ஆய்வறிக்கை

cdt Research

சமூக ஊடகங்களில் தமிழ் தகவல்களை ஒழுங்குபடுத்துதல்

அலியா பாட்டியா
மோனா எல்ஸ்வா

பிப்ரவரி 2026

CENTER FOR DEMOCRACY & TECHNOLOGY

ஜனநாயகம் மற்றும் தொழில்நுட்ப மையம் (CDT) என்பது இந்த டிஜிட்டல் யுகத்தில் சமூக உரிமைகளை முன்னேற்றுவதற்காகப் போராடும் முன்னணி பாரபட்சமற்ற, இலாப நோக்கற்ற அமைப்பாகும். சமத்துவம் மற்றும் ஜனநாயக மதிப்புகளை மையமாகக் கொண்டு தொழில்நுட்பக் கொள்கை, நிர்வாகம் மற்றும் வடிவமைப்பை நாங்கள் வடிவமைக்கிறோம். 1994 இல் நிறுவப்பட்ட இந்த அமைப்பு, இணையத்தின் ஆரம்ப நாட்களிலிருந்தே டிஜிட்டல் உரிமைகளுக்கு குரல் கொடுக்கும் வக்கீலாக இருந்து வருகிறது. இந்த அமைப்பின் தலைமையகம் வாஷிங்டன், டி.சி.யில் உள்ளது, மேலும் பெல்ஜியத்தின் பிரஸ்ஸல்ஸில் ஒரு ஐரோப்பிய அலுவலகத்தைக் கொண்டுள்ளது.

அலியா பாட்டியா
மோனா எல்ஸ்வா

நூலாசிரியர்கள்

சமூக ஊடகங்களில் தமிழ் தகவல்களை ஒழுங்குபடுத்துதல்

CDT ஆய்வறிக்கை

அலியா பாட்டியா
மோனா எல்ஸ்வா

பங்களிப்பு செய்தவர்கள்

சமீர் ஜெயின், மைக்கல் லூரியா, தேவன் எல். ஹங்கர்சன், தனராஜ் தாக்கூர், வருண் ராவ் மற்றும் கேட் ருவான்.

நன்றியுரை

முதலில் இணைய மற்றும் சமூக மையத்திலுள்ள அற்புதமான குழுவிற்கு நன்றி தெரிவிக்க விரும்புகிறோம். அஃபெஃப் அப்ரூகி, ஜுவான் கார்லோஸ் லாரா, நஞ்சிரா சாம்புலி மற்றும் ஜிலியன் யோர்க் ஆகியோரைக் கொண்ட எங்கள் ஆலோசனைக் குழுவிற்கு நன்றி. தருணிமா பிரபாகர் மற்றும் ஆத்மன் வைத்யா ஆகியோருக்கும் நன்றி கூறுகிறோம். இந்த அறிக்கையில் உள்ள அனைத்து கருத்துக்களும் CDT இன் கருத்துக்களாகும்.

எங்களுடைய அழகான அட்டைப்படம் மற்றும் வடிவமைப்பு பணிக்காக ஓஷீன் சிவாவுக்கும், இந்த திட்டத்திற்கான படைப்பாற்றலை இயக்கிய டிம் ஹோக்லாண்டிற்கும் சிறப்பான நன்றிகள் உரித்தாகட்டும்.

ஆய்வில் பங்கேற்று எங்களுடன் தங்களது நேரத்தையும், ஆழமான கருத்துக்களையும் பெருந்தன்மையுடன் பகிர்ந்து கொண்ட பங்கேற்பாளர்களுக்கு நாங்கள் நன்றி கூறுகிறோம். கூறப்பட்ட கருத்துக்கள் அனைத்தும் படிப்பதற்கு எளிதாக இருக்கும் வகையில் சற்றுத் திருத்தப்பட்டுள்ளன.

இணைய சமூக அறக்கட்டளையினால் வழங்கப்பட்ட உதவித்தொகையின் மூலமாகவே இந்தப் பணி சாத்தியமானது.

பரிந்துரைக்கப்பட்ட மேற்கோள்: பாட்டியா, ஏ. & எல்ஸ்வா, எம். (2025). சமூக ஊடகங்களில் தமிழ் உள்ளடக்கத்தை ஒழுங்குபடுத்துதல். ஜனநாயகம் மற்றும் தொழில்நுட்ப மையம்.. https://cdt.org/insights/moderating-tamil-content-on-social-media/

இந்த அறிக்கையில் உள்ள சான்றாதாரங்களில் அசல் இணைப்புகள் (Links) மற்றும் Perma.cc சேவையால் காப்பகப்படுத்தப்பட்ட மற்றும் சுருக்கப்பட்ட இணைப்புகள் அடங்கும். Perma.cc இணைப்புகளில், மீட்டெடுக்கப்பட்ட தேதி மற்றும் காப்பகப்படுத்தல் பற்றிய தகவல்களும் உள்ளடக்கப்பட்டுள்ளன.

இந்த அறிக்கை Creative Commons Attribution-Sharealike 4.0 International Licence இன் கீழ் உரிமம் பெற்றது.

பொருளடக்கம்

அறிமுகம் **5**

குறைந்த வளம் கொண்ட மொழியாகத் தமிழ் **8**

வரலாற்று ரீதியாக அரசியல்மயமாக்கப்பட்ட மொழியாகத் தமிழ் **9**

பிரதான கண்டுபிடிப்புகள் **11**

1. தமிழ் பேசுபவர்கள், சில சமயங்களில் ஒழுங்குபடுத்தல் செயல்முறையைத் தவிர்ப்பதற்காக, குறியீட்டு கலப்புத் தமிழ், கணினி வழித் தமிழ் மற்றும் ஒலிபெயர்ப்பு செய்யப்பட்ட தமிழில் சமூக ஊடக சேவைகள் மற்றும் இணைய தளங்களின் கலவையைப் பயன்படுத்துகின்றனர். 11

2. உலகமயமாக்கப்பட்ட மற்றும் பகுதிசார்ந்த அணுகுமுறைகள்: சமூக ஊடகங்கள் மற்றும் இணைய தளங்கள் தமிழ் மொழியில் தகவல்களைக் கண்காணித்தல் மற்றும் ஒழுங்குபடுத்துவதில் கலவை முறையிலான அணுகுமுறைகளைப் பின்பற்றுகின்றன. 14

3. பல பயனாளர்கள் சில நேரங்களில் ஒழுங்குபடுத்தும் செயல்முறையை அரசியல் காரணங்களுக்காக தங்கள் குரல்களை "அமைதிப்படுத்தும்" முயற்சியாகக் கருதுகின்றனர். 23

4. தமிழ் இயற்கை மொழி செயலாக்க (NLP) ஆராய்ச்சி மற்றும் தமிழ் தகவல் பகுப்பாய்வு திறன்களில் முன்னேற்றங்கள் இருந்தபோதிலும், குறைந்த நிதி முதலீடு மற்றும் தமிழ் இயற்கை மொழி செயலாக்க நிபுணர்களிடம் போதுமான ஈடுபாடு இல்லாதது போன்றவற்றின் காரணமாக, சமூக ஊடக நிறுவனங்கள் தமிழுக்கான தானியங்கி மதிப்பீட்டுக் கருவிகளைப் பயன்படுத்துவதில் மெதுவாகச் செயல்படுகின்றன. 29

பின் இணைப்பு **42**

முறைகள் 42

சான்றாதாரங்கள் **45**

அறிமுகம்

தமிழ் என்பது ஒரு நீண்ட வரலாற்றைக் கொண்ட மொழியாகும். உலகளவில் 80 மில்லியனுக்கும் அதிகமான மக்களால், அதாவது உலக மக்கள்தொகையில் 1% க்கும் அதிகமானவர்களால் பேசப்படும் இந்த மொழியில் உள்ள ஆரம்பக்கால கல்வெட்டுகள் கி.மு 5 ஆம் நூற்றாண்டைச் சேர்ந்தவை (Murugan, 2024). இந்த மொழி இந்தியாவிலும் (பெரும்பாலும் தமிழ்நாடு மற்றும் புதுச்சேரியிலும்), இலங்கையிலும், மலேசியா, தாய்லாந்து, கனடா, இங்கிலாந்து, அமெரிக்கா மற்றும் அதற்கு அப்பால் உள்ள புலம்பெயர் சமூகங்களிலும் பரவலாகப் பேசப்படுகிறது. தமிழ் மொழியின் பரவலான பயன்பாடு இருந்தபோதிலும், முக்கிய சமூக ஊடக தளங்கள் தமிழில் தகவல்களை எவ்வாறு ஒழுங்குபடுத்துகின்றன என்பது குறித்த புரிதல் குறைவாகவே உள்ளது. இந்த அறிக்கை தமிழ் பயனாளர்களின் இணைய அனுபவங்களை ஆராய்கிறது. மேலும், இந்த மொழிக்குத் தொடர்ச்சியாகத் தகவல் ஒழுங்குபடுத்தல் செயல்முறைகளைப் பயன்படுத்துவதில் உள்ள சவால்களையும் ஆராய்கிறது.

இந்த அறிக்கை உலகளாவிய தெற்கத்திய நாடுகளின் குறைந்த வளம் மற்றும் பூர்வீக மொழிகளில் உள்ள தகவல் ஒழுங்குபடுத்துதலை ஆராயும் தொடரின் ஒரு பகுதியாகும். குறைந்த வளம் கொண்ட மொழிகள் என்பது மாதிரிகளைப் பயிற்றுவிப்பதற்குப் போதுமான உயர்தர தரவுகள் காணப்படாத மொழிகளாகும். இதனால் வலுவான தகவல் ஒழுங்குபடுத்தல் அமைப்புகளை, குறிப்பாகத் தானியங்கி அமைப்புகளை உருவாக்குவது கடினமாகிறது (Nicholas & Bhatia, 2023) இந்தத் தொடரில் நடத்தப்பட்ட முந்தைய வழக்கு ஆய்வுகளில், உயர்தர மற்றும் சொந்த மொழி அடிப்படையிலான தரவுத்தொகுப்புகள் இல்லாதது மக்ரெபி அரபு மற்றும் கிஸ்வாஹிலி மொழிகளில் இருந்த தகவல்களைத் திறமையாகவும் துல்லியமாகவும் ஒழுங்குபடுத்துவதைத் தடுத்தது என்பதைக் கண்டறிந்தோம் (Elswah, 2024a; Elswah, 2024b) சீரற்ற மற்றும் துல்லியமற்ற தகவல் ஒழுங்குபடுத்தலானது, உலகளாவிய தெற்கத்திய நாடுகளின் பயனாளர்களிடையே நம்பிக்கையைக் குறைக்கிறது, மேலும் அவர்கள் தங்களைச் சுதந்திரமாக வெளிப்படுத்திக்கொள்வதையும் தகவல்களை அணுகுவதையும் கட்டுப்படுத்துகிறது.

இந்த அறிக்கை இணையத்தில், குறிப்பாக மேற்கத்திய மற்றும் இந்திய நிறுவனங்களால் நடத்தப்படும் பிரபலமான சமூக ஊடக தளங்கள் மற்றும் இணைய தளங்களில் தமிழ் பேசுபவர்களின் அனுபவங்களை விரிவாக ஆராய்கிறது. குறிப்பாக இந்தியா மற்றும் இலங்கையில் ஜனநாயக பின்னடைவு மற்றும் பேச்சு மற்றும் குடிமை பங்கேற்பின் மீதான அதிகரித்து வரும் அடக்குமுறை சூழ்நிலையில் மோசமான தகவல் ஒழுங்குபடுத்தல் குறித்த தமிழ் மொழி பேசுபவர்களின் உணர்வுகளையும் அதன் தாக்கத்தையும் நாங்கள் எடுத்துக்காட்டுகிறோம் (Vesteinsson, 2024; Nadaradjane, 2022). இறுதியாக, இந்த வழக்கு ஆய்வில் வெளிப்படுவது என்னவென்றால், தமிழ் பேசுபவர்கள் அதிகப்படியான தகவல் கட்டுப்பாடுகளை எதிர்கொள்வதாக உணரும் அதே வேளையில், குறைவாக ஒழுங்குபடுத்தப்பட்ட வெறுப்புப் பேச்சு நிறைந்த பதிவுகளையும் எதிர்கொள்கிறார்கள் என்பது ஒரு முழுமையாக்கப்படாத தகவலே ஆகும்.

நாங்கள் ஒரு கலவை முறையிலான அணுகுமுறையைப் பயன்படுத்தினோம். அதில் இந்தியாவிலும் இலங்கையிலும் அடிக்கடி சமூக ஊடகங்களைப் பயன்படுத்தும் 147 பயனாளர்களின் இணையக் கருத்தாய்வு; தகவல் கண்காணிப்பாளர்கள், தகவல் உருவாக்குபவர்கள், தளங்களினுடைய நம்பிக்கை மற்றும் பாதுகாப்பு பிரதிநிதிகள் மற்றும் டிஜிட்டல் உரிமைகள் ஆதரவாளர்களுடன் நடத்தப்பட்ட 17 ஆழமான நேர்காணல்கள்; மற்றும் தமிழ் இயந்திர கற்றல் மற்றும் தரவு நிபுணர்களுடன் ஒரு வட்டமேசை விவாதம் ஆகியவை அடங்கும். பயன்படுத்தப்பட்ட முறைகள் பின்னிணைப்பில் விவரிக்கப்பட்டுள்ளன. இந்த முறைகளின் அடிப்படையில், நாங்கள் கண்டறிந்த தகவல்கள் பின்வருமாறு:

1. தமிழ் பேசுபவர்கள் பல்வேறு மேற்கத்தியச் சமூக ஊடக தளங்களையும் இந்தியத் தளங்களையும் பயன்படுத்துகின்றனர். எங்களுடைய கருத்தாய்வு, தமிழ் பேசுபவர்களிடையே மேற்கத்தியச் சமூக ஊடகத் தளங்கள் மிகவும் பிரபலமாக உள்ளன என்பதையும், அதே நேரத்தில் இந்தியாவின் டிக்டோக் தடை காரணமாக அதற்குப் பதிலாக பயன்படுத்தக்கூடிய உள்ளூர் மாற்றுத் தளங்கள் பிரபலமடைந்து வருகின்றன என்பதையும் சுட்டிக்காட்டுகிறது. இணையத்தில், தமிழ் பேசுபவர்கள் தகவல் ஒழுங்குபடுத்துதலைத் தவிர்ப்பதற்காக உத்திகளைப் பயன்படுத்துகின்றனர். **அவற்றுள் "அல்கோஸ்பீக்" அல்லது கணினி வழித் தகவல்தொடர்புகளைப் பயன்படுத்துதல், மேலும், மற்ற நேரங்களில், இலகுவாகவும் வசதியாகவும் இருப்பதற்காக லத்தீன் எழுத்துக்களைப் பயன்படுத்தி குறியீட்டு கலப்பு மற்றும் ஒலிபெயர்ப்பு செய்யப்பட்ட தமிழைப் பயன்படுத்துதல் போன்றவை அடங்கும்.** இந்த உத்திகள் ஒழுங்குபடுத்தல் செயல்முறையைச் சிக்கலாக்குகின்றன.

2. தொழில்நுட்ப நிறுவனங்கள் இணையத்தில் தமிழ் தகவல்களை ஒழுங்குபடுத்துவதற்காக பல்வேறு அணுகுமுறைகளைப் பின்பற்றுவதற்கு முயற்சிக்கின்றன. ஆனால், பெரும்பாலும் உலகளாவிய அல்லது பகுதிசார்ந்த அணுகுமுறைகளையே கடைப்பிடிக்கின்றன. உலகளாவிய அணுகுமுறை உலகெங்கிலும் உள்ள அனைத்து பயனர்களுக்கும் ஒரே கொள்கைகளைப் பயன்படுத்துகிறது, மேலும் மொழியியல் அல்லது வட்டார நிபுணத்துவத்தின் அடிப்படையில் பணியமர்த்தப்படாத மதிப்பீட்டாளர்கள் மற்றும் கொள்கை உறுப்பினர்களை நம்பியுள்ளது. மதிப்பீட்டாளர்களுக்கு உலகம் முழுவதிலுமிருந்து தகவல் ஒதுக்கப்படுகிறது. இதற்கு நேர்மாறாக, உள்ளூர் அணுகுமுறையானது தமிழ் மொழி சார்ந்த வழிகாட்டுதலைப் பூர்த்தி செய்வதற்காக சில கொள்கைகளை வடிவமைக்கிறது. மேலும், தகவல்களை ஒழுங்குபடுத்துவதில் அதிக தமிழ் பேசுபவர்களை நம்பியுள்ளது. சில இந்திய நிறுவனங்கள் கலப்பின அணுகுமுறையைப் பயன்படுத்துகின்றன, சந்தர்ப்பத்திற்கு ஏற்றவாறு அடிக்கடி தமிழ் பேசுபவர்களுக்காக பகுதிசார்ந்த மாற்றங்களைச் செய்கின்றன.

3. தமிழ் பேசுபவர்கள், மற்றவர்களைப் போலவே, வழக்கமாகச் சீரற்ற **சீரற்ற தகவல் ஒழுங்குபடுத்துதலை எதிர்கொள்கின்றனர், அவர்களின் முதன்மை மொழி ஆங்கிலம் அல்ல என்பதே இதற்குக் காரணம் என**

அவர்கள் நம்புகிறார்கள். ஒருபுறம், ஒழுங்குபடுத்தப்படாத தகவல்களால் தேவை இல்லாத, குறி வைக்கப்பட்ட வெறுப்பு பதிவுகளை மக்கள் காண நேரிட்டாலும், மறுபுறம், குறிப்பாகத் தமிழ் மொழியைக் குறிவைத்து நியாயமற்ற முறையில் மிகைப்படுத்தப்பட்ட ஒழுங்குபடுத்தல் மேற்கொள்ளப்படுகிறது எனச் சந்தேகிக்கிறார்கள்.

