic, Mark. 2020. *Automated Media*. London: Routledge.

Apple, 2014. WWDC. https://www.youtube.com/watch?v=w87fOAG8fjk

Austin, J.L. 1975. *How to Do Things with Words*. Oxford: Oxford University Press.

Banbury, Colby, Vijay Janapa Reddi, Peter Torelli, Jeremy Holleman, Nat Jeffries, Csaba Kiraly, Pietro Montino. 2021. MLPerf Tiny Benchmark.
http://arxiv.org/abs/2106.07597.

Bishop, Todd. 2025. Amazon's Quarterly Profits Soar to a Record $20 Billion, but Cloud Growth Comes up Short. *GeekWire*. February 6.
https://www.geekwire.com/2025/amazons-quarterly-profits-soar-to-a-record-20-billion-but-cloud-growth-comes-up-short/.

Bohn, Dietr. 2017. You Can Finally Say 'Computer' to Your Echo to Command It. *The Verge*. https://www.theverge.com/tldr/2017/1/23/14365338/amazon-echo-alexa-computer-wake-word-star-trek.

Brustein, Joshua. 2016. The Real Story of How Amazon Built the Echo. *Bloomberg*.
https://www.bloomberg.com/features/2016-amazon-echo/.

Butler, Judith. 1997. *Excitable Speech: A Politics of the Performative*. London: Routledge.

Cano, Pedro, Eloi Batlle, Emilia Gómez, Leandro de C.T.Gomes, and Madeleine Bonnet. 2005. Audio Fingerprinting: Concepts and Applications. *Computational Intelligence for Modelling and Prediction*, edited by Saman K. Halgamuge and Lipo Wang, 2: 233–45. Berlin: Springer.

Carroll, Rory. 2015. Goodbye Privacy, Hello 'Alexa': Amazon Echo, the Home Robot Who Hears It All. *The Guardian*, 2015. https://www.theguardian.com/technology/2015/nov/21/amazon-echo-alexa-home-robot-privacy-cloud.

Cavarero, Adriana. 2005. *For More than One Voice: Toward a Philosophy of Vocal Expression*. Stanford: Stanford University Press.

Chhetri, Amit S., Philip Hilmes, Trausti Kristjansson, Robert Ayrapetian, Wai Chu, Mohamed Mansour, Xiaoxue Li, and Xianxian Zhang. 2018. Multichannel Audio Front-End for Far-Field Automatic Speech Recognition. *Amazon Science*. 2018. https://www.amazon.science/publications/multichannel-audio-front-end-for-far-field-automatic-speech-recognition.

Christiansen, R., and C. Rushforth. 1977. Detecting and Locating Key Words in Continuous Speech Using Linear Predictive Coding. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 25 (5): 361–67.

Chun, Wendy Hui Kyong. 2011. *Programmed Visions: Software and Memory*. Cambridge, Mass: MIT Press.

Cohen, Julie E. 2019. *Between Truth and Power: The Legal Constructions of Informational Capitalism*. Oxford University Press, USA.

Couldry, Nick, and Ulises Mejias. 2018. Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Television and New Media*, 1–14.

Crary, Jonathan. 2013. *24/7: Late Capitalism and the Ends of Sleep*. London: Verso.

Cudd, Ann, and Seena Eftekhari. 2021. Contractarianism. *The Stanford Encyclopaedia of Philosophy*, edited by Edward N. Zalta. https://plato.stanford.edu/archives/win2021/entries/contractarianism/.

Deloitte Insights. 2019. Technology, Media, and Telecommunications Predictions 2019.

Derrida, Jacques. 1988. *Limited Inc*. Evanston, IL: Northwestern University Press.

Derrida, Jacques. 1992. *Acts of Literature*. London: Routledge.

Derrida, Jacques. 1998. *Of Grammatology*. Baltimore: JHU Press.

Dolar, Mladen. 2006. *A Voice and Nothing More*. Cambridge, Mass: MIT Press, c2006.

Duncan, Michelle. 2004. The Operatic Scandal of the Singing Body: Voice, Presence, Performativity. *Cambridge Opera Journal* 16 (3): 283–306.

