

Artificial intelligence & endangered languages

The widespread proliferation of virtual digital assistants, which use voice recognition, could increase language endangerment, especially for people from the North-east



One of the last people to speak Purum



A woman from the Monsang community

THONGKHOLAL HAOKIP

Artificial intelligence is changing the world rapidly and has impacted our life and imagination. Big tech companies are increasing their funding on research for the advancement of AI and the integration of such intelligent technology with their products. Indeed, the assistance of AI in transportation, manufacturing, healthcare, education, and media has not only reduced costs and improved productivity but also increased efficiency and quality.

As many benefits as there are to the advancement of AI, there are also inherent risks associated with it. Back in 2014 during an interview with the BBC, British theoretical physicist Stephen Hawking had warned, "The development of full artificial intelligence could spell the end of the human race. It would take off on its own, and re-design itself at an ever-increasing rate. Humans, who are limited by slow biological evolution, couldn't compete, and would be superseded". The development of autonomous weapons and robots, which use AI to eliminate threats, can wreak havoc on humanity.

Keeping aside the existential threat to humans from the uncontrolled development of AI, this technology at large could also pose a risk to those whose work is repetitive and routine in nature such as sweeping, cleaning, washing dishes, etc. It can adversely affect the poor, disenfranchised and marginal groups in different parts of the world and they would only feel the impact slowly and inconspicuously.

Discriminative AI

In recent years, there has been a voice-activated technology revolution which is greatly affecting how a lot of people live in different parts of the world. Many applications these days use AI and machine learning algorithms that analyse large amounts of data to recognise patterns and make decisions on their own. It helps in identifying highly personalised preferences and in turn eases marketing efforts. Virtual assistants such as Apple's Siri, Google Assistant and Amazon's Alexa are powered by AI in recognising speech. Through the recognition of words in speech, AI helps Netflix suggest movies, Spotify recommend songs, and Amazon push product promotions.

Despite such advances, studies have shown that "for people with accents – even the regional lilt, dialects and draws native to various parts of the United States – the artificially intelligent speakers can seem very different: inattentive, unresponsive, and even isolating". It has been indicated that within the U S "the wave of the future has a bias problem, and it's leaving them behind".

In July 2018, the CNN published a report titled "AI is hurting people of colour and the poor. Experts want to fix that". Heather Kelly, who was then with CNN Business, reported about how AI is contributing to greater racial bias and exclusion while also fundamentally changing the world.

She also pointed out how "facial recognition software has trouble identifying women of colour" at the MIT Media Lab.

A similar study that year by *The Washington Post*, in collaboration with two research groups, looked at the problem of smart speakers' accent imbalance by testing thousands of voice commands dictated by more than 100 people across nearly 20 cities. They found that the systems "showed notable disparities in how people from different parts of the U S are understood. People with Southern accents, for instance, were three per cent less likely to get accurate responses from a Google Home device than those with Western accents. And Alexa understood Midwest accents two per cent less than those from along the East Coast".

The study also pointed out that "people who spoke Spanish as a first language, for instance, were understood six per cent less often than people who grew up around California or Washington, where the tech giants are based". Another investigation conducted by ProPublica also revealed that "software used to sentence criminals is biased against black Americans".

The world today, including developing and underdeveloped countries, is being invaded by smart home technologies, which use virtual assistants. In attempting to command the virtual assistant, communication must be made in a dominant language only, otherwise it fails to recognise the command. It can lead to two things in order to use such applications. First, users from marginal communities need to learn dominant languages. Second, the accent has to be imitated for the proper recognition of the command. It will undoubtedly result in increasing homogenisation of languages and accent neutralisation.

Endangered languages in North-east India

On 19 February 2009, the United Nations Educational, Scientific and Cultural Organisation launched the electronic version of the new edition of its "Atlas of the World's Languages in Danger". The interactive digital tool provides updated data about approximately 2,500 endangered languages across the world and can be continuously supplemented, corrected and updated by users. They are further classified into vulnerable, definitely endangered, severely endangered and critically endangered languages. In a few years, many critically endangered languages will become extinct if appropriate measures are not taken to preserve them.

According to the Unesco Atlas, more than 200 languages have become extinct during the last 75 years, 538 are critically endangered, 502 severely endangered, 632 definitely endangered and 607 unsafe or vulnerable.

India has 197 endangered languages. In North-east India, out of about 220 languages spoken, 80 are facing a serious threat and 21 are on the verge of extinction. In Manipur,



Ch Lamtachao, the last surviving man who can sing folk songs in the Tarao language

for instance, at least five languages are listed as critically endangered. Those on the verge of extinction include the Purum language spoken by 276 people, Tarao by 700, Monsang by 1,270, Aimol spoken by 2,400, and Moyon by 2,270.

There are initiatives being taken by different universities and other institutes in the North-east to describe, document and digitise the endangered languages of the region. For instance, the Centre for Endangered Languages was established in 2014 at Tezpur University "with the aim of conducting substantial research on the lesser known and endangered languages of North-east India and to revitalise them with the direct and indirect institutional intervention". The University Grants Commission has recognised Tezpur University as a nodal centre for the cluster which also comprises Rajiv Gandhi University, Itanagar and Sikkim University.

Despite such institutional interventions to protect and preserve endangered languages, there are many forces at play which could still put them on the path to extinction. First, the tendency to neglect native languages and learn dominant languages for better employment prospects. Second, the onslaught of virtual technology and attempts to neutralise accents with the increasing use of voice assistants.