4. கருத்தாய்வில் பதிலளித்தவர்களில் பெரும்பாலானவர்கள் அரசியல் ரீதியாக ஊக்கமளிக்கப்படும் தகவல் ஒழுங்குபடுத்தல் குறித்து கவலை கொண்டுள்ளனர். மேலும் தகவல் நீக்கங்கள் மற்றும் கட்டுப்பாடுகள் இணையத்தில் தங்கள் குரல்களை அடக்கப் பயன்படுத்தப்படுகின்றன என்று நம்புகிறார்கள், குறிப்பாக அவர்கள் அரசியல் பற்றிப் பேசும்போது இது நடத்தப்படுவதாக அவர்கள் நம்புகிறார்கள். **ஒரு சில பயனாளர்கள், குறிப்பாகத் தமிழ்ச் சமூகத்தால் பொதுவாகப் பயன்படுத்தப்படும் அல்லது தமிழ்ச் சமூகத்துடன் தொடர்புடைய சில சொற்கள் அல்லது சின்னங்களைப் பயன்படுத்தும் போது, "நிழல் தடை" (ஷேடோ பான்) அல்லது தளங்களால் பல்வேறு வகையான தெளிவற்ற, வெளிப்படுத்தப்படாத ஒழுங்குபடுத்தல் முடிவுகள் எடுக்கப்படுவதை அனுபவிப்பதாகச் சந்தேகிக்கின்றனர்.**

5. ஒரு துடிப்பான தமிழ்க் கணினி சமூகம் இருந்தபோதிலும், **இந்தியச் சமூக ஊடக நிறுவனங்களுக்குப் போதியளவு அணுகக்கூடிய வளங்கள் இல்லாமை, விருப்பமின்மை மற்றும் நிதி கட்டுப்பாடுகள் காரணமாகத் தமிழில் தானியங்கி ஒழுங்குபடுத்துதலில் செய்யப்படும் முதலீடு இன்னும் கணிசமாகக் குறைவாகவே உள்ளது.**

குறைந்த வளம் கொண்ட மொழியாகத் தமிழ்

ஒரு குறிப்பிட்ட மொழியில் தரவுத்தொகுப்புகள் மற்றும் டிஜிட்டல் மயமாக்கப்பட்ட எழுத்துக்கள் காணப்படுமாயின் "வளங்களைக் கொண்டிருப்பதாகக்" கருதப்படும். இது கணினி அறிவியல் துறையில் ஒரு அளவீடு ஆகும். மேலும் இது தகவல் கொள்கைகளை செயல்படுத்தவும் அவற்றை மீறும் தகவல்களைக் கண்டறியவும் தானியங்கி கருவிகளைப் பயிற்றுவிப்பதற்கும் இது முக்கியமானது (Nicholas & Bhatia, 2023). உலகில் அதிகம் பேசப்படும் மொழிகளில் 17வது இடத்தில் இருந்தாலும், தமிழ் குறைந்த வளம் கொண்ட மொழியாகக் கருதப்படுகிறது. மேலும், பரவலாகப் பேசப்படும் பிற மொழிகளை விடக் கணிசமாகக் குறைவான உயர்தர தரவுத்தொகுப்புகள் மற்றும் தானியங்கி தகவல் ஒழுங்குபடுத்தல் திறன்களைக் கொண்டுள்ளது. மற்ற இந்திய மொழிகளைக் காட்டிலும், தமிழ் குறைந்த வளம் கொண்ட மொழியாகக் காணப்படுகிறது; இந்தி மற்றும் வங்காளம் போன்ற இந்தோ-ஐரோப்பிய மொழிகள் தமிழ் அல்லது பிற திராவிட மொழிகளை விட அதிக டிஜிட்டல் வளங்களைக் கொண்டுள்ளன (Joshi et. al, 2022; Gala, Chitale et al., 2023).[1] குறைந்த வளம் கொண்ட மொழிகளில் காணப்படும் ஆராய்ச்சி மற்றும் தரவுத்தொகுப்புகளின் பற்றாக்குறையானது, தகவல் ஒழுங்குபடுத்தலுக்கான தானியங்கி கருவிகளை உருவாக்குதல், சோதித்தல் மற்றும் மேம்படுத்துவதில் சவால்களை ஏற்படுத்துகிறது. இவை பிரதானமாக ஆங்கிலம் மற்றும் மேற்கத்தியச் சூழல்களில் பொதுவாகப் பேசப்படும் பிற உயர் வள மொழிகளுக்கு ஏற்றவாறு வடிவமைக்கப்படுகின்றன (Shahid & Vashistha, 2023; Nicholas & Bhatia, 2023b; Rowe, 2022).

தமிழின் தனித்துவமான மற்றும் சிக்கலான வடிவம் ஒழுங்குபடுத்தும் செயல்முறையைக் குறிப்பாகத் தானியங்கி ஒழுங்குபடுத்தும் செயல்முறையைச் சிக்கலாக்குகிறது. தமிழ் என்பது ஒரு இரட்டை வழக்கு மொழியாகும், அதாவது இது பல வேறுபாடுகளைக் கொண்டுள்ளது. பேச்சுத் தமிழை விட எழுத்துத் தமிழில் அதிக முறைப்படுத்தப்பட்ட விதிகளும் பெரும்பாலும் வேறுபட்ட சொற்களும் உள்ளன. தற்கால பேச்சுத் தமிழில் பெரும்பாலும் நியமமான இலக்கணம் அல்லது அகராதி இல்லை (Nanmalar, Vijayalakshmi & Nagarajan, 2024), மேலும் பேச்சாளர் இருக்கும் பகுதியைப் பொறுத்து அல்லது அவர்களின் கல்வி நிலையைப் பொறுத்து மாறுபடும் (Schiffman, 1996). இலங்கைத் தமிழ், தமிழ்நாடு மற்றும் இந்தியாவில் பேசப்படும் தமிழிலிருந்து வேறுபடுகிறது. எங்களுடன் பேசிய சில NLP நிபுணர்கள், தமிழ் தரவுத்தொகுப்புக்களில் இந்தியர்கள் பேசும் மற்றும் எழுதும் தமிழே அதிகமாக இடம்பெறுவதாக வாதிடுகின்றனர்.

தமிழ் ஒரு ஒட்டுநிலை மொழியாகவும் இருக்கிறது, அதாவது மொழியின் வசன அமைப்பானது சொற்களை ஒன்றுடனொன்று இணைத்துச் சூழல் சார்ந்த மற்றொரு சொல்லை உருவாக்கும் தன்மையைக் கொண்டிருக்கிறது. இந்த வழியில், தமிழ் மொழியானது ''கணிதக்கழகப்போராட்டமுயற்சி'' போன்ற நம்பமுடியாத வித்தியாசமான சொற்களைக் கொண்டுள்ளது. இந்தச் சொல்லின் அர்த்தம் "கணிதம் கற்பிக்கும் ஒரு நிறுவனத்தில் செய்யப்படுகின்ற ஒரு கலவர முயற்சி" என்பதாகும். இந்த பண்பு மொழியை டிஜிட்டல் மயமாக்குவதை சிக்கலாக்குகிறது. ஏனெனில், பல சூழல் சார்ந்த சொற்கள் பெரும்பாலும் தரவுத்தொகுப்புகளில் குறிப்பிடப்படுவதில்லை.

1 வருடாந்திர கணினி மொழியியல் சங்க மாநாட்டில் சமர்ப்பிக்கப்பட்ட கட்டுரைகளில், இந்தி 16 முறை குறிப்பிடப்பட்டதுடன், தமிழ் ஒரு முறை மட்டுமே குறிப்பிடப்பட்டது (ACL Rolling Review Dashboard, n.d).

வரலாற்று ரீதியாக அரசியல்ம -யமாக்கப்பட்ட மொழியாகத் தமிழ்

தமிழ் ஆவணக்காப்பாளர்கள் தமிழ் மொழி வரலாற்று ரீதியாக ஓரங்கட்டப்படுவதாகவும், அரசியல் ரீதியாக எதிர்ப்புக்குள்ளாவதாகவும், இது குறிப்பாக இலங்கையில் நடைபெறுகிறது எனவும் கூறுகின்றனர் (Dassanayake, 2024). 1956 ஆம் ஆண்டு இலங்கையில் 'சிங்களம் மட்டும்' என்ற சட்டம் நிறைவேற்றப்பட்டதனால் நாட்டில் தேசிய அளவில் அங்கீகரிக்கப்பட்ட ஒரே மொழியாகச் சிங்கள மொழி மாறியது, இது தமிழ் எழுத்துப் பதிப்புகள் உருவாக்கப்படுவதைத் தடுத்ததோடு, தமிழ் இலக்கியங்களை இறக்குமதி செய்வதைக் கூட குற்றமாக மாற்றியது (Anandakugan, 2020). 1983 க்கும் 2009 க்கும் இடையில் நடந்த உள்நாட்டுப் போரின் போது நடைமுறைப்படுத்தப்பட்ட மொழிக் கொள்கைகள் தமிழ் பயன்பாட்டைக் குறைத்ததாக அகதிகளுக்கான ஐக்கிய நாடுகளின் உயர் ஆணையம் கூறுகிறது (UNHCR IRIN, 2012). 1981 ஆம் ஆண்டு, ஆசியாவின் மிகப்பெரிய நூலகங்களில் ஒன்றாகக் கருதப்பட்ட யாழ்ப்பாண நூலகம் எரிக்கப்பட்டதில், இலங்கைத் தமிழ் எழுத்துக்கள் மற்றும் கலைப்பொருட்களின் முக்கியமான தொகுப்பு அழிக்கப்பட்டது.[2] யாழ்ப்பாண நூலகத்தின் முன்னாள் தலைமை நூலகர், "ஆறு அறைகள் நிறைந்த 97,000 புத்தகங்கள் சாம்பலாகிவிட்டன" என்று கூறினார் (McCarthy, 2015). இதன் விளைவாக, இலங்கைத் தமிழில் மற்ற தமிழ் பேச்சுவழக்குகளை விடக் குறைவான வளங்களே காணப்படுகின்றன.

இந்தியாவிலுள்ள தமிழ் பேசுபவர்களும் தமிழ் மொழியின் பெருக்கத்திற்கு அச்சுறுத்தல் இருப்பதாக வலியுறுத்துகின்றனர், மேலும் அவர்கள், பொது நிறுவனங்கள் இரண்டு ஆதிக்க தேசிய மொழிகளான இந்தி மற்றும் ஆங்கிலத்தில் செயல்பட வேண்டும் என்று சிலர் கூறும் முன்மொழியப்பட்ட சட்டங்களை நீண்ட காலமாக எதிர்த்துள்ளனர். இந்த மசோதாக்கள் தங்கள் மொழியியல் மற்றும் கலாச்சார பாரம்பரியத்தை அழிப்பதாக வாதிடுகின்றனர் (Venkatachalapathy, 2022; Times of India, 2024); மொழி ஆர்வலர்கள் இந்த மசோதாக்கள் நாட்டின் பன்மொழித் தன்மையைத் தரைமட்டத்திற்கு வீழ்த்துவதாகக் கருதுகின்றனர், மேலும் பலர் பரவலான போராட்டங்களையும் நடத்தியுள்ளனர் (Subramaniam, 2024).[3]

குறிப்பாக சில இந்திய அரசு அதிகாரிகள், இந்தியாவில் தமிழ் மொழிப் பயன்பாடு ஆபத்தில் இருப்பதாகவும், இதனால் அவர்கள் 20 ஆம் நூற்றாண்டின் முற்பகுதியில் இருந்து மொழிக் கொள்கைகள் தொடர்பான முயற்சிகளை எதிர்த்து வருவதாகவும் தெரிவித்தார் (Sivapriyan, 2025; Venkatachalapathy, 2022; Times of India, 2024).இதற்குப் பதிலளிக்கும் விதமாக, தமிழ் மொழிக்கான பொது டிஜிட்டல் மயமாக்கப்பட்ட அணுகலை அதிகரிப்பதற்காகத் தமிழக அரசு தமிழ் இலக்கியத்தை தேசியமயமாக்க முயன்றுள்ளது (Digital Tamil Studies,n.d).

2 நாங்கள் பேட்டி கண்ட ஒரு தமிழ் ஆவணக்காப்பாளர் யாழ்ப்பாண நூலகம் எரிக்கப்பட்டது, "தமிழரின் கலாச்சார அடையாளம், பாரம்பரியம் மற்றும் தொடர்ந்து எதிர்க்கும் மனவுறுதி என்பவற்றை அழிப்பதற்காக முறையாகத் திட்டமிடப்பட்ட ஒரு அணுகுமுறை" என்று கூறினார்.

3 அரசியல் கட்சிகளும் ஆர்வலர்களும் நீண்ட காலமாக இந்த முயற்சிகளை எதிர்த்து வருகின்றனர். இந்தி-முதல்மொழி எனும் மொழிக் கொள்கைகளை எதிர்க்கும் முதலாவது ஆவணப்படுத்தப்பட்ட எதிர்ப்பு நடவடிக்கைகள் 1938 இல் நடந்தன. சமீபத்தில் 2022 இல் மற்றும் கடந்த ஆண்டு செப்டம்பர் 2024 விலும் இவை நடைபெற்றன.

தமிழ் மொழி ஆபத்தில் உள்ளது அல்லது தவறாக நடத்தப்படுகிறது என்ற தமிழ் மொழி பேசுபவர்களின் பரவலான கருத்து, அதனுடன் தொடர்புடைய இணையம் மற்றும் அச்சு ஆகிய இரண்டிலும் தமிழ் மொழியைப் பாதுகாக்க வேண்டும் என்ற விருப்பம் மற்றும் உயர்தர தமிழ் வளங்களின் பற்றாக்குறை ஆகியவற்றைப் புரிந்துகொள்வதற்கு பல்வேறு சூழ்நிலைகளில் தமிழ் மொழியின் அரசியல் வரலாற்றைக் கருத்தில் கொள்வது முக்கியமாகும்.

தமிழ் தகவல்களை இணையத்தில் ஒழுங்குபடுத்தும் செயல்முறையானது, குறிப்பாக மொழியின் பல்வேறு வேறுபாடுகள் மற்றும் அரசியல் வரலாறு தொடர்பான விடயங்களில், குறைவாகவே ஆராயப்பட்டுள்ளது. இந்த அறிக்கை தமிழ் தகவல் ஒழுங்குபடுத்தலுக்குப் பயன்படுத்தப்படும் தற்போதைய அமைப்புகளை ஆராய்வதற்காக பண்பியல் (qualitative) மற்றும் அளவையியல் (quantitative) தரவுகளைப் பயன்படுத்துகிறது, மேலும் தமிழ் பயனாளர்களுக்குச் சிறந்த அனுபவங்களை வழங்கக்கூடிய நியாயமான மற்றும் சமமான மனித மற்றும் தானியங்கி ஒழுங்குபடுத்தல் செயல்முறைகளை நிறுவுவதற்கான பரிந்துரைகளை வழங்குகிறது.

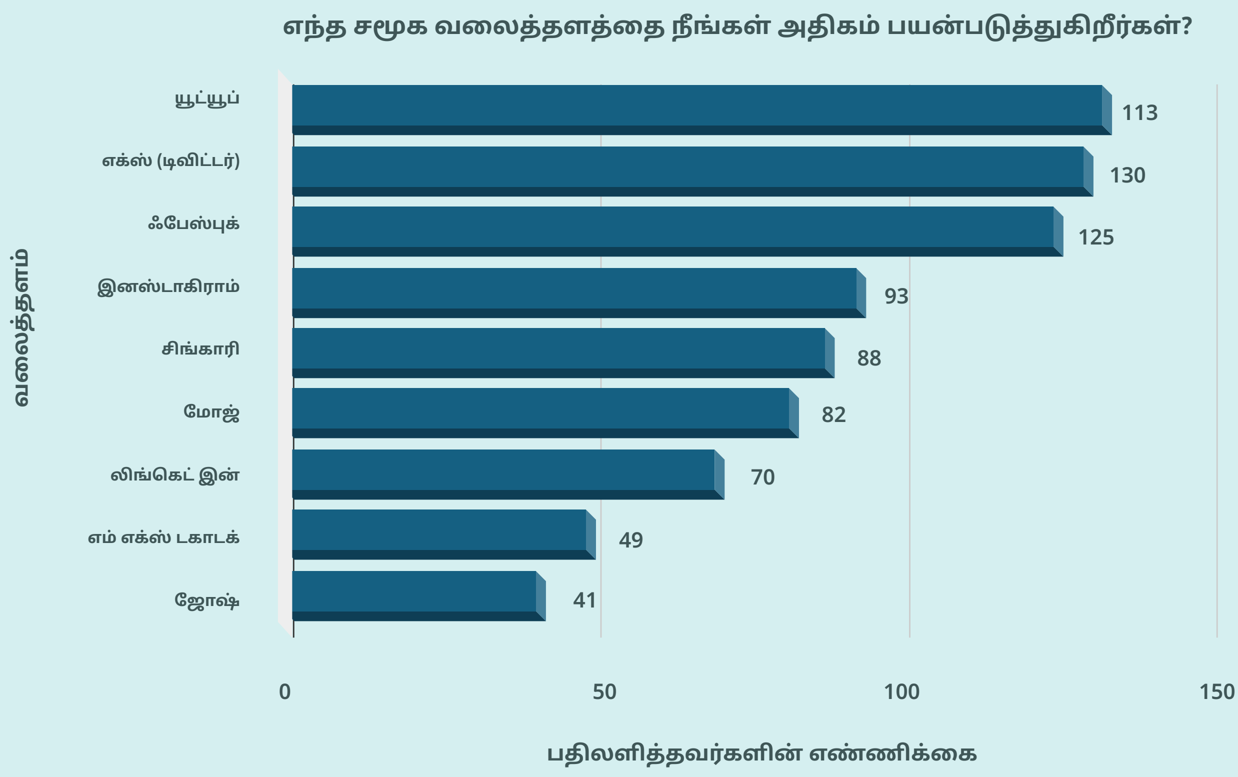
பிரதான கண்டுபிடிப்புகள்

1. தமிழ் பேசுபவர்கள், சில சமயங்களில் ஒழுங்குபடுத்தல் செயல்முறையைத் தவிர்ப்பதற்காக, குறியீட்டு கலப்புத் தமிழ், கணினி வழித் தமிழ் மற்றும் ஒலிபெயர்ப்பு செய்யப்பட்ட தமிழில் சமூக ஊடக சேவைகள் மற்றும் இணைய மன்றங்களின் கலவையைப் பயன்படுத்துகின்றனர்.

அமெரிக்காவை மையமாகக் கொண்ட சமூக ஊடக தளங்கள் மற்றும் இந்தியச் சமூக ஊடக தளங்கள் ஆகிய இரண்டுமே தமிழ் பேசுபவர்கள் மற்றும் இந்தியத் தமிழ் பயனாளர்களிடையே பிரபலமாக உள்ளன. நாங்கள் கணக்கெடுத்த 147 சமூக ஊடக பயனாளர்களிடையே யூடியூப், எக்ஸ், பேஸ்புக் மற்றும் இன்ஸ்டாகிராம் போன்ற அமெரிக்க சமூக ஊடக தளங்கள் மிகவும் பிரபலமானவையாக இருக்கின்றன. இந்தியாவில் இந்தியத் தளங்களைப் பயன்படுத்துவதாகவும் ஆய்வில் பங்கேற்றவர்கள் குறிப்பிட்டனர். பங்கேற்பாளர்களில் 60% (88 பேர்) பேர் வீடியோ பகிர்வு மொபைல் செயலியான சிங்காரியைப் பயன்படுத்துவதாகத் தெரிவித்தனர். 2020ம் ஆண்டு ஜூன் மாதம் இந்தியாவில் டிக்டோக் தடைசெய்யப்பட்டதைத் தொடர்ந்து, அதற்கு மாற்றாக உருவாக்கப்பட்ட மோஜ், ஜோஷ் மற்றும் எம்எக்ஸ் டகாடக் ஆகியவை தமிழ் பேசுபவர்களால் பயன்படுத்தப்படுகின்றன (படம் 1 ஐப் பார்க்கவும்).