Eidsheim, Nina Sun. 2019. *The Race of Sound: Listening, Timbre, and Vocality in African American Music*. Durham: Duke University Press.

ElliQ, 2025. Meet ElliQ. https://elliq.com/pages/features

Engst, Adam. 1992. Casper Speaks. TidBITS. https://tidbits.com/1992/03/09/casper-speaks/.

Felker, Nick. 2020a. Voice Tech in Star Trek: TNG Supercut, Season 4. https://www.youtube.com/watch?v=6CDhEwhOm44&t=790s

Felker, Nick. 2020b. Voice Tech in Star Trek: TNG Supercut, Season 7. https://www.youtube.com/watch?v=qotCgA26Fp8&t=9s

Fidler, David P. 2015. *The Snowden Reader*. Bloomington: Indiana University Press.

Froomkin, Dan. 2015. How the NSA Converts Spoken Words Into Searchable Text. *The Intercept*. May 5. https://theintercept.com/2015/05/05/nsa-speech-recognition-snowden-searchable-text/.

Fuller, Matthew, and Florian Cramer. 2008. Interface. *Software Studies: A Lexicon*. Cambridge, Mass: MIT Press.

Galloway, Alexander R. 2012. *The Interface Effect*. Cambridge: Polity.

Gibbs, Samuel. 2016. Amazon Launches Echo Voice-Controlled Speaker and Alexa Assistant in the UK. *The Guardian*. https://www.theguardian.com/technology/2016/sep/14/amazon-echo-voice-controlled-speaker-alexa-assistant-uk.

Google, 2025. Use quick phrases on Google Nest Hub Max. https://support.google.com/googlenest/answer/11187357?hl=en

Gustafsson, Clara. 2019. Sonic Branding: From Brand Image to Brand Imagination. *The Oxford Handbook of Sound and Imagination, Volume 1*, edited by Mark Grimshaw-Aagaard, Mads Walther-Hansen, and Martin Knakkergaard. Oxford: Oxford University Press.

Harcourt, Bernard E. 2015. *Exposed: Desire and Disobedience in the Digital Age*. Cambridge, Mass: Harvard University Press.

Harding, Scharon. 2024. Alexa Had 'No Profit Timeline,' Cost Amazon $25 Billion in 4 Years. *Ars Technica*. https://arstechnica.com/gadgets/2024/07/alexa-had-no-profit-timeline-cost-amazon-25-billion-in-4-years/.

Harris, Mark. 2016. Virtual Assistants Such as Amazon's Echo Break US Child Privacy Law, Experts Say. *The Guardian*. https://www.theguardian.com/technology/2016/may/26/amazon-echo-virtual-assistant-child-privacy-law.

Hookway, Branden. 2014. *Interface*. Cambridge, Mass: MIT Press.

Kamuni, Navin, Sathishkumar Chintala, Naveen Kunchakuri, Jyothi Swaroop Arlagadda Narasimharaju, and Venkat Kumar. 2024. Advancing Audio Fingerprinting Accuracy with AI and ML: Addressing Background Noise and Distortion Challenges. *IEEE 18th International Conference on Semantic Computing*.

King, Hope. 2015. Amazon Built the Star Trek Computer for Your House. *CNN Business*. https://money.cnn.com/2015/07/24/technology/amazon-echo-review/.

Kinsella, Bret. 2016. Will Amazon Echo Sales Top 7 Million Through 2016? *Voicebot.Ai*. https://voicebot.ai/2016/12/27/will-amazon-echo-sales-top-7-million-through-2016/.

Kinsella, Bret. 2018. PRO Voice Headphones with Alexa Launched by 66 Audio. *Voicebot.Ai*. January 5. https://voicebot.ai/2018/01/05/pro-voice-headphones-alexa-launched-66-audio/.

Knox, Robert. 2021. Law. *The Sage Handbook of Marxism*, edited by Sara R. Farris, Beverley Skeggs, Alberto Toscano. London: Sage.

Koenecke, Allison, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha, Connor Toups, John R. Rickford, Dan Jurafsky, and Sharad Goel. 2020. Racial Disparities in Automated Speech Recognition. *Proceedings of the National Academy of Sciences* 117 (14): 7684–89.