My personal experience

As indicated, the increasing use of virtual assistants in different technological products is not only discriminative; it can also lead to increasing homogenisation. Accents are a part of cultural identities that need to be cherished rather than neutralised. And with such acknowledgement of diversity, there should be an increasing awareness campaign to celebrate it.

Recent advances in natural lan-

guage processing, which allows machines to read and analyse large amounts of natural language data and understand human languages including the contextual nuances within them, can be an answer to reduce, if not totally remove, this discrimination which poses a threat to endangered languages. To demonstrate the advancement in natural language processing, let me share my experience in the last three years in this field.

The basic idea of this article was presented in a global symposium on "Artificial Intelligence in Governance and Disaster Management" held at the Special Centre for Disaster Research, Jawaharlal Nehru University, in March 2019. Before presenting the paper, I tested the accent recognition of Siri on my iPad. At the first attempt, I spoke in my original North-east accent and Siri recognised about 60 per cent of what was said. At the second attempt, I moderately imitated the American accent and it was able to recognise about 80 per cent of the speech. At the third attempt, I completely imitated the American accent and Apple's virtual assistant application was able to recognise my commands cent per cent. During the symposium, the product manager of Alexa AI, who had come all the way from Boston in the U S, agreed on the concerns that I shared and promised to work on reducing such accent discrimination.

Today different AI-assisted home applications can be used to improve the quality of life and increase comfort. Google Assistant or Alexa can be asked to turn on smart televisions. Such televisions, particularly those which use android operating systems, have voice search features. Google Assistant can communicate with the user and a voice command can be made to play the desired video or select new videos in popular Over The Top media services.

Since the coronavirus disease was declared a pandemic by the World Health Organisation and restrictions were imposed by governments, classes and meetings in academic institutions have been conducted online. Google Meet and Zoom video-communication service were commonly used for such classes and meetings.

In Google Meet, speeches can be translated into text through "turn on captions". It is designed "to help participants who may be deaf or hard of hearing". During faculty meetings, I tested Google's speech-to-text technology which provided live captions and found that the speeches were recognised almost cent per cent.

With all such technological advancements that use voice command, and the integration of natural language processing in AI and machine learning, will systems be adaptive enough or quick in recognising accents of minority communities?

Even though technology is advancing faster, including natural language processing, they are restricted to areas that have market value and profit. It is highly doubtful and unlikely that AI and its natural language processing will be advanced to such a level in areas that have no market value. Given that, accents must be imitated in dominant languages by people categorised in the endangered languages section, like millions from the North-east.

As such the increasing use of modern applications and services such as search by voice or Google Translate and others could hasten the extinction of endangered languages, until natural language processing is advanced to a level that recognises all accents, including those from critically endangered languages.

The writer is assistant professor, Centre for the Study of Law and Governance, Jawaharlal Nehru University, New Delhi

PLUS POINTS

Protective measures



Assam chief minister Sarbananda Sonowal, while recently interacting with district magistrates, top health officials and other stakeholders associated with Covid-19 management, emphasised on creating micro-containment zones along with extensive testing to combat the resurgence of the pandemic.

Commenting that the second wave of the pandemic has pushed up the daily surge of positive cases in many parts of India unlike Assam, Sonowal directed state officials to enforce protective guards to help check the spike of positive cases so that a major health disaster can be averted. Asking the DMs to support the Prime Minister's call to make "Tika Utsav" a success, he stated that inoculation drives in all districts should be intensified for eligible beneficiaries.

At the same time, Sonowal asked them to use their machineries to speed up testing, precisely in important locations like railway stations, airports and other places of mass concentration. Insisting on more public awareness over corona-screening and vaccination, he asked the DMs to strengthen their existing infrastructure like corona care centres, isolation beds, and oxygen cylinders to face any exigency. Extending Rongali (Bohag) Bihu greetings to the people, he appealed to everyone to adhere to corona-appropriate behaviour during the festival to ensure health and happiness for all.

-NANA THAKURIA

Virus protocols



The Mizoram government has ordered closure of schools for students up to Class VIII following a surge in Covid-19 cases, says a report in the Guwahati-based *Assam Tribune*. Fresh guidelines issued last week said that the state government has again banned night church service and extended the duration of night curfew by two hours in all the district headquarters, including the state capital as the second wave of Covid-19 has hit Mizoram. The curfew will now be in force from 8 p.m. to 4 a.m. instead of 10 p.m. to 4 a.m.

The ongoing examinations for classes IX to XII, however, will continue as scheduled. Schools for students of classes V to VIII had reopened on 1 March and for students of classes III and IV on 1 April. Classes in universities and colleges, however, will be subject to the decision taken by the state's higher and technical education department, the order said.

A PTI report from Aizawl adds: Churches, cinema halls, auditoria, community halls, picnic spots and other places of entertainment like resorts are allowed to open with 50 per cent of their sitting capacity. The number of attendees at a funeral, wedding and any social or political gathering was fixed at 50, the order said. Shops, hotels, restaurants and other business establishments will open under strict adherence to Covid-19 protocols.

The guideline said that all returnees from other states would undergo rapid antigen tests and those testing negative would be placed under home isolation for seven days. However, they will not require home quarantine if they test negative in RT-PCR and TrueNat laboratories.

Returnees from abroad or new Covid-19 variant-affected states will have to be quarantined at designated facilities. But if they have already stayed 10 days in the country after arriving and tested negative during screenings at the entry points here, they will be allowed seven days home isolation.

Quarantine is not mandatory for people who visit the state for a short stay not longer than 96 hours provided they possess a Covid-19 negative certificate not older than 96 hours upon their arrival.