இணையத்தில், தமிழ் பல வழிகளில் பயன்படுத்தப்படுகிறது. எழுத்துத் தமிழ் பெரும்பாலும் முறையான வடிவமாகக் காணப்படுகிறது, அதே நேரத்தில் பேச்சுத் தமிழ் பெரும்பாலும் பேச்சுவழக்காகவும் பொதுவானதாகவும் கருதப்படுகிறது. இதன் விளைவாக, சமூக ஊடகங்களில் பதிவிடப்படும் தமிழானது பெரும்பாலும் பேச்சுத் தமிழின் எழுத்து வடிவமாக இருக்கிறது, இதில் குறிப்பிட்ட விதிகள் மற்றும் நெறிமுறைகள் குறைவாகவே காணப்படுகின்றன. தமிழ் பேசுபவர்கள் பெரும்பாலும் குறியீட்டு கலப்புச் சொற்கள் அல்லது செய்திகளை எழுதுவதாக நேர்காணலில் பங்கேற்றவர்கள் தெரிவித்தனர், இதில் தமிழ் மற்றும் ஆங்கிலத்தை இணைத்து தமிலிஷ் அல்லது தங்கிலிஷில் பேசுவதும் அடங்கும்.[4] உதாரணமாக, அவர்கள் ஏதாவது ஒன்றைச் சிறந்தது என விவரிப்பதற்காக"super-a-irukku (சூப்பர்-ஆ-இருக்கு)" என்று கூறலாம், அல்லது பொதுவாக ஒரே வாக்கியத்தில் தமிழையும் ஆங்கிலத்தையும் மாறி மாறி பயன்படுத்தலாம். குறிப்பாக பொதுவான சொற்களுக்கு புதிய அர்த்தங்களை உருவாக்கியுள்ள ஜென்ஸீ மற்றும் ஜென் ஆல்பா

4 தமிலிஷ் அல்லது தங்கிலிஷ் ஆனது உரையாடுவதற்காகத் தமிழ் மற்றும் ஆங்கில வார்த்தைகளை வாக்கியங்களில் பயன்படுத்தும் பயனர்களையும், தமிழ் பேச்சை ஒலிபெயர்ப்பதற்கு லத்தீன் எழுத்துக்களைப் பயன்படுத்துபவர்களையும் உள்ளடக்கியது.



▲ **படம் 1:** ஆய்வில் பதிலளித்த, தமிழில் சமூக ஊடக சேவைகளை வழக்கமாக (தினசரி அல்லது வாராந்திரம்) பயன்படுத்தும் நபர்களின் எண்ணிக்கை (n=147). பங்கேற்பாளர்கள் பல விருப்பத் தெரிவுகளைத் தேர்ந்தெடுக்க அனுமதிக்கப்பட்டனர்).
இந்தத் தகவல் CDT நடத்திய இணையக் கருத்துக் கணிப்பிலிருந்து பெறப்பட்டது (நவம்பர்–டிசம்பர் 2024) பெறப்பட்டது.

என அழைக்கப்படும் இளம்தலைமுறையினரிடையே தமிலிஷ்/தங்கிலிஷ் பரவலாகக் காணப்படுகிறது. ஒரு உதாரணம் "மேட்டர்" (matter), அதாவது ஒரு சம்பவம் அல்லது மோதல், பொதுவாக இது "konjam matter iruku (கொஞ்சம் மேட்டர் இருக்கு) என்ற விதத்தில் ஒரு வாக்கியத்தில் பயன்படுத்தப்படுகிறது, இதனுடைய அர்த்தம் "கொஞ்சம் பிரச்சனை இருக்கிறது" என்பதாகும்.

தமிலிஷ் அல்லது தங்கிலிஷ் மொழியின் பயன்பாட்டில், லத்தீன் எழுத்துக்களைப் பயன்படுத்தி தமிழ் மொழியை ஒலிபெயர்க்கும் பேச்சு முறைகளும் அடங்கும். பல பங்கேற்பாளர்கள் லத்தீன் எழுத்துக்களைப் பயன்படுத்தி தமிழை ஒலிபெயர்த்தும் தமிழ் சொற்கள் மற்றும் வாக்கியங்களின் உச்சரிப்பை எழுதியும் இணையத்தில் தமிழில் பேசினர். பங்கேற்பாளர்கள் இந்த நடைமுறையை, லத்தீன் எழுத்து விசைப்பலகைகளைக் கொண்ட இணையத்துடன் இணைக்கப்பட்ட சாதனங்கள் எங்கும் நிறைந்திருப்பதன் அறிகுறி எனவும், தமிழ் விசைப்பலகைகள் ஒப்பீட்டளவில் அணுக முடியாத அல்லது கிடைக்காததன் அறிகுறி எனவும் விவரித்தனர்.

இதன் விளைவாக, இணையத்தில் எந்த வகையான தமிழையும் பேசும் பயனர்கள் தானியங்கி ஒழுங்குபடுத்தல் முறைகளைத் தவிர்ப்பதற்காக ஆக்கப்பூர்வமான உத்திகளை உருவாக்கியுள்ளனர். எங்கள் நேர்காணல்களில், பங்கேற்பாளர்கள் குறியீட்டு-கலப்பு, ஒலிபெயர்ப்பு அல்லது "அல்கோஸ்பீக்" ஆகியவற்றைப் பயன்படுத்தி ஒழுங்குபடுத்தலை தவிர்க்கலாம் என்று குறிப்பிட்டனர். அல்கோஸ்பீக் அல்லது கணினி-வழித் கருத்து பரிமாற்றம் என்பது தகவல் நீக்கப்படுவதை அல்லது ஒழுங்குபடுத்தல் அமைப்புகளால் தரவரிசைப்படுத்துவதில் கீழே தள்ளப்படுவதைத்

தவிர்ப்பதற்காகக் குறியீட்டு வார்த்தைகளை உருவாக்குதல் மற்றும் பயன்படுத்துவதைக் குறிக்கிறது (Lorenz, 2022; Treem et. al, 2020). எங்கள் நேர்காணல்களில், அன்றாடம் இணையத்தில் தொந்தரவுகளை எதிர்கொள்ளும் ஒரு LGBTQ+ ஆர்வலர், மக்கள் ஒழுங்குபடுத்தல் நடவடிக்கையிருந்து தப்பிப்பதற்காக நட்சத்திரக் குறிகளை பயன்படுத்துவதாகக் கூறினார்.

மற்றொரு பங்கேற்பாளர் கூறியதாவது:

> “நாய் என்ற பேச்சுவழக்கு சொல்லைப் பயன்படுத்துவதற்குப் பதிலாக, தரமான வார்த்தையைப் பயன்படுத்துவார்கள். அல்லது வாசகர்கள் புரிந்துகொள்ளும் வகையில் “நாய்” என்ற தரமான வார்த்தையின் எழுத்துக்களைப் புரட்டுவார்கள், ஆனால் இதனை அல்கோரிதத்தினால் (algorithm) புரிந்துகொள்ள முடியாது. அல்லது அல்கோரிதத்தினைத் தவிர்ப்பதற்காக “நா (na)” அல்லது [நட்சத்திரக் குறியீடு] அல்லது சின்னங்களைப் பயன்படுத்துவார்கள்” (Tamil language advocate, November 2024, India).

இலங்கையிலிருந்து காப்பகப்படுத்தப்பட்ட தமிழ் ஆவணங்களைப் பகிர விரும்பும் தமிழ் பயனாளர்கள் மற்றும் ஆவணக்காப்பாளர்களால் பயன்படுத்தப்படும் மற்றொரு வகையான உத்தியாக, பல்லூடக ஒழுங்குபடுத்தல் நுட்பங்களைத் தவிர்ப்பதற்காகச் சின்னங்களை மங்கலாக்குதல் அல்லது பதிவிடும் பதிவுகளைச் சூழ்நிலைக்கு ஏற்றவாறு வெட்டிப் பயன்படுத்துதல் என்பவற்றைக் குறிப்பிடலாம். பல்லூடகங்களை ஒழுங்குபடுத்தும் நுட்பங்கள், படங்கள், ஆடியோ மற்றும் வீடியோ போன்றதகவல்களைப் பகுப்பாய்வு செய்து ஒழுங்குபடுத்தக்கூடிய தானியங்கி கருவிகள் பயன்படுத்தப்படுகின்றன. (Thakur & Llansó, 2021). யாழ்ப்பாண நூலகம் எரிக்கப்பட்டதில் ஏராளமான தமிழ் ஆவணங்களை இழந்ததால், இலங்கையின் உள்நாட்டுப் போருக்குப் பின்னர், இலங்கையிலும் புலம்பெயர் சமூகங்களிலும் உள்ள தமிழ் பயனாளர்கள் தமிழ் வரலாற்றையும் ஆவணங்களையும் இணையத்தில் பாதுகாக்க முயன்றனர். நேர்காணலில் பங்கேற்ற பலர், உள்நாட்டுப் போரிலிருந்து காப்பகப்படுத்தப்பட்ட படங்கள் அல்லது ஆவணங்களை இணையத்தில் பகிர்ந்ததற்காக, குறிப்பாக இந்த ஆவணங்கள் தமிழீழ விடுதலைப் புலிகள் (LTTE) பற்றிக் குறிப்பிட்டதற்காகப் பதிவு நீக்கங்களை எதிர்கொண்டதாகக் கூறினர்.[5] 2009 ஆம் ஆண்டு போர் முடிவடைந்ததிலிருந்து செயலிழந்து இருக்கும் LTTE இயக்கம் சில தளங்கள் மற்றும் அரசாங்கங்களால் பயங்கரவாத அமைப்பாக அறிவிக்கப்பட்டுள்ளது (Cronin-Furman & Arulthas, 2021). செய்தி தளங்கள் அல்லது அறிஞர்கள் LTTE தொடர்பான தகவல்களைப் பதிவிடும்போது, கடுமையான ஒழுங்குபடுத்தலை எதிர்கொள்கின்றனர். நேர்காணலில் பங்கேற்ற பலர், இலங்கையில் விடுதலைப் புலிகள் இயக்கத்தைப் பற்றி

5 1990களில் இரண்டு உயர்மட்ட அரசியல் படுகொலைகள் மற்றும் பல படுகொலை முயற்சிகளுக்குப் பிறகு, 33 நாடுகளால் LTTE ஆனது பயங்கரவாத அமைப்பாக அறிவிக்கப்பட்டது. 2009 ஆம் ஆண்டுக்குப் பிறகு, உள்நாட்டுப் போர் முடிவடைந்த பின்னர் LTTE கலைக்கப்பட்டதைத் தொடர்ந்து, பல நாடுகள் இந்த சட்டத்தை அமல்படுத்துவதைத் தளர்த்தின. கடந்த பத்து ஆண்டுகளில், பெரும்பான்மையான தமிழ் புலம்பெயர்ந்தோர் வசிக்கும் நாடுகளான கனடா மற்றும் ஐக்கிய இராச்சியத்தில் தற்போது பதவியிலுள்ள பல அரசியல்வாதிகள் LTTE சின்னங்கள் மற்றும் கொடிகளுடன் நிகழ்வுகளில் கலந்து கொண்டு புகைப்படம் எடுத்துள்ளனர். (Tamil Guardian, 2023).

செய்தி வெளியிட்ட பின்னர் செய்தி ஊடக நிறுவனமான ‘தமிழ் கார்டியனின்’ பேஸ்புக் மற்றும் இன்ஸ்டாகிராம் கணக்குகள் நீக்கப்பட்டமையை, ஒரு செய்தி நிறுவனம் என்ற அந்தஸ்தை பெற்றிருந்தாலும் கூட இந்த அளவுக்கு மீறிய தகவல் நீக்கம் செய்யும் முறை கடுமையாகவே செயல்படுகிறது என்பதற்கான ஒரு உதாரணமாகச் சுட்டிக்காட்டினர் (Biddle, 2022; Amarsingham & Nandakumar, 2021). நேர்காணலில் பங்கேற்றவர்கள், விடுதலைப் புலிகள் பற்றிக் குறிப்பிடும் செய்தித்தாள் துணுக்குகள் அல்லது ஆவணங்களை வெளியிட்டாலும் கூட, அவர்களின் பதிவுகள் நீக்கப்படும் என்று கூறினர்.

> "நான் 1980களின் பழைய செய்தித்தாள் துணுக்குகளை வெளியிடுவதை வழக்கமாகக் கொண்டிருந்தேன், அவர்கள் அதைத் தணிக்கை செய்வார்கள். இவை எல். டி.டி.இ பற்றிய துணுக்குகள் அல்லது சுவரொட்டிகள் அல்லது அவற்றை ஒத்த விஷயங்களாகும். "நாங்கள் எம்மை விடுவித்துக் கொள்வோம், சங்கிலிகளை உடைப்போம்” போன்ற விஷயங்கள். சின்னங்கள் போன்ற விஷயங்கள் கூட இல்லை. அது தணிக்கை செய்யப்படும். இதற்குப் பதிலளிக்கும் விதமாக, மக்கள் இன்ஸ்டாகிராமில் உள்ள பதிவுகளை மங்கலாக்குவார்கள். சுவரொட்டி தமிழில் இருந்தது, அவர்களால் எப்படியோ அதை அடையாளம் காண முடிந்தது. நான் என்னுடைய படிப்பின் போதோ அல்லது ஆராய்ச்சி செய்யும் போதோ நான் கண்ட வரலாற்று விஷயங்களைப் பகிர்ந்து கொள்ள விரும்பினேன். என்னால் அதைச் செய்ய முடியாதபோது, எந்தக் காரணத்திற்காகவும் நான் இந்த தளத்தில் இருக்க வேண்டிய அவசியமில்லை என்று தோன்றியது. நான் அதனைப் பற்றி அதிக அக்கறை கொள்ளவில்லை. யாழ்ப்பாண நூலகத்தைப் பற்றி உங்களுக்குத் தெரியுமா? அது 1981 இல் எரிக்கப்பட்டது... மிகவும் அதிர்ச்சிகரமானது. ஏனென்றால், நிறைய வரலாற்று விஷயங்கள் அங்கே இருந்தன. மக்கள் இன்றும் அதை நினைவில் வைத்திருக்கிறார்கள். அதற்குப் பதிலளிக்கும் விதமாக, தமிழர்கள் இணையக் காப்பகங்களை உருவாக்கினர்... ஆனால் நீங்கள் அதை இன்ஸ்டாகிராம் மற்றும் பேஸ்புக்கில் அவ்வளவாகப் பார்ப்பதில்லை.” (தமிழ் பத்திரிகையாளர் மற்றும் டிஜிட்டல் உரிமைகள்களுக்கு குரல் கொடுப்பவர், நவம்பர் 2024, இந்தியா).

2. உலகமயமாக்கப்பட்ட மற்றும் பகுதிசார்ந்த அணுகுமுறைகள்: சமூக ஊடகங்கள் மற்றும் இணைய மன்றங்கள் தமிழ் மொழியில் தகவல்களைக்

கண்காணித்தல் மற்றும் ஒழுங்குபடுத்துவதில் கலவை முறையிலான அணுகுமுறைகளைப் பின்பற்றுகின்றன.

மேற்கத்திய அல்லது இந்திய இணையச் சேவைகளில் பணிபுரியும் பங்கேற்பாளர்களுடனான நேர்காணல்கள் மூலம், தமிழ்த் தகவல் ஒழுங்குபடுத்துதலைப் பொறுத்தவரையில் நிறுவனங்கள் கலவையான அணுகுமுறைகளைப் பின்பற்றுவதைக் கண்டறிந்தோம். பொதுவாக, நிறுவனங்கள் தகவல்களை ஒழுங்குபடுத்துவதற்கு "உலகளாவிய" அல்லது மொழிச்சார்பற்ற அணுகுமுறையைப் பின்பற்றுகின்றன, இது இந்தத் தொடரில் முந்தைய வழக்கு ஆய்வில் காணப்பட்டது (Elswah, 2024a). இருப்பினும், சில நிறுவனங்கள் தமிழ் பயனாளர்களுக்கு தங்கள் சேவையை வழங்கத் தொடங்கியபோது "பகுதிசார்ந்த" அல்லது மொழி சார்ந்த அணுகுமுறையை ஏற்றுக்கொண்டதையும் நாங்கள் கண்டறிந்தோம்.

உலகளாவிய அணுகுமுறை

எங்களுடன் பேசிய பெரும்பாலான அமெரிக்காவை மையமாகக் கொண்ட பெரிய சமூக ஊடக நிறுவனங்கள், உலகளாவிய தகவல் ஒழுங்குபடுத்தல் அணுகுமுறையைத் தேர்ந்தெடுத்துள்ளன, பயனாளர்கள் எந்த மொழியைப் பயன்படுத்துகிறார்கள் மற்றும் அவர்கள் எங்கு இருந்தார்கள் என்பதைப் பொருட்படுத்தாமல் அனைவருக்கும் ஒரே மாதிரியான தகவல் கொள்கைகளைப் பயன்படுத்துகின்றன. இந்த அணுகுமுறை மொழி சார்ந்த நிபுணத்துவம் இல்லாத மதிப்பீட்டாளர்கள் அல்லது கொள்கை நிபுணர்களை ஆட்சேர்ப்பு செய்வதை உள்ளடக்கியது. அமெரிக்காவை மையமாகக் கொண்ட சமூக ஊடக நிறுவனக் கொள்கைத் தலைவர் ஒருவர், தங்கள் நிறுவனத்தின் மொழிகளுக்கான தகவல் ஒழுங்குபடுத்தல் அணுகுமுறையை "கவரேஜ் மாதிரி" என்று குறிப்பிட்டார், இதில் ஒரு குறிப்பிட்ட மொழியில் தகவல் ஒழுங்குபடுத்துவதற்கு போதுமான வளங்கள் காணப்படுகின்றதா என்ற கேள்வி பெரும்பாலும் நெருக்கடிக் காலங்களிலேயே பரிசீலிக்கப்பட்டது. அவர் குறிப்பிட்டது பின்வருமாறு.

> "ஒரு நெருக்கடி ஏற்படும் போது மொழி என்பது மேலெழும். அதாவது உண்மையில் ஒரு பாதுகாப்பு அச்சுறுத்தல் இருந்தாலோ... அல்லது, இணையதளத்தில், ஏதாவது ஒரு வன்முறையை தொடர்புப்படுத்திப் பதிவு இருந்தாலோ, ஏதேனும் ஒரு மோதல் வெடித்து, அதுபற்றி சமூகத்திலோ அல்லது ஊடகங்களிலோ உரையாடல் இருந்தாலோ, "அப்போதுதான் போதுமான மொழிவளங்கள் உள்ளனவா, வெறுப்புப் பேச்சின்

> வகைப்படுத்திகள் உள்ளனவா என்பதை மறுபரிசீலனை செய்வது குறித்து ஆலோசிக்கப்படும்." (அமெரிக்காவை தளமாகக் கொண்ட சமூக ஊடகக் கொள்கை பிரதிநிதி, டிசம்பர் 2024)

உலகளாவிய அணுகுமுறையில், மதிப்பீட்டாளர்கள் பெரும்பாலும் தங்கள் பணிகளை ஆங்கிலத்தில் செய்யவே பணியமர்த்தப்படுகிறார்கள். ஒரு பங்கேற்பாளர் கூறியதாவது:

> "தகவல் மதிப்பாய்வாளர்கள் ஒரு குறிப்பிட்ட அளவுக்கு ஆங்கில மொழியைப் பற்றிய புரிதலைக் கொண்டிருக்க வேண்டும் என்று எதிர்பார்க்கப்படுகிறது. அவர்கள் கொள்கைகளைப் புரிந்துகொண்டு அவை [கொள்கைகள்] எதிர்பார்க்கப்பட்ட விளைவைக் கொண்டிருப்பதை உறுதிசெய்ய வேண்டும்." (அமெரிக்காவை தளமாகக் கொண்ட சமூக ஊடக நிறுவனத்தில் கொள்கை, டிசம்பர் 2024, இந்தியா)

சில நேரங்களில் தகவல் தானாகவே ஆங்கிலத்தில் இயந்திர ரீதியாக மொழிபெயர்க்கப்படும், மதிப்பாய்வாளர்களுக்கு அசல் மொழி என்னவென்று சொல்லப்படுவதில்லை. மற்ற நேரங்களில், பதிவுகள் மதிப்பாய்வாளர்கள் மற்றும் நம்பிக்கை மற்றும் பாதுகாப்புக் குழுக்களுடன் பகிரப்படுகின்றன. இயந்திர மொழிபெயர்ப்பு கருவிகள் பிழை ஏற்படக்கூடியவை என்று அறியப்பட்டிருந்தாலும், குறிப்பாகக் குறைந்த வளம் கொண்ட மொழிகளில், அவர்கள் தகவல்களின் பகுதிகளை மொழிபெயர்க்க இயந்திர மொழிபெயர்ப்பு கருவிகளை நம்பியிருக்குமாறு கேட்டுக்கொள்ளப்படுகிறார்கள். (Nicholas & Bhatia, 2023). ஒரு கொள்கைத் தலைவர் பின்வருமாறு விளக்கினார்:

> "தகவல் தானாகவே மொழிபெயர்க்கப்படுகிறது... இந்த தளத்திலேயே உள்ளமைக்கப்பட்ட மொழிபெயர்ப்புக் கருவிகள் உள்ளன, மேலும் அவை உண்மையில் பிராந்திய மொழிகளில் சிறப்பாகச் செயல்படுவதில்லை என்பது அறியப்படுகிறது. இந்தியாவில் ஒழுங்குபடுத்தல் நடைபெறுகிறது எனின், தமிழ் ஆங்கிலத்திற்கு மொழிபெயர்க்கப்படும், பின்னர் மதிப்பீட்டாளர் அதைப் பார்ப்பார்... தளத்தின் மொழிபெயர்ப்பு உங்களுக்கு ஆங்கிலத்தில் தோராயமாகச் சொல்லக்கூடிய அளவுக்குத்தான் இருக்கும்." - (U.S.-based social media policy lead, December 2024, India).