Kofman, Ava. 2018. Interpol Rolls Out International Voice Identification Database Using Samples From 192 Law Enforcement Agencies. *The Intercept*. June 25. https://theintercept.com/2018/06/25/interpol-voice-identification-database/.

Ku, Pin-Jui, I-Fan Chen, Chao-Han Huck Yang, Anirudh Raju, Pranav Dheram, Pegah Ghahremani, Brian King, Jing Liu, Roger Ren, and Phani Sankar Nidadavolu. 2024. Hot-fixing wake word recognition for end-to-end ASR via neural model reprogramming. In *2024 IEEE International Conference on Acoustics, Speech and Signal Processing*.

Lawrence, H.M. 2019. Siri Disciplines. *Your Computer Is on Fire*. Cambridge, Masss: MIT Press.

Li, Xiaochang. 2017. Divination Engines: A Media History of Text Prediction. *Thesis* (PhD). New York University.

Licklider, J.C.R. 1960. Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics*, March, 4–10.

Liu, Hongyi, Apurva Abhyankar, Yuriy Mishchenko, Thibaud Sénéchal, Gengshen Fu, Brian Kulis, Noah D. Stein, Anish Shah, and Shiv Naga Prasad Vitaladevuni. 2020. Metadata-Aware End-to-End Keyword Spotting. *Interspeech 2020*, 2282–86. ISCA.

Metz, Cade. 2013. Amazon's Invasion of the CIA Is a Seismic Shift in Cloud Computing. *Wired*. https://www.wired.com/2013/06/amazon-cia/.

Mills, Mara. 2010. Deaf Jam. *Social Text* 28 (1): 35–58.

Navarro, Joan, Ester Vidaña-Vila, Rosa Ma Alsina-Pagès, and Marcos Hervás. 2018. Real-Time Distributed Architecture for Remote Acoustic Elderly Monitoring in Residential-Scale Ambient Assisted Living Scenarios. *Sensors* 18 (8). https://doi.org/10.3390/s18082492.

Neville, Stephen. 2020. Eavesmining: A Critical Audit of the Amazon Echo and Alexa Conditions of Use. *Surveillance & Society* 18 (3): 343–56.

Obar, Jonathan A., and Anne Oeldorf-Hirsch. 2020. The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services. *Information, Communication & Society* 23 (1): 128–47.

Parker, James E.K. 2015. *Acoustic Jurisprudence: Listening to the Trial of Simon Bikindi*. Oxford: Oxford University Press.

Parker, James E.K., and Sean Dockray. 2023. 'All Possible Sounds': Speech, Music, and the Emergence of Machine Listening. *Sound Studies*, 1–29.

Paranoid, 2025. Products. https://paranoid.com/products

Reuters. 2024. Amazon Doubles down on AI Startup Anthropic with $4bn Investment. *The Guardian*, November 22. https://www.theguardian.com/technology/2024/nov/22/amazon-anthropic-ai-investment.

Roemelle, Brian. 2017. Casper. https://www.youtube.com/watch?v=8De_KxYt1pQ

Rohlicek, J.R. 1995. Word Spotting. *Modern Methods of Speech Processing*, 123–58. Boston, MA: Springer.

Rohlicek, J.R., W. Russell, S. Roukos, and H. Gish. 1989. Continuous Hidden Markov Modeling for Speaker-Independent Word Spotting. *International Conference on Acoustics, Speech, and Signal Processing*, 627–30.

Rose, Richard C. 1996. Word Spotting from Continuous Speech Utterances. *Automatic Speech and Speaker Recognition*, edited by Chin-Hui Lee, Frank K. Soong, and Kuldip K. Paliwal, 355:303–29. Boston, MA: Springer. https://doi.org/10.1007/978-1-4613-1367-0_13.

Sadowski, Jathan. 2019. When Data Is Capital: Datafication, Accumulation, and Extraction. *Big Data & Society* 6 (1): 205395171882054.