அமெரிக்காவை மையமாகக் கொண்ட சமூக ஊடக தளங்கள் தமிழ் பேசும் தகவல் மதிப்பீட்டாளர்களை பணியமர்த்தும்போது, அவர்கள் எப்போதும் தமிழ் தகவல்களை மதிப்பாய்வு செய்யும் பணியை அவர்களிடம் கொடுப்பதில்லை. பெரும்பாலும், இந்த பெரிய சமூக ஊடக நிறுவனங்களில் மனித மதிப்பாய்வு செய்யப்படும் தகவல்களை, அவர்களின் மொழிப்புலமையைக் கருத்திற்கொள்ளாமல், ஒழுங்கற்ற முறையில் மதிப்பீட்டாளருக்கு வழங்கப்படுகிறது. இந்த நடைமுறை உலகம் முழுவதும் தகவல் ஒழுங்குபடுத்தும் விதிகளில் நிலைத்தன்மையை உறுதி செய்வதை நோக்கமாகக் கொண்டது என்று அமெரிக்காவை மையமாகக் கொண்ட சமூக ஊடக நிறுவனமொன்றின் கொள்கைக் குழு உறுப்பினர் ஒருவர் கூறினார்:

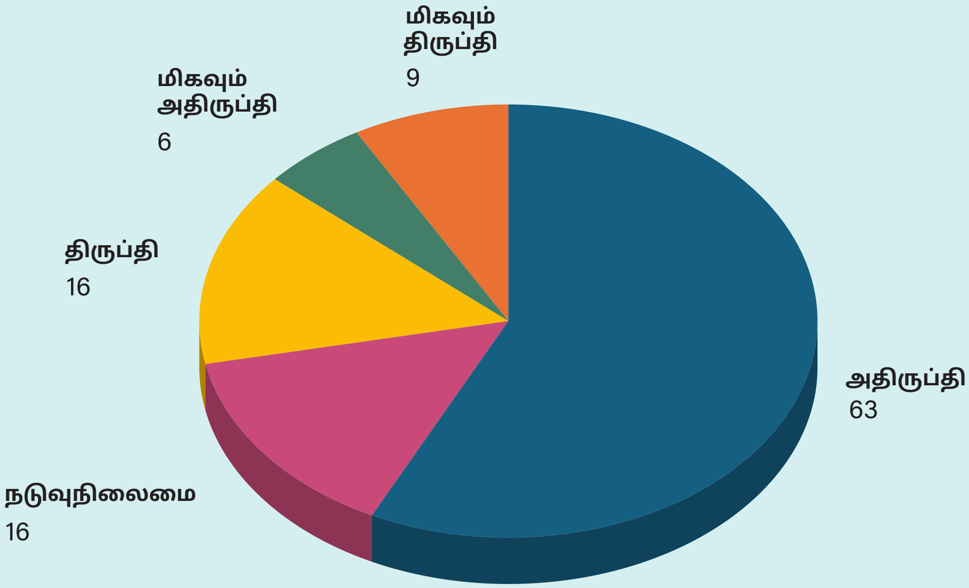
“பிராந்திய விவரக்குறிப்புகள் இருந்தால், அது அனைத்து தகவல் மதிப்பாய்வாளர் வழிகாட்டுதலிலும் சேர்க்கப்பட்டு, அனைவருக்கும் கிடைக்கும். இதற்குக் காரணம், ஒரு பதிவு, அது தெற்காசியாவிலிருந்து வெளிவந்தாலும் கூட, EMEA [ஐரோப்பா, மத்திய கிழக்கு மற்றும் ஆப்பிரிக்காவைக் குறிக்கிறது] இல் உள்ள ஒரு மதிப்பாய்வாளருக்குச் செல்ல வாய்ப்புள்ளது. (அமெரிக்காவை தளமாகக் கொண்ட சமூக ஊடக நிறுவனத்தில் கொள்கை, டிசம்பர் 2024, இந்தியா).

பகுதிசார்ந்த அணுகுமுறை

சில இந்தியச் சமூக ஊடக நிறுவனங்களும் அமெரிக்க இணைய மன்றங்களும் தமிழ் பேச்சைக் கட்டுப்படுத்துவதற்கு பகுதிசார்ந்த அணுகுமுறைகளை ஏற்றுக்கொண்டுள்ளன. எங்களுடன் பேசிய எந்த அமெரிக்கச் சமூக ஊடக நிறுவனங்களும் தகவல் ஒழுங்குபடுத்தலுக்கு பகுதிசார்ந்த அணுகுமுறையைப் பயன்படுத்தவில்லை. மக்ரெபி அரபு மொழியில் தகவல் ஒழுங்குபடுத்தல் எவ்வாறு மேற்கொள்ளப்படுகிறது என்பது பற்றிச் செய்யப்பட்ட முந்தைய வழக்கு ஆய்வில் பகுதிசார்ந்த அணுகுமுறைகள் ஆராயப்பட்டன (Elswah, 2024a). இந்த தமிழ் விஷயத்தில், தளங்கள் தமிழ்ச் சூழல்களுக்கு ஏற்றவாறு கொள்கைகளை மாற்றுவது மட்டுமல்லாமல், வலைப்பதிவுகள் மூலம் பயனர்களுடன் தொடர்பு கொள்ளவும், வழிகாட்டுதல் வழங்கவும், நிறுவனத்தின் கொள்கை குழுக்களுடன் கருத்துக்களைப் பகிர்ந்து கொள்ளவும் மதிப்பீட்டாளர்களுக்குக் கூடுதல் அதிகாரத்தினை வழங்குகின்றன. பகுதிசார்ந்த அணுகுமுறைகளைப் பயன்படுத்திய மேற்கத்தியத் தொழில்நுட்ப நிறுவனங்கள் பெரும்பாலும் பெரிய சமூக ஊடக நிறுவனங்களை விட இணைய தளங்களாக இருந்தன. அவர்கள் தமிழ் நிபுணர்களை பணியமர்த்துவதன் மூலம் பகுதிசார் அணுகுமுறையைப் பயன்படுத்தினர், மேலும் தமிழ் மொழி சார்ந்த வழிகாட்டுதலை உருவாக்கக் கொள்கை குழுக்களுடன் பேசுவதற்கு அடிக்கடி சந்திப்புகள் அல்லது செய்தி அனுப்பும் இடங்கள் போன்ற வழிமுறைகளை அவர்களுக்கு வழங்கினர், மேலும் தமிழ் பேசுபவர்களுக்குச் சேவையை வழங்குவதற்கு முன்பாக சேவை விதிமுறைகளையும், தொடர்புகொள்வதற்கான வளங்களையும் உருவாக்கினர் கொள்கை குழுக்களுடன் அவர்கள் அடிக்கடி எவ்வாறு தொடர்பு கொள்கிறார்கள் என்பது பற்றி ஒரு தமிழ் மதிப்பீட்டாளர் பேசினார்:

“[இணைய தளம்] இந்திய மொழிகளில் தொடங்கப்பட்டபோது, [இணைய தளத்தின்] ஆங்கிலப் பதிப்பில் கிடைத்த அனைத்துக் கொள்கைகளும் அந்தந்த மொழிகளுக்கு ஏற்றவாறு மொழிபெயர்க்கப்பட்டன... எங்கள் பெரும்பாலான கொள்கைகள் மொழிபெயர்க்கப்பட்டதால் அவை தெளிவற்றவையாக இருந்தன. "அப்போது, அந்த கொள்கை கவனிக்கப்பட வேண்டியதா என்பதைப் புரிந்துகொண்டு கொள்கையை மாற்றக் கலந்துரையாடல் அவசியமானது" (Tamil community moderator for an online forum, India, December 2024).

சமூக வலைதளங்களில் நீங்கள் அளித்த புகாருக்கு வந்த பதில் எவ்வளவு திருப்திகரமாக இருந்தது?



▲ **படம் 2**: புகார் அமைப்புகள் மற்றும் செயல்முறைகளின் செயல்திறன் பற்றி கணக்கெடுப்பில் பதிலளித்தவர்களின் கருத்துக்கள் (பதிவை ஒரு முறையாவது புகாரளித்ததாகக் கூறிய பங்கேற்பாளர்கள் %. n=110).

இந்த தகவல் CDT நடத்திய இணையக் கருத்துக் கணிப்பிலிருந்து (நவம்பர்–டிசம்பர் 2024) பெறப்பட்டது.

சில பகுதிசார்ந்த தகவல் ஒழுங்குப்படுத்தல் அணுகுமுறைகள், மீறல்களை முறைப்படுத்தப் பயனர்களையே பெரிதும் நம்பியுள்ளன. மேலும், இவை பயனர்களின் கருத்துக்களைக் கொண்டு ஒழுங்குபடுத்தல் விதிகளை உருவாக்குகின்றன. இம்முறையில், உயர்தர பயனாளர் அறிக்கைகளை உருவாக்க, சமூகத்திற்குள் அறிவையும் நம்பிக்கையையும் வளர்ப்பது அவசியம். சமூக மதிப்பீட்டாளர்கள் இதைச் செய்வதற்கான ஒரு வழி, ஒரு பதிவு சேவை விதிமுறைகளை மீறும் போது, ஒழுங்குபடுத்தலுக்கான மிகவும் சரியான வழிகாட்டுதலைப் பகிர்வதாகும். ஒரு மதிப்பீட்டாளர் பின்வருமாறு குறிப்பிட்டார்:

> "குறிப்பிட்ட கொள்கை மீறல்கள் அதிக எண்ணிக்கையில் இருப்பதை நான் காணும்போதெல்லாம், கொள்கை மீறல்கள் குறித்து [தளத்தில் வலைப்பதிவின் வடிவத்தில்] ஒரு அறிவிப்பை வெளியிடுவதை உறுதிசெய்வேன். மக்களின் பெயர்களைக் குறிப்பிடாமல், பொதுவாக [பதிவுகளை] சுட்டிக்காட்டுகிறேன். அதைப் பயனருக்கு நேரடிச் செய்தியாகவும் அனுப்பி அவர்களுக்குத் தெரியப்படுத்துவேன்..."நாங்கள் எல்லா கருத்துகளையும் அனுமதித்தோம்; ஏதாவது நாகரீகமற்றதாகவோ, எதையாவது தூண்டுவதாகவோ அல்லது வெறுப்பூட்டுவதாகவோ இருந்தால், சமூகம் கருத்துகளை எங்களுக்குத் தெரிவிக்கும் என்று நாங்கள் நம்பினோம், அதன்படி சமூகத்திலிருந்து அத்தகைய அறிக்கைகளைப் பெறுகிறோம். அதன் அடிப்படையிலேதான் நாங்கள் செயல்படுகிறோம்." - (ஆன்லைன் மன்றத்திற்கான ஒரு தமிழ் சமூக மதிப்பீட்டாளர்,டிசம்பர் 2024, இந்தியா).

இந்தியச் சமூக ஊடக நிறுவனங்களும் தங்கள் தகவல் கொள்கைகளை பகுதிசார்ந்து அணுகுகின்றன. ஆனால் அனைத்து நேரங்களிலும் இல்லை. பெரும்பாலும்,

இந்தியத் தொழில்நுட்ப நிறுவனங்கள் தகவல்களை ஒழுங்குபடுத்துவதற்காக ஒரு அகில இந்திய அணுகுமுறையைப் பின்பற்றுகின்றன. அனைத்து பயனர்கள் மற்றும் பகுதிகளின் பேச்சையும் ஒரே மாதிரியாகக் கருதுகின்றன. இருப்பினும், சில சமயங்களில், கலாச்சார வேறுபாடுகளைக் கணக்கில் எடுத்துக்கொண்டு, பெரும்பாலும் தற்காலிக அடிப்படையில் தங்கள் கொள்கைகளை பகுதிசார்ந்து மாற்றுகின்றன:

> “இந்த தளத்தில் உள்ள மிகப்பெரிய மொழிகளில் தமிழும் ஒன்றாகும். இந்தியை விட பெரியது...இதற்கான பல காரணங்களில் ஒன்று மிகப் பெரிய பொழுதுபோக்குத் துறை...குழுவில் உள்ள ஒரு இளையவர், தமிழர், அந்த ஆவணங்களை எழுதி, அவற்றை பகுதிசார்ந்து மாற்றியமைத்தார்...உதாரணமாக, தமிழ்ப் பதிப்பில் பெரியாரைப்[6] பற்றி நீங்கள் எதுவும் சொல்ல முடியாது... இது ஒரு எழுதப்படாத வீதி... ஒவ்வொரு மொழியிலும் சில சொற்களை நாங்கள் அமைதிப்படுத்தியிருக்கிறோம். எந்த வார்த்தைகள் சரி, எவை சரியல்ல என்பதைத் தகவல்களை ஒழுங்குபடுத்தும் குழுக்கள் முடிவு செய்யும். இருந்தபோதிலும் விதிகளைச் சமமாக அமல்படுத்த முடியாது. வெவ்வேறு மொழிகளில், வார்த்தைகள் மிகவும் வித்தியாசமான தாக்கத்தை ஏற்படுத்தும். உதாரணத்திற்கு, ஆங்கிலத்தில், f*ck என்ற வார்த்தையை நாம் சுதந்திரமாகப் பயன்படுத்தலாம். இந்தியில் அந்த வார்த்தையை, நாம் அவ்வளவு சுதந்திரமாகப் பயன்படுத்தமுடியாது. தமிழில் அந்த வார்த்தை என்னவென்று எனக்குத் தெரியவில்லை, அந்த வார்த்தை ஒருவேளை சரியானதாக இருந்தால், நாம் அதை அனுமதிக்கலாம்.” - (இந்திய சமூக ஊடக நிறுவனக் கொள்கை , அக்டோபர் 2024, இந்தியா)

இந்திய நிறுவனங்கள் கொண்டிருக்கின்ற ஒரு பெரிய சவால் மொழி சார்ந்த ஒழுங்குபடுத்தலைப் பெறுவதற்கான செலவு ஆகும். தகவல் மீறலைக் கண்டறிந்து தகவல் கொள்கைகளைச் செயல்படுத்துவதற்காகத் தானியங்கி கருவிகளை உருவாக்குவது அல்லது ஒவ்வொரு மொழியிலும் மதிப்பாய்வாளர்களை பணியமர்த்துவது என்பது நிறையப் பணம் செலவாகக்கூடிய அணுகுமுறை என்று நிறுவன பிரதிநிதிகள் தெரிவித்தனர். தகவலைக் கட்டுப்படுத்த தானியங்கி கருவிகளில் முதலீடு செய்யும்போது, நிறுவனத்திற்கு சட்டச்சிக்கல் ஏற்படக்கூடிய சில பதிவுகளுக்கு முன்னுரிமை அளிக்கப்படுகிறது.

ஆய்வில் பதிலளித்த பெரும்பாலானோர், தவறாக வழிநடத்தும் அல்லது தீங்கு விளைவிக்கும் பதிவுகள் நிறைந்த இணையச் சூழல்களை எதிர்கொண்டதாகக் கூறினர். கணக்கெடுப்பில் பதிலளித்தவர்களில் சுமார் 81% பேர் தவறாக வழிநடத்தும் பதிவுகளின் மிகைமைக் குறித்து கவலை தெரிவித்தனர், மேலும் 82% பேர், வெறுப்பு பேச்சு குறித்து கவலை தெரிவித்தனர். அதேபோல், எங்கள் கணக்கெடுப்பில் பங்கேற்றவர்களில்

---

6 ஈரோடு வெங்கடப்ப ராமசாமி என்ற இயற்பெயர் கொண்ட பெரியார், ஒரு இந்தியச் சமூக ஆர்வலர் மற்றும் தத்துவஞானி ஆவார், அவர் சமூக நீதி மற்றும் சாதி ஒடுக்குமுறை எதிர்ப்பு இயக்கத்தின் முன்னணி நபராகப் பரவலாக அறியப்படுகிறார்.

கிட்டத்தட்ட 75% பேர், ஒரு முறையாவது இணையத்தில் தீங்கு விளைவிக்கும் பதிவுகளைப் பற்றி புகார் செய்ததாகக் குறிப்பிட்டனர். வரம்புகளை மீறுவதாகப் புகாரளித்தவர்களில் சுமார் 63% பேர், தங்களின் புகார்கள் பயனளிக்கவில்லை என்றும், அந்த பதிவு அகற்றப்படவில்லை என்றும் குறிப்பிட்டனர். தளத்தின் பதில் நடவடிக்கை குறித்து சுமார் 57% பேர் அதிருப்தி அடைந்தனர். மேலும், 5% பேர் மிகவும் அதிருப்தி அடைந்தனர் (படம் 2 ஐப் பார்க்கவும்).

தனது பதிவு தீங்கு விளைவிப்பதாகக் கூறி புகார் செய்யப்பட்ட நிலையில், வெறுப்புப் பேச்சினைக் கொண்டிருந்த இன்னுமொரு பதிவு தொடர்பாகத் தான் புகாரளித்தும் எந்தவொரு நடவடிக்கையும் எடுக்கப்படவில்லை என்று டிஜிட்டல் உரிமை செயற்பாட்டாளர் ஒருவர் தனது விரக்தியை வெளிப்படுத்தினார். அவர் கூறியதாவது.

> நான் எக்ஸ், மெட்டா, இன்ஸ்டாகிராம் க்கு எப்போது புகார் செய்தாலும் எதுவுமே நடப்பதில்லை. நான் நெறிமுறைகளைப் பின்பற்றுவதால் இதைச் செய்கிறேன். தகவல் மதிப்பீட்டுக் கொள்கைகளையே நான் பின்பற்றுகிறேன். ஏனென்றால் நான் சிறிய விஷயத்தை எழுதும்போது கூட அது பெருமளவான புகார்களைச் சந்திக்கிறது. "உங்களுடைய பதிவு சமூக வழிகாட்டுதல்களுக்கு எதிரானது" என்று எனக்கு ஒரு அறிவிப்பு வரும். நான் புண்படுத்தும் விஷயங்களைப் பதிவிடுவதில்லை. நான் ஒரு பயிற்சியாளர், மனித உரிமை ஆர்வலர், நான் மிகவும் சாதாரண விஷயங்களையே பதிவிடுகிறேன். ஆனால் அதற்காகக் கூட, நான் தகவல் விதிகளை மீறியதாக் கூறப்படுகிறேன். ஆனால் மக்கள் "நான் உன்னைக் கொல்லப் போகிறேன்" என்று தமிழில் சொல்லும்போது, நான் அதைப் புகார் செய்தாலும் அவர்கள் அதை அகற்றுவதில்லை."அதில் எந்த விதி மீறலையும் நாங்கள் காணவில்லை" என்று அவர்கள் கூறுகிறார்கள்... இது என்ன ஒரு கொள்கை? அதை மொழிபெயர்த்துப் பாருங்கள் ... நாங்கள் உங்களுக்கு மொழிபெயர்ப்புகளை வழங்கியுள்ளோம், உங்களுக்குச் சாப வார்த்தைகளின் பட்டியலை வழங்கியுள்ளோம், தமிழில் சாதி அடிப்படையிலான பழிச்சொற்களை வழங்கியுள்ளோம். உலகம் முழுவதிலும் இருந்து நாங்கள் ஒரு பட்டியல் உருவாக்கி, அதைத் தளங்களுக்கு அனுப்பியிருக்கிறோம். ஆனால் அவர்கள் இன்னும் எதையும் செய்யவில்லை. மக்கள் எனக்கு மிரட்டல் விடுக்கும்போது, எதுவும் நடக்காது. அந்த மிரட்டலை நான் ஒரு புகைப்படமாக (screenshot) பதிவிடுகிறேன். அது "தகவல் விதிமீறல்" என்று கூறி அகற்றப்படுகிறது. அவர்கள் என்னை ஒரு விபச்சாரி என்று அழைத்தார்கள். அதையே நான் ஒரு புகைப்படமாகப் பதிவிடுகிறேன். என் பதிவு அகற்றப்படும் ஆனால் அந்த அசல் பதிவு, அகற்றப்படாது. இந்த செயல் மிகவும் விரக்தியடையச் செய்கிறது” (தமிழ் பத்திரிகையாளர் மற்றும் டிஜிட்டல் உரிமைகள்களுக்கு குரல் கொடுப்பவர், நவம்பர் 2024, இந்தியா).

தமிழ் பேசும் இணைய மன்றங்களில் பாலியல், ஓரினச்சேர்க்கை மற்றும் சாதி அடிப்படையிலான துன்புறுத்தல்கள் மற்றும் அவதூறுகள் பரவலாக இருப்பதாக நேர்காணலில் பங்கேற்பாளர்கள் கூறுகின்றனர். இருப்பினும், நிறுவனங்களினால் வழங்கப்படும் பதில், மொழி மற்றும் யார் பாதிக்கப்படுகிறார்கள் என்பதைப் பொறுத்து மாறுபடும் என்று நேர்காணல் பங்கேற்பாளர்கள் நம்புகின்றனர். இருமொழி அவதூறுகள், ட்ரோலிங் மற்றும் மீம்ஸ்களின் பயன்பாட்டைக் குறித்து நேர்காணலில் பங்கேற்ற ஒருவர் பின்வருமாறு கூறினார்:

> "பெண்களுக்கு எதிரான அவதூறுகள் புகார் செய்யப்படலாம் அகற்றப்படலாம், ஆனால் சாதி அடிப்படையிலான அவதூறுகள் ஒருபோதும் அகற்றப்படாது. அது ஒருபோதும் நீக்கப்படாது. குறிப்பாகத் தமிழில், இந்தியில் கூட, நீங்கள் ஒருவரை 'சாப்ரி'[7] என்று அழைத்தால் அது புகாரளிக்கப்பட்டால், அது அகற்றப்படும். ஆனால், மக்கள் தமிழில் [இதே போன்ற] சொற்களைப் பயன்படுத்தும்போது, குறிப்பாக 'பறையர்' போன்ற அவதூறுகள், இது ஒரு குறிப்பிட்ட சாதியைச் சேர்ந்த பெண்களை விபச்சாரி என்று அவதூறாகப் பேசுவதற்குப் பயன்படும் ஒரு சொல்லாகும், அது எனக்கு எதிராக நிறையப் பயன்படுத்தப்பட்டுள்ளது... இந்த வார்த்தைகள் ஒருபோதும் அகற்றப்படுவதில்லை." (Tamil journalist and digital rights advocate, November 2024, India).

சாதி அடிப்படையிலான துன்புறுத்தல் தொடர்பான செய்திகளில், சமூக ஊடகத் தளங்கள் சாதி அடிப்படையிலான ஒழுங்குபடுத்தல் முறையை மேம்படுத்துவதில் முதலீடு செய்யத் தவறிவிட்டதாக நேர்காணலில் பங்கேற்றவர்கள் பரவலாக நம்பினார்கள். சாதி அடிப்படையில் இலக்கு வைத்துத் தாக்குதல்கள் மேற்கொள்ளப்படும்போது தகவல் ஒழுங்குபடுத்தல் கொள்கைகள் மற்றும் செயல்முறைகள் பெரும்பாலும் சீரற்ற முறையில் செயல்படுத்தப்படுகின்றன என்பதையும் ஆராய்ச்சி கண்டறிந்துள்ளது (Kain et. al, 2021; Yatharth, 2025; Soundararajan et al., 2019). 2019 ஆம் ஆண்டில், இந்தியாவில் பேஸ்புக்கில் காணப்படும் வெறுப்பூட்டும் பதிவுகளில் 13 சதவீதமானவை சாதி அடிப்படையிலான வெறுப்பூட்டும் பேச்சைக் கொண்டிருந்ததாக ஒரு ஆய்வு கண்டறிந்துள்ளது (Soundararajan et al., 2019).[8]

நிறுவனங்கள் சாதி அடிப்படையிலான தாக்குதல்களைக் கண்டறிவதற்காக தங்கள் அமைப்புகளைப் பயிற்றுவிப்பதில் உதவுவதற்கு, குழுக்களால் கூட்டாக உருவாக்கப்பட்ட மற்றும் நிபுணர்களால் உருவாக்கப்பட்ட அவதூறு வார்த்தைகள் கொண்ட

7 ஆன்லைனில் பரவலாகப் பயன்படுத்தப்பட்டதன் காரணமாக அதன் அர்த்தம் உருமாறிய ஒரு அவதூறு, மற்ற அவதூறுகளை விட அதிக ஆராய்ச்சிக்கான இணையத்தில் கவனத்தை அதிகம் பெற்றுள்ளது. (Mukherjee & Desai, 2024).

8 இந்த ஆய்வின் ஆசிரியர் 2022 ஆம் ஆண்டு ஒரு தொழில்நுட்ப நிறுவனத்தில் சாதி பற்றிப் பேச அழைக்கப்பட்டார். ஆனால், தலைப்பின் "பிரிவினையை ஏற்படுத்தும் மற்றும் புண்படுத்தும்" தன்மை காரணமாகப் பேச்சு ரத்து செய்யப்பட்டது. பேச்சை ரத்து செய்த நிறுவனத்தின் இந்த முடிவு குறிப்பிடத்தக்க எதிர்ப்புகளையும் ஊடகக் கவனத்தையும் பெற்றது (Tiku, 2022).

பட்டியல்கள் மற்றும் சொல்லகராதிகளை பயன்படுத்துகின்றன.[9] உதாரணமாக, மெட்டா நிறுவனம், ஆங்கிலம் அல்லாத மொழிகளில் அவதூறு பட்டியல்களை உருவாக்குவதற்கு கூட்டாண்மைகளை (partnerships) நம்பியிருப்பதாக தங்கள் வலைத்தளத்தில் வெளிப்படுத்துகிறது, ஆனால் அவர்கள் எப்படி, யாருடன் ஈடுபடுகிறார்கள் என்பது தெளிவாகத் தெரியவில்லை. மெட்டா நிறுவனம் இதனை தங்கள் பிராந்திய குழுக்களால் உருவாக்கப்பட்ட "சந்தைக்கே உரியப் பட்டியல்கள்" எனவும் "அவர்களின் பகுதி அல்லது சமூகத்தின் மொழி மற்றும் கலாச்சாரம் குறித்த தொடர்ச்சியான பண்பியல் மற்றும் அளவையியல் பகுப்பாய்வு" என்றும் அழைக்கிறது (Meta, 2022).[10] இந்த குழுக்கள் குறிப்பிட்ட துறையில் நிபுணத்துவம் பெற்றவர்களுடன் பேச்சுவார்த்தையில் ஈடுபடுகின்றனவா – அல்லது தகவல்களை ஒழுங்குபடுத்துவதில் உதவுவதற்காக வெளிப்படையாக உருவாக்கப்பட்ட திறந்த மூல பட்டியல்களைப் பயன்படுத்துகின்றனவா, அல்லது அந்த பட்டியலை மாநாடுகள் போன்ற பல நம்பிக்கை மற்றும் பாதுகாப்பு அரங்கங்களில் நிறுவனங்களுக்குக் கிடைக்கச் செய்கின்றனவா என்பது தெளிவற்றதாகவே காணப்படுகிறது.

ஆய்வில் பதிலளித்த பெரும்பாலானவர்கள் தாங்கள் பதிவுகளைப் பற்றி முன்பே புகாரளித்ததாகக் கூறினர், ஆனால் சில நேர்காணல் பங்கேற்பாளர்கள் அதை எப்படிச் செய்வது என்று தெரியவில்லை என்று கூறினர். இந்தியாவில் உள்ள ஒரு LGBTQ+ கருத்து படைப்பாளர், "சமூக ஊடகங்களின் உதவியை எவ்வாறு பெறுவது என்று எனக்குத் தெரியவில்லை" என்று கூறினார். இந்தக் கண்டுபிடிப்பு, உலகெங்கிலும் உள்ள பல பயனர்கள், குறிப்பாகப் பெண்கள், பாலின வேறுபாடு கொண்டவர்கள் மற்றும் ஓரங்கட்டப்பட்ட பயனர்களுக்கு, பிரபலமான இணையத் தளங்களில் பதிவுகளைப் பற்றிப் புகாரளிப்பது சவாலாக இருப்பதாகவும், இந்த செயல்முறைகள் ஏதோ ஒரு வகையில் குறைவாக, சீரற்றவையாக அல்லது குறைபாடுள்ளவையாக இருக்கின்றன என்பதைக் காட்டும் முந்தைய ஆராய்ச்சியுடன் ஒத்துப்போகிறது (Sehgal & Nambiar, 2024; Vilk 2023).

அதுமட்டுமின்றி, பல நேர்காணல் பங்கேற்பாளர்கள் இலக்கு வைக்கப்பட்ட வெறுப்பு பேச்சுக்கும், அவர்கள் புகாரளிக்கும் பதிவுகளுக்குப் பதில் கிடைக்காததற்கு, கேள்விக்குரிய பதிவுகள் ஆங்கிலத்தில் இல்லாதது மற்றும் ஒரு பயனராக அவர்கள் தங்களின் முதல் மொழியாக ஆங்கிலம் பேசாததுதான் காரணம் என்று கூறுகின்றனர். தமிழ் மற்றும் சாதி அடிப்படையிலான அவதூறுகள் அரிதாகவே கட்டுப்படுத்தப்படுகின்றன என்றும்,

9 Tattle ஒரு பட்டியலை உருவாக்கியது, இது தற்போதுள்ள நிலையில், தமிழ் உட்பட இந்திய மொழிகளில் 600க்கும் மேற்பட்ட அவதூறுச் சொற்களைக் கொண்டு உருவாக்கப்பட்ட மிகப்பெரிய பட்டியலாகும் (Tattle, 2024). புதிய மற்றும் வேறுபட்ட குறியீடு கலந்த சொற்கள் மற்றும் எழுத்துப்பிழைகளைச் சேர்ப்பதன் மூலம் அந்த குழு தொடர்ந்து பட்டியலைப் புதுப்பித்து, ஆராய்ச்சியாளர்கள் மற்றும் நம்பிக்கை மற்றும் பாதுகாப்புக் குழுக்களால் பயன்படுத்தக் கூடியவாறு திறந்த தரவு உரிமத்தின் கீழ் GitHub இல் இதனைக் கிடைக்கச் செய்கிறது.

10 இந்த ஆய்வு எழுதப்பட்ட காலத்தில், மெட்டா நிறுவனம் அதன் வெறுக்கத்தக்க நடத்தை கொள்கையில் மாற்றங்களைச் செய்துள்ளது. இந்த புதிய வெறுக்கத்தக்க நடத்தை கொள்கை மற்றும் பிற கொள்கைகளைச் செயல்படுத்த, முன்கூட்டியே கண்டறிதல் மற்றும் ஒழுங்குபடுத்தல் என்ற செயன்முறையை விட பயனாளர் புகார்களை நம்பியிருப்பதை அதிகரிக்க முயன்றது.

அதனால், சிலர் அவற்றைத் தண்டனையின்றி அல்லது ட்ரோல் செய்வதற்கான உரிமத்துடன் பயன்படுத்துகிறார்கள் என்றும் பல பங்கேற்பாளர்கள் நம்புகிறார்கள். இந்த அடையாளங்களின் ஒன்றோடு ஒன்று சேரும் தன்மை, ஒரு தனிமனிதனின் அடையாளங்களின் (மொழி, சாதி, வர்க்கம் மற்றும் பாலினம்) தனித்துவமான ஒடுக்குமுறைக்கு வழிவகுக்கிறது. இது ஒழுங்குபடுத்தல் செயல்முறையின் தாக்கங்களைப் புரிந்துகொள்வதற்கு முக்கியமானது:

> "உங்களுக்குத் தெரியும். எந்தவொரு இந்தியப் பெண்ணையும் நடத்தும் விதம் மிகவும் வித்தியாசமானது... ஆங்கிலம் பேசும் தலித் பெண்ணை நடத்தும் விதம் தமிழ் பேசும் தலித் பெண்ணை நடத்தும் விதத்திலிருந்து மிகவும் வித்தியாசமாக இருக்கும்." (தமிழ் பத்திரிகையாளர் மற்றும் டிஜிட்டல் உரிமைகள்களுக்கு குரல் கொடுப்பவர், நவம்பர் 2024, இந்தியா).

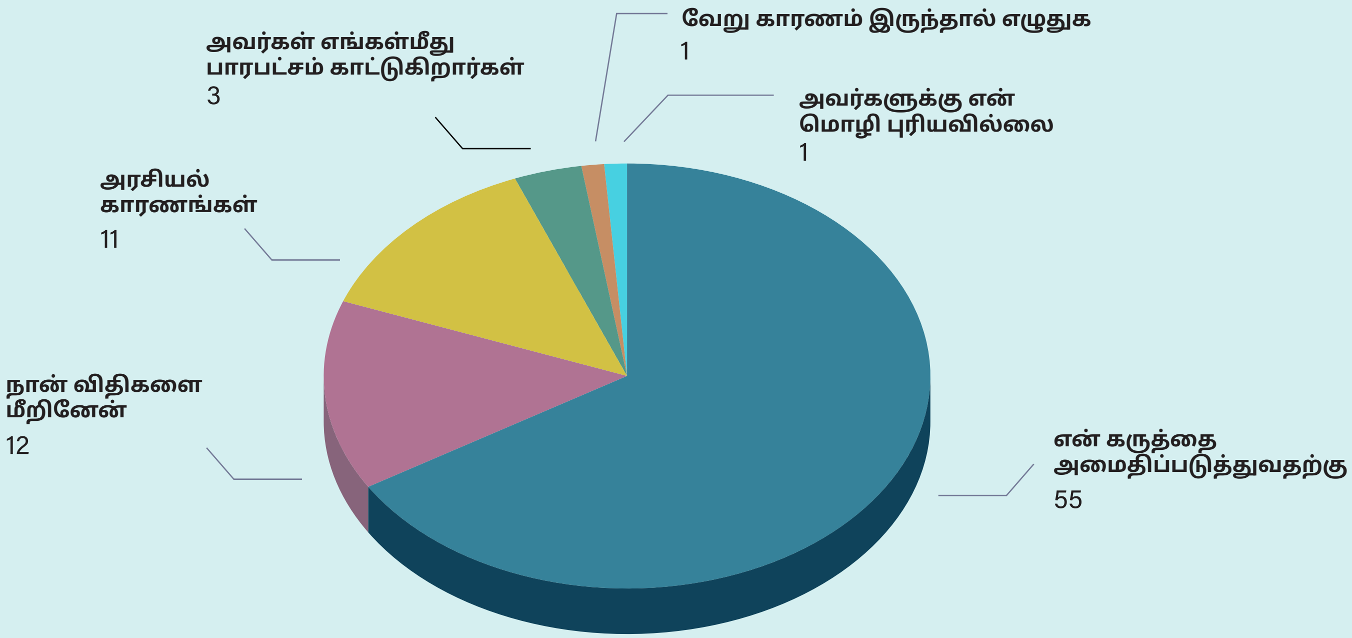
> "இந்த தளங்கள் எங்களுக்காக உருவாக்கப்படவில்லை. இது உலகில் நான் வசிக்கும் பகுதியிலிருக்கும் ஒரு குறிப்பிட்ட வகுப்பு பின்னணியைச் சேர்ந்தவர்களுக்காக வடிவமைக்கப்பட்டது. இது மேல் வகுப்பு பற்றியதும் கூட. முழுமையாக ஆங்கிலத்தில் கற்பிக்கப்படும் பள்ளிக்குச் செல்லக்கூடியவர்களைப் பற்றியது. அது இன்றும் ஒரு சிறப்புரிமையாகவே உள்ளது." (டிஜிட்டல் பாதுகாப்பு பயிற்சியாளர், நவம்பர் 2024, இந்தியாa).

3. பல பயனாளர்கள் சில நேரங்களில் ஒழுங்குபடுத்தும் செயல்முறையை அரசியல் காரணங்களுக்காக தங்கள் குரல்களை "அமைதிப்படுத்தும்" முயற்சியாகக் கருதுகின்றனர்.

கணக்கெடுப்பில் பதிலளித்தவர்களில் சுமார் 57% பேர் தங்கள் பதிவு ஒரு முறையாவது நீக்கப்பட்டதாகக் குறிப்பிட்டனர். பெரும்பான்மையானவர்கள் பதிவு அரசியல் இயல்புடையது என்று அடையாளம் கண்டனர். கணக்கெடுப்பில் பதிலளித்தவர்கள் தமது பதிவுகள் நீக்கப்படுவதையும் அவை மற்றவர்களிடம் சென்றடைவதை நிறுத்துவதையும், தளங்களில் தங்களின் குரல்களை "அமைதிப்படுத்துவதற்காக" மேற்கொள்ளப்படும் நடவடிக்கைகள் என நம்புகின்றனர். இந்த சந்தேகங்கள் பரவலாகக் காணப்பட்டன. தமிழ் படைப்பாளிகள், டிஜிட்டல் உரிமை ஆர்வலர்கள் மற்றும் தகவல் மதிப்பீட்டாளர்களுடனான நேரடி நேர்காணல்களில் அவை மீண்டும் மீண்டும் கூறப்பட்டன.

நீண்டகால தமிழ்க் கணினி நிபுணரும் முன்னாள் மதிப்பீட்டாளருமான ஒருவர், ஒழுங்குபடுத்தல் நடவடிக்கையை எதிர்கொண்ட தனது சொந்த அனுபவம் குறித்துக் கேட்டபோது பின்வருமாறு குறிப்பிட்டார்:

சமூக வலைதளங்கள் உங்கள் பதிவுகளை ஏன் நீக்கியதாக நினைக்கிறீர்கள்?