Salvador, Stan Weidner, Jeffrey Paul Lilly, Frederick V Weber, Jeffrey Penrod Adams, and Ryan Thomas. 2016. Wake Word Evaluation. US 9.275,637 B1, filed November 6, 2012, and issued March 1, 2016.

Sensory, 2025. Wake word. https://www.sensory.com/wake-word/

Soundhound. 2021. Why Your Brand's Voice Assistant Should Have a Custom Wake Word. *SoundHound*. 2021. https://www.soundhound.com/voice-ai-blog/why-your-voice-assistant-should-have-a-custom-wake-word-d9df31baff14/.

Soundhound, 2025a. Wake word. https://www.soundhound.com/voice-ai-products/wake-word/

Soundhound, 2025b. About. https://www.soundhound.com/about/

Srnicek, Nick. 2017. *Platform Capitalism*. London: Polity.

Statt, Nick. 2018. Amazon Adds Follow-up Mode for Alexa to Let You Make Back-to-Back Requests. *The Verge*. https://www.theverge.com/2018/3/9/17101330/amazon-alexa-follow-up-mode-back-to-back-requests.

Sterne, Jonathan. 2011. The Theology of Sound: A Critique of Orality. *Canadian Journal of Communication* 36 (2): 207.

Sterne, Jonathan, and Mehak Sawhney. 2022. The Acousmatic Question and the Will to Datafy: Otter. Ai, Low-Resource Languages, and the Politics of Machine Listening. *Kalfou* 9 (2): 288–306.

Stone, Brad. 2021. The Secret Origins of Amazon's Alexa. *Wired*. https://www.wired.com/story/how-amazon-made-alexa-smarter/.

Taylor, Pete. 2015. Edward Snowden Interview: 'Smartphones Can Be Taken Over.' *BBC News*, October 5. https://www.bbc.com/news/uk-34444233.

Titlow, John. 2017. Sonos's First Alexa-Powered Smart Speaker Is Finally Here. *Fast Company*. https://www.fastcompany.com/40476573/sonoss-first-alexa-powered-smart-speaker-is-finally-here.

Tuohy, Jennifer. 2022. How to Create an Alexa Routine on Your Amazon Echo Smart Speaker. *The Verge*. 2022. https://www.theverge.com/23312859/amazon-alexa-echo-routine-smart-speaker-how-to.

Turow, Joseph. 2021. *The Voice Catchers*. New Haven: Yale University Press.

Virilio, Paul. 2007. *The Original Accident*. London: Polity.

Vismann, Cornelia. 2013. Cultural Techniques and Sovereignty. *Theory, Culture & Society* 30 (6): 83–93.

Vlahos, James. 2019. *Talk to Me: Amazon, Google, Apple and the Race for Voice-Controlled AI*. New York: Penguin Random House.

Voicebot. 2018. Voice Assistant Consumer Adoption Report.

Waddell, Kaveh. 2016. Why Digital Assistants Are a Privacy Nightmare. *The Atlantic*. May 24. https://www.theatlantic.com/technology/archive/2016/05/the-privacy-problem-with-digital-assistants/483950/.

Wang, Joe, Rajath Kumar, Mike Rodehorst, Brian Kulis, and Shiv Naga Prasad Vitaladevuni. 2020. An Audio-Based Wakeword-Independent Verification System. *Interspeech 2020*, 1952–56. ISCA.

Wang, Yiming, Hang Lv, Daniel Povey, Lei Xie, and Sanjeev Khudanpur. 2021. Wake Word Detection with Streaming Transformers. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 5864–68.

Wark, McKenzie. 2019. *Capital Is Dead*. London: Verso.

Weise, Elizabeth. 2016. Bezos: 'Star Trek' Was Inspiration for Amazon Echo. *USA Today*. https://www.usatoday.com/story/tech/2016/05/18/jeff-bezos-amazon-transformers-washington-post/84548082/.

Wolfson, Sam. 2018. Amazon's Alexa Recorded Private Conversation and Sent It to Random Contact. *The Guardian*, May 24. https://www.theguardian.com/technology/2018/may/24/amazon-alexa-recorded-conversation.

Zuboff, Shoshana. 2018. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.