▲ **படம் 3:** சமூக ஊடகத் தளங்கள் ஏன் தங்கள் பதிவுகளை அகற்றின அல்லது கட்டுப்படுத்தின என்பதற்கான காரணங்களைப் பங்கேற்பாளர்கள் அடையாளம் கண்டுள்ளனர் (பதிவு நீக்கங்கள் அல்லது கட்டுப்பாடுகளை அனுபவித்ததாகக் கூறிய கணக்கெடுப்பில் பங்கேற்றவர்களின் %, n= 83).
இந்த தகவல் CDT நடத்திய இணையக் கருத்துக் கணிப்பிலிருந்து (நவம்பர்–டிசம்பர் 2024) பெறப்பட்டது.

“அது தமிழிலோ அல்லது ஆங்கிலத்திலோ இருந்தாலும் பரவாயில்லை, நீங்கள் அரசாங்கத்திற்கு எதிராகப் பேசும்போதும் அதேதான் நடக்கும். இது மொழியைப் பற்றியதல்ல, பாரபட்சத்தைப் பற்றியது. இந்தப் பதிவு ஸ்பேம் (Spam) எனப் புகாரளிக்கப்படும், மேலும் சமூக வழிகாட்டுதல்களை மீறுவதாகக் கருதி அகற்றப்படும் [...] சமூக வழிகாட்டுதல்கள் மிகவும் தெளிவற்றவை. இது யாரோ ஒருவர் அங்கே அமர்ந்து என்ன தவறு நடந்தது என்பதைப் பற்றி எங்களுக்குப் பதிலளிப்பது கிடையாது. "நீங்கள் சமூக வழிகாட்டுதல்களை மீறினீர்கள்" என்று அவர்கள் சொல்வார்கள். அதற்கு நாங்கள் எல்லா வழிகாட்டுதல்களையும் வாசிப்போம். நாங்கள் எதை மீறினோம் என்பது எங்களுக்குப் புரியாது. அது அவர்களின் சொந்த சட்டப் பாதுகாப்புக்காகத்தான் என்று நான் நினைக்கிறேன்." (முன்னாள் மதிப்பீட்டாளர், அக்டோபர் 2024, இந்தியா).

நேர்காணலில் பங்கேற்ற வேறொருவர், அவர்கள் உள்நாட்டுப் போரின் போது இலங்கையிலிருந்து துன்புறுத்தலில் இருந்து தப்பிச் செல்லும் தமிழர்கள் குறித்துப் பதிவிட்டதற்காக மேற்கொள்ளப்பட்ட ஒழுங்குபடுத்தல் நடவடிக்கை அரசியல் ரீதியானது எனச் சந்தேகித்தார். இலங்கையில் மேற்கொள்ளப்படும் மிகையான ஒழுங்குபடுத்தல் நடவடிக்கைகள் சமூக இயக்கங்களைத் தடுக்கிறது மற்றும் "பொறுப்புணர்வுக்கு இடையூறு விளைவிக்கிறது" என்று நீண்ட காலமாக வாதிடும் ஆராய்ச்சி மற்றும் மனித உரிமைகள் வாதத்தைப் பல நேர்காணல் பங்கேற்பாளர்கள் சுட்டிக்காட்டினர் (Amarasingam & Nandakumar, 2021; Ethirajan, 2021). நேர்காணலில் பங்கேற்ற இன்னுமொருவர், "பிரபாகரன்"

என்ற பொதுவான பெயரைக் கொண்ட கடை முகப்புகளின் படங்களை மக்கள் பதிவிட்டால் கூட, அது கொல்லப்பட்ட புலிகளின் தலைவரின் பெயராக இருப்பதனால், அவர்களின் பதிவுகள் நீக்கப்படும் என்று கூறினார். தமிழ் தேசியவாத இயக்கத்துடன் தொடர்புடையதாகத் தோன்றும் சின்னங்கள் சமூக ஊடகங்களில் கண்டறியப்பட்டு அகற்றப்பட்டதாக இன்னொரு நேர்காணல் பங்கேற்பாளர் கூறினார். உதாரணமாக கார்த்திகைப் பூ.[11] அரசாங்கத்திடமிருந்து அடிக்கடி கோரிக்கைகள் வருவதாக இலங்கை மற்றும் மேற்கத்திய ஊடக அறிக்கைகள் தெரிவிக்கின்றன போதிலும், அவர்களின் பதிவு நீக்கப்பட்டதற்கான காரணம் நீக்கப்படவேண்டும் என்ற அரசின் கோரிக்கையா அல்லது முன்கூட்டிய ஒழுங்குபடுத்தல் செயல்முறையா என்பது அவர்களுக்கு அறிவிக்கப்படவில்லை (Constine, 2018; Tamil Guardian, 2024; Mallwarachi, 2024):

> “வெகுஜன நிகழ்வுகள் இல்லாமல் இருந்த “COVID காலத்தில் ஒரு விஷயம் இருந்தது. நவம்பர் 27 ஆம் திகதிக்கான சில இணையச் செயல்பாடுகள் இருந்தன. மக்கள் தங்கள் மரியாதையை இணையத்தில் பகிர்ந்து கொண்டனர். இந்தக் காலகட்டத்தில் மக்கள் [LTTE] கொடியையோ அல்லது உறுப்பினர்களின் உண்மையான படங்களையோ அல்லது சீருடைகளையோ பதிவிடவில்லை, ஆனால் பதிவுகள் அகற்றப்பட்டுக் கொண்டிருந்தன. கார்த்திகைப் பூ, எனது தோழி ஒருவர் தனது தோழியுடன் லில்லி மலர்களை பிடித்துக்கொண்டு "நாங்கள் ஞாபகப்படுத்துகிறோம், எதிர்க்கிறோம்" என்று கூறும் புகைப்படத்தைப் பதிவிட்டார். அந்தப் பதிவு நீக்கப்பட்டது. அது அதிர்ச்சியாக இருந்தது. ஒரு பூ எப்படி வன்முறையாக இருக்க முடியும்? இதை அவர்கள் எப்படிக் கண்டுபிடித்தார்கள்?" (இலங்கை தமிழ் பத்திரிகையாளர் மற்றும் டிஜிட்டல் உரிமைகள்களுக்கு குரல் கொடுப்பவர், ஜனவரி 2025, UK).

சில நேர்காணல் பங்கேற்பாளர்கள் தமிழ் அல்லது வேறு மொழியில் சில சொற்களைப் பயன்படுத்தும்போது, "நிழல் தடை (Shadowbanning)" என்று பொதுவாக அழைக்கப்படும் மறைமுகமான ஒழுங்குபடுத்தல் நடவடிக்கையை அனுபவித்ததாகச் சந்தேகித்தனர். நிழல் தடை என்பது பரந்த அளவிலான வெளிப்படுத்தப்படாத தகவல் ஒழுங்குபடுத்தல் நடவடிக்கைகளை உள்ளடக்கிய ஒரு செயல்முறைக்கு வழங்கப்படும் பேச்சுவழக்கு சொல்லாகும். இதில் ஒரு பயனரின் பதிவுகளை மற்ற பயனர்களிடமிருந்து மறைத்தல், தேடலில் இருந்து ஒரு பயனரின் அடையாளம் அல்லது பதிவுகளை அகற்றுதல், அல்லது பயனர்களின் பதிவுகளை பரிந்துரை செய்யும் அமைப்பில் மிகக் குறைவாகத் தரவரிசைப்படுத்துதல் ஆகியவை அடங்கும். இதனால் அந்த பதிவுகளைக் குறைவாகவே காண முடியும் (Nicholas, 2022). ஒரு பயனரின் பதிவுகள் எந்தெந்த சூழ்நிலைகளில் அவருக்குத் தெரியாமல் ஒழுங்குபடுத்தப்படுகின்றன என்பதனை பயனாளர்களுக்குத் தளமானது தெளிவாக வெளிப்படுத்துவதில்லை. எனவே,

11 கார்த்திகைப் பூ தமிழ் தேசியவாத இயக்கத்தின் அடையாளமாகவும் தேசிய மலராகவும் பார்க்கப்படுகிறது, இது தமிழீழ விடுதலைப் புலிகள் (LTTE) உடன் ஒத்ததாக மாறியுள்ளது. ஏனெனில், இது தமிழீழ தேசியக் கொடியில் உள்ள அனைத்து வண்ணங்களையும் கொண்டுள்ளது.

பயனர்கள் தங்கள் பதிவுகள் ஏன் குறைவாக மற்றவர்களைச் சென்றடைகின்றன என்பதற்கான காரணத்தைத் தாமே உருவாக்கி, இந்த வகையான தலையீட்டைத் தவிர்ப்பதற்கான வழிகளைக் கண்டுபிடிக்கின்றனர் என்பதனை முந்தைய ஆய்வு வெளிக்காட்டியது. (Savolainen, 2022; Nicholas, 2022). நாங்கள் நேர்காணல் செய்த ஒரு தமிழ் அரசியல் விமர்சகர், அவர்களின் சில அரசியல் பதிவுகள் குறைவான ஈடுபாட்டைப் பெறுகின்றன என்பதை வலியுறுத்தினார்:

> "குறிப்பாக நான் RSS[12] பற்றி நிறைய எழுதும்போது, அது கீழ்நிலைஊட்டத்திற்குச் செல்கிறது. என்னை 54,000 பின்தொடர்கிறார்கள். இருந்தாலும், என் பதிவு மிகக் குறைவாகவே சென்றடைகிறது. காசா அல்லது பாலஸ்தீனம் அல்லது RSS அல்லது BJP என்ற வார்த்தையை நீங்கள் பயன்படுத்தும் சில பதிவுகள், உங்களுக்கு மட்டுமே கிடைக்கும் என்பது உங்களுக்குத் தெரியும்… நான் கண்களை மூடிக்கொண்டு 10 நிமிடங்களில் எனக்கு மூன்று லைக்குகள் கிடைக்கும் என்று சொல்ல முடியும். அது நிச்சயம். அது கீழ்நிலை ஊட்டத்திற்கு நகர்த்தப்பட்டுள்ளது என்பதை நான் அப்படித்தான் அறிவேன்." (தமிழ் பத்திரிகையாளர் மற்றும் டிஜிட்டல் உரிமைகள்களுக்கு குரல் கொடுப்பவர், நவம்பர் 2024, இந்தியா).

தகவல் ஒழுங்குபடுத்தலை வடிவமைக்கத் தளங்கள் அரசாங்கத்தின் உள்ளீட்டை நம்பியிருப்பதையும், இதனை அவர்கள் பயனர்களுக்கு வெளியிடுவதில்லை என்பதையும் நாங்கள் கண்டறிந்தோம். தளங்களின் முன்னாள் மற்றும் தற்போதைய பிரதிநிதிகளுடனான எங்கள் நேர்காணல்களின் அடிப்படையில், தமிழில் தகவல் ஒழுங்குபடுத்தல் செயல்முறையைப் பின்வரும் வழிகளில் வடிவமைக்கத் தளங்கள் பெரும்பாலும் அரசாங்க வழிகாட்டுதலை நம்பியுள்ளன.

முதலாவதாக, இந்திய மற்றும் மேற்கத்தியச் சமூக ஊடக தளங்கள் பெரும்பாலும் பயனர்களுக்குச் சிறிய அறிவிப்புடன், பதிவை நீக்குதல் அல்லது பயனரின் தகவல்களை வழங்குதல் போன்ற அரசாங்கத்தின் கோரிக்கைகளுக்கு இணங்குகின்றன. ஒரு மதிப்பீட்டாளர் பின்வருமாறு விவரித்தார்:

> “ஒரு குறிப்பிட்ட பதிவை அகற்றுவதற்கான கோரிக்கை, அரசாங்கத்திடமிருந்து வந்தால், சட்டப்பூர்வமாக அந்த கோரிக்கையை நிறைவேற்ற வேண்டும் என்று அந்த நாடு கேட்டுக்கொண்டால், [தளம்] அந்த பதிவை அகற்றக் கடமைப்பட்டிருக்கிறது.” (சமூக ஊடக தளத்தில் மதிப்பீட்டாளர், டிசம்பர் 2024, இந்தியா).

இரண்டாவதாக, இடைத்தரகர்கள் தொடர்பான இந்திய மற்றும் இலங்கை சட்டங்கள் பெரும்பாலும் இந்திய மற்றும் மேற்கத்திய சமூக ஊடக தளங்களால் மிகைப்படுத்தப்பட்ட மற்றும் சொந்த வழிகளில் விளக்கப்படுகின்றன. இதன் விளைவாகத் தமிழ்

---

12 ராஷ்ட்ரிய சுயம்சேவக் சிங் அல்லது ஆர்.எஸ்.எஸ் என்பது ஒரு இந்து தேசியவாத அமைப்பாகும், மேலும் பாரதீய ஜனதா கட்சியின் (பிஜேபி) தாய் அமைப்பாகும். இது ஒரு அரசியல் கட்சியாகும், தற்போது இந்தியாவில் மத்திய ஆட்சியில் இருக்கும் கட்சியாகும். (Britannica, n.d.).

பதிவு நீக்கப்படுகிறது. இந்தியாவில் ஐடி சட்டம் மற்றும் இலங்கையில் 9ம் இலக்க இணையப் பாதுகாப்பு சட்டம் போன்ற சட்டங்கள், அந்தந்த அதிகார வரம்புகளில் பயனர் உருவாக்கிய பதிவை வழங்கும் இணையத் தளங்களுக்குச் சட்டவிரோத அல்லது தீங்கு விளைவிக்கும் பதிவுகளை அகற்ற வேண்டும் என்று கோரிக்கை விடுகின்றன. எனினும், அவை வெவ்வேறு அளவுகளில் காணப்படுகிறது. இந்தத் தேவைகள் இணையச் சேவைகளால் வெவ்வேறு வழிகளில் விளக்கப்பட்டுள்ளன. மேலும், அவை பெரும்பாலும் சிறிய, உள்ளூர் இடைத்தரகர்களுக்குச் சுமையாக இருக்கின்றன (Kumar et. al, 2022). இதன் விளைவாக இந்தியத் தளங்களோடு சேர்ந்து , மேற்கத்தியத் தளங்கள் கூட, அவற்றை நேரடியாக தங்கள் சமூக வழிகாட்டுதல்களிலும், மதிப்பீட்டாளர்களுக்கான வழிகாட்டுதலிலும் இணைத்துக்கொள்கின்றன. இது பின்வருமாறு ஒரு இந்தியச் சமூக ஊடக ஊடக நிறுவனத்தின் கொள்கைக் குழுவின் தலைவர் குறிப்பிட்டதைப் போன்றதாகும்.

> “பொதுவான கொள்கை ஐடி சட்டத்திலிருந்து வந்தது… பொது ஒற்றுமையைச் சீர்குலைக்கும் அல்லது முரண்பாட்டை உருவாக்கும் கருத்துகளை நீங்கள் கூற முடியாது. அதாவது, கடவுள்களைப் பற்றி நீங்கள் கருத்துகளைக் கூற முடியாது…உதாரணமாக, இந்திய சட்டம் தேசியக் கொடியை இழிவுபடுத்த அனுமதிக்காது, எனவே தேசியக் கொடியை இழிவுபடுத்தும் பதிவை நீங்கள் பதிவிட முடியாது. எங்கள் சமூக தரநிலைகள் பெரும்பாலும் அதன் ஒரு துணைக்குழுவாகும்… ஒரு தளத்தில் ஏதாவது நடந்தால், காவல்துறை உங்களைத் தொடர்பு கொள்ளும். நீங்கள் அதைப் பதிவிட்டால், நீங்கள் ஒரு கலவரத்தைத் தொடங்கினால், காவல்துறை உங்கள் விவரங்களை எங்களிடம் கேட்பார்கள். உங்கள் விவரங்களை நாங்கள் வழங்குவோம்… இதேபோல், பதிவுகளை அகற்றவோ அல்லது கவனிக்கவோ நிறுவனங்களைக் கேட்ட பல்வேறு சந்தர்ப்பங்களும் உள்ளன…நாங்கள் இதைப் பற்றி இன்னும் கொஞ்சம் முன்கூட்டியே செயல்படுவோம்." (கொள்கைத் தலைவர், இந்திய சமூக ஊடக நிறுவனம், அக்டோபர் 2024, இந்தியா)

இந்தச் சட்டங்கள் பெருமளவில் பதிவுகள் நீக்கப்படுவதற்குக் காரணமாகின்றன, இதன் விளைவாகத் தளத்தின் ஒழுங்குபடுத்தும் செயல்முறைகளில் அவநம்பிக்கை ஏற்படுகிறது. இது பயனாளர்கள் மேல்முறையீட்டுச் செயல்முறைகளையும் தகவல் ஒழுங்குபடுத்தல் செயல்முறைகளையும் கூட முற்றிலுமாகத் தவிர்க்கக் காரணமாகிறது. நம்பிக்கையைப் பேணுவதற்கும், இணக்கத்திற்கும், பயனர் நம்பிக்கைக்கும் இடையிலான ஒரு சாமர்த்தியமான சமநிலையை வழிநடத்துவதற்கும் பயனர்களுக்கு வெளிப்படையாக இருப்பது அவசியம் என்று ஒரு இணைய மன்றத்தின் மதிப்பீட்டாளர் கூறினார்:

> "நான் கூறுவது தவறாக இருக்கலாம், ஆனால் அரசாங்கத்தின் வேண்டுகோளின் பேரில் நான் தொடர்ந்து பதிவுகளை நீக்கிக்கொண்டிருந்தால், பயனர்கள் வந்து பேசுவதற்கு இது ஒரு பாதுகாப்பான

> இடம் என்று நம்ப மாட்டார்கள். எனவே நான் பதிவுகளை நீக்கினால், அவர்களிடம், "அரசாங்கம் அதை நீக்க வேண்டும் என்று கூறியது" என்று கூறுவேன். இது இரண்டாவது முறையாக நடந்தால், "பாருங்கள், இது ஒருவித அரசாங்க விதி. நீங்கள் விதி மீறாமல் பார்த்துக் கொள்ள வேண்டிய கொள்கைகள் இவை" என்று சொல்ல எனக்கு ஒரு நிலையான கொள்கை இருக்க வேண்டும். அரசாங்கம் மீண்டும் மீண்டும் உங்களிடம் வந்து, அதையும் இதையும் அகற்று என்று கூறுவதையும் நீங்கள் விரும்பமாட்டீர்கள். (ஒரு சிறிய ஆன்லைன் மன்றத்தில் மதிப்பீட்டாளர், நவம்பர் 2024, இந்தியா).

அமெரிக்காவை மையமாகக் கொண்ட ஒரு சமூக ஊடக நிறுவனத்தில் பணிபுரியும் ஒரு ஊழியர் கூறுகையில், நிறுவனங்கள் அரசாங்கத்துடன் தங்களை இணைத்துக் கொள்கின்றன என்ற சில சமூக ஊடக பயனர்களின் கருத்தின் விளைவாக, தளத்தின் ஒழுங்குபடுத்தல் செயற்பாட்டில் அவநம்பிக்கை ஏற்படுகிறது எனவும், இது கல்வியாளர்கள், சிவில் சமூகம் மற்றும் ஊடகங்களால் நீண்ட காலமாக ஆவணப்படுத்தப்பட்டுள்ளது எனவும் கூறினார். (Sombatpoonsiri & Mahapatra, 2024; Horowitz & Purnell, 2020; Mirza, 2023).

இறுதியாக, மேற்கத்தியச் சமூக ஊடக தளங்கள் பெரும்பாலும் அரசாங்கங்களின் அழுத்தத்திற்குப் பிறகுதான் பதிவுகளை ஒழுங்குபடுத்தும் திறனில் அதிகமாக முதலீடு செய்ய ஆரம்பித்தன. நம்பிக்கை மற்றும் பாதுகாப்பு தொடர்பான ஊழியர்களுடனான நேர்காணல்களில், குழுக்கள் அரசாங்கத்தைத் தவிர வேறு யார் சொல்வதையும் கேட்பதில்லை என்ற உணர்வு இருந்தது. உதாரணமாக, 2019 ஆம் ஆண்டு ஈஸ்டர் ஞாயிறு தாக்குதலுக்குப் பிறகு, இலங்கை அரசாங்கம் அவசரக்கால நடவடிக்கையின் கீழ் சமூக ஊடக தளங்களுக்கான அணுகலைக் கட்டுப்படுத்தியது, இதன்போது தங்கள் நிறுவனம் தமிழ் தகவல்களை ஒழுங்குபடுத்துவதற்காக "அவசரக்கால தீவிர நடவடிக்கைகளை" எடுத்ததாக ஒரு முன்னாள் கொள்கை ஊழியர் குறிப்பிட்டார்:

> “அவசரக்கால நடவடிக்கைகள் மேற்கொள்ளப்பட்டன , உதாரணமாக, ஈஸ்டர் ஞாயிறு தாக்குதல்களுக்குப் பிறகு, நிறுவனத்திற்குத் தமிழ் மற்றும் சிங்களத்தில் வகைப்படுத்திகள் இல்லாததால் வெறுப்புப் பேச்சுக்கள் அனைத்தும் கட்டுப்படுத்தப்படாமல் போனது என்றும் பிரதானமாக விளக்கம் கூறப்பட்டது. நிச்சயமாக நடந்ததும் அதுதான். ஆனால் அப்போது ஒரு கட்டத்தில், நிறுவனம், தனக்குள்ளாக 'கிரீன் லான்டர்ன்' என்று அழைக்கப்படும் ஒன்றைப் பயன்படுத்தியது. இது வெறுப்புப் பேச்சு வகைப்படுத்தி என்று நான் நினைக்கிறேன். அதனால், பகுதிசார் மொழிகளில் வெறுப்புப் பேச்சைக் கட்டுப்படுத்த நிறுவனத்தால் முடிந்தது. நிறுவனங்கள் தங்களுக்கு இந்தத் திறன்கள் உள்ளன என்பது வெளியில் தெரியப்படுவதை விரும்புவதில்லை என்பதால், இதைச் சுற்றி நிறைய ரகசியம் இருந்தது. எனவே, மொழிவளங்களில் உண்மையான இடைவெளி உள்ளதா, அல்லது அது நிறுவனத்தின் விருப்பமின்மையா அல்லது அவர்கள்

> அந்த வளங்களைப் பயன்படுத்தவேண்டும் என்பது பற்றித் திட்டமிடவில்லையா என்பதில் சிறிது தெளிவின்மை உள்ளது." (ன்ெடுத்ப்ெடுவற்கு ஒரு வருடத்ிற்கு முன்பு, நிறுவனங்கள்
> பைிதானா கண்டுபிடிப்புகள்
>
> அமெரிக்காவை தளமாகக் கொண்ட சமூக ஊடக நிறுவனத்தில் கொள்கை, டிசம்பர் 2024, இந்தியா.).

2018 ஆம் ஆண்டில், அதாவது அந்த அவசரக்கால நடவடிக்கைகள் பயன்படுத்தப்படுவதற்கு ஒரு வருடத்திற்கு முன்பு, நிறுவனங்கள் பதிவுகளைக் கட்டுப்படுத்தத் தவறியதால், இலங்கை மக்களுக்கு எதிராகக் கலவரங்கள் தூண்டப்பட்டதாகக் குற்றம் சாட்டப்பட்டது. அந்தக் கலவரங்களுக்குப் பதிலளிக்கும் விதமாக, இலங்கை அரசாங்கம் அவசர உத்தரவின் கீழ் சமூக ஊடக சேவைகளுக்கான அனுமதியைத் தடுத்தது. 2020 ஆம் ஆண்டில், தனது சேவைகள் தடை செய்யப்பட்ட நிறுவனங்களில் ஒன்றான, மெட்டா நிறுவனத்தின் விசாரணைக்குப் பிறகு, நிறுவனம், இதனால் ஏற்பட்ட உண்மையான மனித உரிமை தாக்கங்களை ஏற்றுக்கொண்டு மன்னிப்பு கோரியது (Brustein, 2020). அதிகமாகச் சிங்களம் மற்றும் தமிழ் பேசும் ஊழியர்களை பணியமர்த்தவும், பாதிக்கப்படக்கூடிய குழுக்களான இஸ்லாமியர்கள் மற்றும் தமிழர்களைப் பாதுகாக்கச் சிங்கள மொழியில் கண்டறிதல் தொழில்நுட்பத்தைப் பயன்படுத்தவும் நிறுவனம் உறுதிபூண்டது (பேஸ்புக் இலங்கை மனித உரிமைகள் தாக்க மதிப்பீடு, 2020).[13]

இறுதியில், நேர்காணலில் பங்கேற்றவர்கள், உண்மையான ஆபத்து, வளர்ந்து வரும் சுய-தணிக்கை முறையில் உள்ளது என்றும், இது பேச்சு சுதந்திரத்தையும், பிராந்தியத்தில் அரசியல் சூழலையும் அச்சுறுத்துகிறது என்றும் வாதிட்டனர்.

> "நீங்கள் எழுதுவதற்கும் பேசுவதற்கும் முன்பு நிறையச் சிந்திக்க வேண்டும். மக்கள் தானாகவே சுய தணிக்கை செய்து கொள்வார்கள். தளங்கள் அவர்களைத் தணிக்கை செய்ய விரும்பவில்லை என்றால், அவர்கள் தங்களைத் தாங்களே தணிக்கை செய்து கொள்வார்கள். ஒரு தளம் நன்றாக இல்லை என்றால், நீங்கள் மற்றொரு தளத்தைக் நாடலாம். அல்லது உங்கள் சொந்த வலைத்தளம் அல்லது உங்கள் சொந்த சர்வரை வைத்துக்கொள்ளலாம். தளங்கள் உங்களை வீழ்த்துவது உண்மையில் மிகச் சிறிய விஷயம். உண்மையான ஆபத்து நாட்டின் ஜனநாயக சூழலாகும். நாம் நம்மை ஆபத்தில் ஆழ்த்துகிறோம். தளங்களை விட, உண்மையான ஆபத்து நீங்கள் செயல்படும் நாட்டின் ஜனநாயக சூழலில்தான் உள்ளது" (Former moderator, October 2024, India).

4. தமிழ் இயற்கை மொழி செயலாக்க (NLP) ஆராய்ச்சி

13 மெட்டா நிறுவனம் இந்தியாவில், மனித உரிமைகளின் தாக்கம் குறித்த மதிப்பீடு ஒன்றை நடத்தியது, சர்வதேச சிவில் சமூகம் கேட்டுக்கொண்ட போதிலும் அதை ஒருபோதும் வெளியிடவில்லை (Access Now, 2022).

மற்றும் தமிழ் தகவல் பகுப்பாய்வு திறன்களில் முன்னேற்றங்கள் இருந்தபோதிலும், குறைந்த நிதி முதலீடு மற்றும் தமிழ் இயற்கை மொழி செயலாக்க நிபுணர்களிடம் போதுமான ஈடுபாடு இல்லாதது போன்றவற்றின் காரணமாக, சமூக ஊடக நிறுவனங்கள் தமிழுக்கான தானியங்கி மதிப்பீட்டுக் கருவிகளைப் பயன்படுத்துவதில் மெதுவாகச் செயல்படுகின்றன.

தமிழ்ச் சமூகம் ஒரு வளமான மற்றும் ஒருங்கிணைக்கப்பட்ட கணினி பாரம்பரியத்தைக் கொண்டுள்ளது, மேலும் சமூக ஒழுங்குபடுத்தல் அல்லது திறந்த அறிவுப் பகிர்வில் ஈடுபட்டுள்ள பலர் குறைந்தபட்சம் 2000களின் முற்பகுதியிலிருந்தே ஒழுங்குபடுத்தும் செயல்முறையை உருவாக்கி விவாதித்து வருகின்றனர். நாங்கள் பேசிய பல தமிழ்க் கணினி வல்லுநர்கள் மதிப்பீட்டாளர்களாக அல்லது சமூக நிர்வாகிகளாக தமது பணியைத் தொடங்கியவர்கள். மேலும், இவர்கள் ஆரோக்கியமான இணைய சமூகங்களை வளர்ப்பதன் மூலம் ஈடுபாட்டை ஊக்குவிப்பதன் நுட்பமான சமநிலையை நன்கு அறிந்தவர்கள்:

> “அந்தக் காலத்தில், தமிழ் வலைப்பதிவுகளின் வளமான சூழல் அமைப்பு இருந்தது. வேறு இந்திய மொழிகளை விட முன்னரே தமிழ் வலைப்பதிவு சமூகமொன்று இருந்தது. தமிழ் டெவலப்பர்களால் உருவாக்கப்பட்ட வலைப்பதிவு திரட்டிகள் பிரபலமாக இருந்தன. நான் 2003, 2002, 2004 பற்றிப் பேசுகிறேன். இந்த வலைப்பதிவு சமூகம் முழுவதும் திறந்த இணையத்தில் செழித்து வளர்ந்தது… பலதரப்பட்ட பங்களிப்பாளர்கள் பொறுத்தவரை, இந்தியாவில் தமிழ் கணினிச் சமூகம் முன்னோடி சமூகங்களில் ஒன்று என்று நான் கூறுவேன்.” (தமிழ் NLP நிபுணர் மற்றும் முன்னாள் மதிப்பீட்டாளர், நவம்பர் 2024, இந்தியா).

தமிழ் மதிப்பீட்டாளர்கள் மற்றும் கணினி வல்லுநர்கள் தங்கள் மொழியின் மீதான அபரிமிதமான ஆர்வத்தால் உந்தப்படுகிறார்கள். ஒரு கணினி நிபுணர், தனது அடையாளம் மற்றும் உள்நாட்டுப் போரின் துன்பத்திலிருந்து தப்பியோடிய அனுபவம் பற்றிப் பேசினார். தனது வேலையைப் பற்றிப் பேசியதுடன், தமிழ் புலமையில் முதலீடு செய்வதன் முக்கியத்துவத்தைச் சுட்டிக்காட்டினார்:

> “பாரம்பரிய (அச்சு) முறைகளுக்கு வெளியே, மொழி வளங்களையும், கலாச்சார பாரம்பரியத்தையும் பாதுகாக்க வேண்டியதன் அவசியத்தில் நாம் செய்யும் பல பணிகள் உள்ளன. எனவே, நூலகம்[14] என்பது டிஜிட்டல் மயமாக்குவதற்கான முயற்சியாகத் தொடங்கியது. "நூலகம்" என்பது அச்சிடப்பட்ட பொருட்கள் மற்றும் பல்லூடக ஆவணங்களின் வரலாறுகளை டிஜிட்டல் மயமாக்கும் ஒரு திட்டமாகும். டிஜிட்டல் தொழில்நுட்பங்கள் வளங்களைப் பரவலாகக் கிடைக்கச் செய்வதற்கு எங்களுக்கு உதவியது.” (தமிழ் NLP நிபுணர் மற்றும் முன்னாள் மதிப்பீட்டாளர், அக்டோபர் 2024, இந்தியா).

தமிழ் மதிப்பீட்டாளர்கள் மற்றும் கணினி ஆர்வலர்களின் வலுவான மற்றும் துடிப்பான சமூகம் இருந்தபோதிலும், மேற்கத்தியச் சமூக ஊடக நிறுவனங்களின் தமிழ் மொழி கணினி நிபுணர்களுடனான ஈடுபாடு மிகக் குறைவாகவே உள்ளது. மேலும், தமிழ் மொழி தொடர்ந்தும் குறைந்த வளம்கொண்ட மொழியாகவே உள்ளது. தானியங்கி தகவல் ஒழுங்குபடுத்தல் அமைப்புகளின் வளர்ச்சியின் அடிப்படையில் இது குறிப்பாக உண்மையாகவே இருக்கிறது.

இணையச் சேவைகள் தகவல்கள் பரிமாறும் அளவைக் கருத்தில் கொண்டு, தானியங்கி தகவல் ஒழுங்குபடுத்தல் அமைப்புகளை பொதுவாகப் பதிவுகளை ஒழுங்குபடுத்தும் கருவிகளாகப் பயன்படுத்துகின்றன (Duarte & Llansò, 2017). இருப்பினும், பயிற்சித் தரவுகள் குறைவாகக் கிடைப்பது, இந்தக் கருவிகளைப் பெறுவதற்கான செலவு (ஒரு இந்தியச் சமூக ஊடக நிறுவனம் குறிப்பிட்டது போல) அல்லது இந்திய மொழிகளில் முதலீடு செய்வதற்கான குறைந்த விருப்பம் காரணமாக, இணையச் சேவைகள் தமிழில் தானியங்கி அமைப்புகளை உருவாக்குவதில் முதலீடு செய்வதில்லை. தமிழ் இயற்கை மொழி செயலாக்க வட்டமேசை உரையாடலின் போது நிபுணர்கள் எங்களிடம் கூறியதாவது, மொழியின் மொழியியல் சிக்கலைத் துல்லியமாகப் பிரதிபலிக்கும் தரவுத்தொகுப்புகள் காணப்படாமல் இருத்தல், தமிழ் பேசுபவர்கள் தங்கிலீசு அல்லது தமிழிஷ் மற்றும் இணையத்தில் அல்கோ-ஸ்பீக் போன்றவற்றை அதிகமாக பயன்படுத்துதல் மற்றும் மொழியின் தனித்துவமான சூழல்-இயல்பு காரணமாக தற்போதுள்ள தகவல் ஒழுங்குபடுத்தல் மற்றும் உணர்வு பகுப்பாய்வு அமைப்புகள் பெரும்பாலும் தோல்வியடைகின்றன. அதுமட்டுமன்றி, சில தானியங்கி தகவல் ஒழுங்குபடுத்தல் அமைப்புகள் சொந்தமொழி பேசுபவர்களால் உருவாக்கப்பட்ட மதிப்பீட்டு அளவீடுகளைப் பயன்படுத்திச் சோதிக்கப்படுகின்றன.

ஒரு தமிழ் நிபுணர் சொன்னதாவது, தமிழ் நிபுணர்களுடன்

---

14 தமிழ் கலைப்பொருட்களைப் பாதுகாப்பதற்கான நிலையான முறைகளுக்கு டிஜிட்டல் திறந்த அணுகல் காப்பகங்கள் ஒரு சிறந்த அமைப்பாக இருக்கும் என்று காப்பக வல்லுநர்கள் கூறுகின்றனர். "நூலகம்" அறக்கட்டளை அத்தகைய ஒரு காப்பகமாகும், இது இலங்கையில் தகவல் ஆதாரங்களுக்கான அணுகலை வழங்கும் மற்றும் "அறிவு சார்ந்த வளர்ச்சியை வளர்க்கும்" ஒரு இலவச இணைய நூலகமாகச் செயல்படுகிறது.. இது புத்தகங்கள், பத்திரிகைகள், செய்தித்தாள்கள், துண்டுப்பிரசுரங்கள், கையெழுத்துப் பிரதிகள் மற்றும் பலவற்றை உள்ளடக்கிய 100,000 க்கும் மேற்பட்ட பதிவுகளைக் கொண்டுள்ளது (Noolaham Foundation, n.d; Daniel, 2016).

நிறுவனங்கள் இணைந்து செயல்படும் தன்மை குறைவாக இருப்பது சமீபத்திய போக்கு. சேவைகளின் முகப்பக்கத்தை மொழிபெயர்க்க அமெரிக்காவைத் தளமாகக் கொண்ட ஒரு சமூக ஊடக நிறுவனம் முன்பு அவரை ஈடுபடுத்தியது.[15] இதற்கு நேர்மாறாக, நாங்கள் பேசிய ஒரு தமிழ் உண்மைச் சரிபார்ப்பாளர் (fact-checker), தாங்கள் உண்மைச் சரிபார்ப்பு செய்த ஒரு நிறுவனத்திடம் அதன் வழங்கப்பட்ட தானியங்கி கருவிகளை மேம்படுத்துமாறு கேட்டபோது, அந்த நிறுவனம் பின்வருமாறு பதிலளித்ததாகக் குறிப்பிட்டார்:

> "உதாரணமாக, நான்கு ஆண்டுகளுக்கு முன்பு, [சமூக ஊடக நிறுவனம்] மொழிபெயர்ப்பு துல்லிய மொழி கருவிகளையும் பின்னர் படியெடுத்தலையும் (Transcription) செய்ய முயற்சித்தது. படியெடுத்தல் கருவிகள் ஆங்கிலத்தில் இருந்து தமிழுக்கு வசன வரிகள் இட (Subtitles) [உதவி] செய்தன, ஆனால் நான்கு ஆண்டுகளுக்கு முன்பு ஆங்கிலத்தில் இருந்து தமிழுக்கு வசன வரிகள் இட [எந்த கருவிகளும் இல்லை]. கலிஃபோர்னியாவிலிருந்து ஒரு கருவியைப் பற்றி [நான் அறிந்தேன்]. சம்பந்தப்பட்ட நிறுவனத்திற்கு நான் ஒரு மின்னஞ்சல் அனுப்பினேன். தமிழில் ஏன் ஆதரவு அளிக்கக்கூடாது என்று கேட்டேன். அந்தப் பதில்தான் என் கண்களைத் திறந்து வைத்தது. அவர்கள், "உலக மக்கள்தொகையில் தமிழ் மொழி 2% க்கும் குறைவாக உள்ளது. எங்கள் நிறுவனம் ஏன் 2% க்கும் குறைவான மக்களுக்கு ஒரு கருவியைத் தயாரிக்க வேண்டும்? நீங்கள் விரும்பினால், நீங்கள் அதைச் செய்யலாம்" என்று சொன்னார்கள். (Tamil fact-checker, December 2024, India).

தமிழில் தரவுகள் கிடைப்பதை மேம்படுத்தவும், மொழியில் தானியங்கி கருவிகளை உருவாக்கவும் பல தமிழ் பேசுபவர்களும் இயற்கை மொழி செயலாக்க நிபுணர்களும் தனிப்பட்ட முறையில் முதலீடு செய்துள்ளனர் மேலும் மாநில அரசுகளிடமிருந்து மானியங்களைக் கோரியுள்ளனர். தமிழ்த் தேசிய மொழி செயலாக்க ஆராய்ச்சி மையம் போன்ற பல ஆராய்ச்சி கூட்டமைப்புகள் மற்றும் திட்டங்கள், மைக்ரோசாஃப்ட் ஆராய்ச்சி மற்றும் ப்ராஜெக்ட் வாணி தலைமையிலான முயற்சிகள் தமிழில் தானியங்கி தொழில்நுட்பங்களை உருவாக்குவதிலும் மேம்படுத்துவதிலும் வெளிப்படையாகக் கவனம் செலுத்துகின்றன. கணினி மொழியியல் சங்கத்தின் திராவிட மொழி தொழில்நுட்பப் பட்டறை மற்றும் தமிழ்க் கணினிக்கான சர்வதேச மாநாடு போன்ற அதிநவீன தமிழ் மொழி தொழில்நுட்பங்களை மேம்படுத்துவதற்குப் பல அரங்கங்கள் உள்ளன. பயிற்சி தரவுத்தொகுப்புகளில் இந்திய மொழிகளிடையே தமிழ் பிரதிநிதித்துவத்தை மேம்படுத்துவதில் கவனம் செலுத்துகின்ற AI for Bharat மற்றும் Karya போன்ற முயற்சிகளும் உள்ளன. ரூபா பிரியதர்ஷினி மற்றும் பாரதி சக்ரவர்த்தி போன்ற முன்னணி கல்வியாளர்கள், தமிழ் இணையத்தில் பயன்படுத்துவதைப் பற்றி ஆய்வு செய்து, தமிழில் உணர்வு பகுப்பாய்வு மற்றும் வெறுப்புப் பேச்சு

15 இந்த சேவையின் மொழிபெயர்க்கப்பட்ட பதிப்புகளுக்கு நிபுணர்களின் ஒன்றிணைந்த ஒத்துழைப்பு பெறப்பட்டது, ஆராய்ச்சியாளர்களின் கூற்றுப்படி இந்தத் திட்டம் நிராகரிக்கப்பட்டதாகச் சந்தேகிக்கப்படுகிறது (u/diegomrosa, 2023).

கண்டறிதலை எளிதாக்கும் வகையில் தமிழ் உரையின் தொகுப்பை உருவாக்கிய முன்னணி கல்வியாளர்களாக மீண்டும் மீண்டும் குறிப்பிடப்படுகிறார்கள். (Chakravarthi et. al, 2020). சமூக ஊடக நிறுவனங்கள் தங்கள் தமிழ் கருவிகளை மேம்படுத்த முயற்சிக்கும் வழிகளுக்கான சில எடுத்துக்காட்டுகள் இவை எனத் தமிழ் இயற்கை மொழி செயலாக்க நிபுணர்கள் கூறியுள்ளனர்.

நேர்காணலில் பங்கேற்றவர்கள் நிறுவனங்கள் இயற்கை மொழி செயலாக்க நிதிகளை முதலீடு செய்வதில் முன்னுரிமை தரவேண்டிய பகுதிகளை அடையாளம் கண்டனர். முதலாவது, வளர்ந்து வரும் தமிழ் பேச்சு முறைகளை, குறிப்பாக இணையத்தில், நன்கு புரிந்துகொள்ளச் செய்வதற்கான ஆய்வு ஆகும். தமிழ் ஆவணங்களை தேசியமயமாக்கி டிஜிட்டல் மயமாக்கும் முயற்சியைத் தமிழ்நாடு அரசு தொடங்கியுள்ள நிலையில், மொழியின் வளங்களை ஆன்லைனில் மேம்படுத்தும் நோக்கில், இணையத் தமிழ் பேச்சைப் புரிந்துகொள்வதிலும், தமிழ் தகவல்களை நிர்வகிப்பதற்கான ஆராய்ச்சி அடிப்படையிலான தானியங்கி தொழில்நுட்பங்களை உருவாக்குவதிலும் அதிக முதலீடு செய்வது, தமிழ் பேச்சாளர்களின் இணைய ஒழுங்குபடுத்தல் அனுபவங்களை மேம்படுத்துவதற்கு மிக முக்கியமானது.

இறுதியாக, தானியங்கி அமைப்புகளைச் சோதிக்க அளவுகோல்கள் மற்றும் கட்டமைப்புகளை உருவாக்குவது, தேசியம், சாதி, பாலினம் மற்றும் வர்க்கம் ஆகிய அனைத்திலும் தானியங்கி ஒழுங்குபடுத்தல் தொழில்நுட்பங்கள் செயல்படுவதை உறுதி செய்வதற்கு அவசியமாகிறது. BHASA மற்றும் Pariksha போன்ற நம்பிக்கைக்குரிய முயற்சிகள் ஏற்கனவே உருவாகியுள்ளன. தமிழ் மொழியின் பன்முக புரிதல்களைப் பற்றி மாதிரிகள் (Models) தெரிவிக்கப்பட வேண்டும் என்பதால், பங்கேற்பு முறைகளைப் பயன்படுத்துவதன் மூலம் வரையறைகளை உருவாக்கி மாறுபட்ட விளக்கவுரையாளர்களின் கண்ணோட்டங்களை ஒருங்கிணைப்பதற்கான அதிக முயற்சிகளை ஆதரிப்பது மிகவும் முக்கியமாகும் (Arora et. al, 2024; Gordon et al., 2022).

பரிந்துரைகள்

1. கொள்கைகளை உருவாக்கி மதிப்பாய்வு செய்யும் போது, துறைசார் நிபுணர்களுடனான ஈடுபாட்டை அதிகரிப்பதன் மூலம், தேசியம், பாலினம், சாதி மற்றும் வர்க்கம் ஆகிய அனைத்து தரப்பிலும் காணப்படும் தமிழ் பேச்சின் பன்முகத்தன்மையை உள்ளடக்கிய தமிழ் தகவல் ஒழுங்குபடுத்தல் வழிகாட்டுதலை உறுதி செய்தல்.

தமிழ் பேசுபவர்கள் இணையத்தில் தொடர்புகொள்வதற்காகப் பயன்படுத்தும் பல்வேறு வழிகளை, குறிப்பாக நிறுவனங்கள் பதிவுகளை ஒழுங்குபடுத்துவதற்கான உலகளாவிய அணுகுமுறையைப் பின்பற்றும்போது, பெரும்பாலும் கணக்கில் எடுத்துக்கொள்ளத் தவறிவிட்டதாக நேர்காணலில் பங்கேற்றவர்கள் பலமுறை குறிப்பிட்டனர். பல்வேறு சமூகப் பொருளாதார நிலை, பாலினம், பாலியல் நோக்குநிலை மற்றும் சாதி ஆகியவற்றின் பரிசீலனைகளை ஒழுங்குபடுத்தும் முடிவுகள் பொருத்தமான முறையில் உள்ளடக்கவில்லை என்றும் பலர் நம்பினர். ஆங்கிலம் பேசத் தெரியாத காரணத்தினாலோ அல்லது மேற்கத்தியம் அல்லாத சூழல்களில் தீங்கு விளைவிக்கும் பேச்சுக்குத் தளங்கள் அதிக சகிப்புத்தன்மையைக் கொண்டிருப்பதன் காரணத்தினாலோ, தாம் மோசமான ஒழுங்குபடுத்தல் நடவடிக்கையை எதிர்கொண்டதாகப் பலர் உணர்ந்தனர். மற்ற ஆராய்ச்சியாளர்கள் அடையாளம் கண்டுள்ளபடி, தளங்கள், தமது ஒழுங்குபடுத்தல் செயல்முறைகள் வெவ்வேறு பயனர்களுக்குத் துன்புறுத்தல் மற்றும் தீங்கு நிகழக்கூடிய வழிகளுக்கு உணர்திறன் கொண்டவை என்பதை உறுதி செய்ய வேண்டும், மேலும் ஓரங்கட்டப்பட்ட பயனர்கள் தாங்கள் சந்திக்கும் துன்புறுத்தலை முறையீடு செய்வதற்கான வழிமுறைகளைக் கடினமாக்க கூடாது (Diaz, 2023). தனித்துவமான மொழியியல் நுணுக்கங்களைக் கருத்தில் கொள்ளாமல் தகவல் ஒழுங்குபடுத்தலுக்கு ஒரே மாதிரியான அணுகுமுறையைப்

பயன்படுத்துவதை நிறுவனங்கள் மறுபரிசீலனை செய்ய வேண்டும். ஏனென்றால் இந்த அணுகுமுறை தமிழ் பேசுபவர்கள் மற்றும் பிற பயனாளர்கள் தாங்கள் தேர்ந்தெடுத்த மொழியில் தங்களை வெளிப்படுத்திக் கொள்ளும் திறனைக் கடினமாக்குகிறது. மொழி மற்றும் பகுதி முழுவதும் தகவல்களை ஒழுங்குபடுத்துவதற்காக ஒரே அணுகுமுறையைப் பின்பற்றும்போது கூட, இந்தியா, இலங்கை மற்றும் புலம்பெயர் நாடுகளில் தமிழ் மொழி சூழல்களுக்குள் கொள்கைகள் விளக்கப்படுகின்ற அல்லது ஒழுங்குபடுத்தப்பட்டுள்ள வெவ்வேறு வழிகளைப் புரிந்துகொள்வதற்கு, தளங்கள் மொழி மற்றும் துறைசார் நிபுணர்களை ஈடுபடுத்த வேண்டும். அவ்வாறு செய்வது, விதிமீறுகின்ற துன்புறுத்தும் பேச்சுக்கள், ஓரங்கட்டப்பட்ட பயனர்களைக் குறிவைக்காது என்பதையும், ஓரங்கட்டப்பட்ட பயனர்களின் பேச்சு இணையத்தில் தவறான விதத்தில் ஒடுக்கப்படாது என்பதையும் தளங்கள் உறுதிசெய்ய உதவும்.

நிறுவனங்கள் ஏற்கனவே செய்துள்ள தமது எல்லைக்குட்பட்ட நம்பிக்கை மற்றும் பாதுகாப்பு முதலீடுகளைத் திரும்பப் பெறுவதால், இந்த முயற்சிகள் மிகவும் முக்கியமானவை (Goggin, 2024). நம்பிக்கை மற்றும் பாதுகாப்பு திறனில் இந்த இடைவெளியை நிரப்புவதற்கு நிறுவனங்கள் அதிக தானியங்கி தீர்வுகளைப் பயன்படுத்தத் தேர்வு செய்கின்றன. ஆனால் நாங்கள் நிரூபித்தபடி, தமிழ் தகவல் ஒழுங்குபடுத்தல் அமைப்புகள் தற்போது பல இடைவெளிகளையும் குறைபாடுகளையும் கொண்டுள்ளன. இணையத்தில் தீங்குகள் எவ்வாறு வெளிப்படுகின்றன மற்றும் தமிழ் எவ்வாறு பயன்படுத்தப்படுகிறது என்பதை நன்கு புரிந்துகொள்ளக்கூடிய மொழி கணினி நிபுணர்கள் மற்றும் துறைசார் நிபுணர்கள் போன்ற தமிழ் நிபுணர்களுடன் சேர்ந்து செயல்படுவது, நிறுவனங்கள் தானியங்கி தொழில்நுட்பங்கள் மற்றும் தகவல் ஒழுங்குபடுத்தல் செயல்முறைகளை ஒட்டுமொத்தமாக மேம்படுத்த உதவும். உதாரணமாக, இந்திய, குறிப்பாகத் தமிழ்ச் சூழல்களில் பாலின அடிப்படையிலான தீங்குகள் வெளிப்படும் விதம், பிற இந்திய மொழிகள் அல்லது ஆசியா அல்லாத சூழல்களிலிருந்து பெரிதும் வேறுபடுகிறது (Aneja et al., 2024).

2. துன்புறுத்தல்களைத் தடுப்பதற்கான கொள்கைகளை அமல்படுத்துவதை மேம்படுத்துதல் மற்றும் தமிழில் முறையீடு செய்வதற்கான வழிகளில் அதிக முதலீடு செய்தல்.

நேர்காணல்களில், தமிழில் துன்புறுத்தல் அல்லது வெறுப்புப் பேச்சை **முறையீடு** செய்வதற்கான கருவிகள் போதுமானதாக இல்லை, பற்றாக்குறையாக உள்ளன அல்லது தமிழ் பேசுபவர்களுக்கு அவற்றைப் பயன்படுத்துவது கடினமாக உள்ளது என்று கண்டறியப்பட்டது. தமிழ்ச் சூழலை நன்கு அறிந்த குழுக்களால் நடத்தப்பட்ட முந்தைய ஆராய்ச்சி, குறிப்பாக தெற்காசியச் சூழல் மற்றும் விளிம்புநிலை சமூகங்களுக்குள் ஏற்படும் நுணுக்கமான துன்புறுத்தல்களை மற்றும் புகாரளிப்பதில் உள்ள முரண்பாடுகள் மற்றும் குறைபாடுகளைக் கண்டறிந்தது (Sehgal & Nambiar, 2024). முதலாவதாக, தளத்தின் கொள்கைகள் எப்போதும் தமிழ் உள்ளிட்ட பிற மொழிகளில் துல்லியமாக மொழிபெயர்க்கப்படுவதில்லை. இதனால் பயனர்கள் எவ்வாறான செய்திகளைப் புகாரளிக்க முடியும், மற்றும் அவற்றை எப்படிப் புகாரளிக்க முடியும் என்பதை அறிந்து கொள்வது கடினமாகிறது(CITE). மேலும், பயனர்கள் பல பதிவுகளைப் புகாரளிப்பதில் கட்டுப்பாடுகள் காணப்படுகின்றன. இருப்பினும் சில நேரங்களில் தொடர்ச்சியாக இலக்கு வைக்கப்பட்ட பதில்களைக் காண்பிப்பது துன்புறுத்தலை நிரூபிப்பதற்கு அவசியமாகிறது. பயனர்கள் தளத்தின் பட்டியலில் காணப்படும் துன்புறுத்தல் வகைகளுடன் தொடர்புபட்ட பதிவுகளை மட்டுமே புகாரளிக்க முடியும். அவை பெரும்பாலும் உலகளாவிய தெற்கு நாடுகளில் நடைபெறும் பல வகையான துன்புறுத்தல்களை போதுமான அளவு விவரிக்கவில்லை (Sehgal & Nambiar, 2024). மேலும், பயனர்கள் புகாரளிப்பதில் வெற்றி பெற்றாலும், அவர்களுக்கு பெரும்பாலும் தளங்களில் இருந்து எந்த பதிலும் கிடைப்பதில்லை.

பயனர்களின் அனைத்து முறையீடுகளுக்கும் பயனர் தேர்ந்தெடுத்த மொழியில் பதிலளிக்கப்படுவதைத் தளங்கள் உறுதி செய்ய வேண்டும். மேலும், தளம் செயல்படும் அளவைப் பிரதிபலிக்கும் வகையில் சரியான நேரத்தில் பதிலளிக்க வேண்டும் மற்றும் பயனர்களின் மன ஆரோக்கியம் மற்றும் நல்வாழ்வில் மீண்டும் மீண்டும் துன்புறுத்தலுக்கு உள்ளாவதால் ஏற்படும் தாக்கத்தை ஒப்புக்கொள்ள வேண்டும் (Thakur & Hankerson-Madrigal, 2022). ஒரு பதிவு ஏன் அகற்றப்பட்டது அல்லது அகற்றப்படவில்லை என்பது குறித்த சிறந்த விளக்கங்களை வழங்கப் பெரிய மொழி மாதிரிகளை (LLMs) பயன்படுத்துவதற்கான தற்போதைய திட்டங்களைக் கருத்தில் கொண்டு, புகாரளிக்கும் செயல்முறையின் அனைத்து படிமுறைகளையும் அனைத்து பயனாளர்களும், அவர்கள் எந்த மொழியைப் பேசினாலும், அணுக முடியும் என்பதனை

உறுதிசெய்வதற்கு அவை தாய்மொழி பேசுபவர்களால் சோதிக்கப்படுவதை நிறுவனங்கள் உறுதி செய்ய வேண்டும் (டிஜிட்டல் நம்பிக்கை மற்றும் பாதுகாப்பு கூட்டாண்மை, 2024).

3. தமிழ் மொழிபெயர்ப்பாளர்களை பணியமர்த்துவதன் மூலமோ அல்லது அவர்களின் சேவைகளை ஒப்பந்தம் செய்வதன் மூலமோ, போதுமான மொழித் திறனுடன் இயந்திர மொழிபெயர்ப்பு கருவிகளின் பயன்பாட்டை மேம்படுத்துதல்.

தமிழ் தகவல்களை ஒழுங்குபடுத்துவதற்கு முன், அவற்றைத் தமிழிலிருந்து ஆங்கிலத்திற்கு மொழிபெயர்ப்பதற்காக, மதிப்பீட்டாளர்களும் மதிப்பீட்டுக் குழுக்களும் பெரும்பாலும் இயந்திர மொழிபெயர்ப்பு தொழில்நுட்பங்களை நம்பியுள்ளனர். இருப்பினும், ஆராய்ச்சியாளர்கள் மற்றும் பத்திரிகையாளர்கள், இந்த தானியங்கி மொழிபெயர்ப்பு கருவிகள் பெரும்பாலும் பிழைகளுக்கு ஆளாகின்றன. குறிப்பாகத் தமிழ் போன்ற குறைந்த வள மொழிகளில், அவை பெரும்பாலும் சொற்களைப் புறக்கணிக்கின்றன. முற்றிலும் புதியவற்றை உருவாக்குகின்றன. மேலும், இந்த மாதிரிகளைப் பயிற்றுவித்துச் சோதிப்பதற்குத் தமிழ் தரவு பற்றாக்குறை காணப்படுவதன் காரணமாகத் தமிழ் பதிவுகளை மொழிபெயர்க்கும்போது ஒரு சொல்லை முற்றிலும் தவறாகப் புரிந்துகொள்கின்றன என்பதைக் கண்டறிந்துள்ளனர் (Choudhary et al., 2018; Ramesh et al., 2020; Deck, 2023). மொழிபெயர்ப்பு கருவிகளைப் பயன்படுத்த விரும்பும் மதிப்பீட்டுக் குழுக்கள், தமது நிறுவனத்தின் மொழிபெயர்ப்பு கருவிகளை மேம்படுத்துவதற்கு வாணி (Vaani) போன்ற தாய்மொழிப் பேசுபவர்களால் உருவாக்கப்பட்ட மொழித் தொகுப்புகளைப் பயன்படுத்துவதைக் கருத்தில் கொள்ள வேண்டும்.ஒரு பதிவை ஆராய்வதற்குத் தேவையான அத்தியாவசிய சூழலை இழக்காமல் இருப்பதற்கு, குறிப்பாக எல்லைக்கோட்டின் அருகில் உள்ள பதிவு மற்றும் முறையீடுகள் என்று வரும்போது, தாய்மொழி பேசுபவர்களால் பதிவு மதிப்பாய்வு செய்யப்படுவதையும் அவர்கள் உறுதி செய்ய வேண்டும்.

4. தமிழ் இயற்கை மொழி செயலாக்க ஆராய்ச்சியாளர்களுடன் இணைந்து பின்வரும் இடங்களில் செயல்பட வேண்டும். தமிழ் ஒழுங்குபடுத்தல் கருவிகளை

மேம்படுத்துவதற்காக, வெறுப்புப் பேச்சு மற்றும் சாதி அடிப்படையிலான பேச்சைக் கண்டறிவாதற்காகக, தமிழ் பேச்சை நன்கு புரிந்துகொள்ளும் தாய்மொழி மூல அளவுகோல்களைப் உருவாக்குவாதற்காகத்

தானியங்கி தகவல் உள்ளடக்க ஒழுங்குபடுத்தல் மற்றும் மொழிபெயர்ப்பு தொழில்நுட்பங்கள் பேச்சுக்களை ஒழுங்குபடுத்துவதற்காக அதிகளவில் பயன்படுத்தப்படுகின்றன, குறிப்பாக உலகளாவிய அளவில் சேவைகளை வழங்கும் தளங்களால் இவை பயன்படுத்தப்படுகின்றன. இருப்பினும், இந்த அமைப்புகள் நன்றாகச் செயல்படுகின்றனவா என்பதைச் சோதிப்பதற்கு உயர்தர பயிற்சி தரவுத்தொகுப்புகள், வரையறைகள் மற்றும் மதிப்பீட்டு அளவீடுகள் போன்ற வளங்களின் பற்றாக்குறை காரணமாக அவை பல வழிகளில் தோல்வியடைகின்றன (Nicholas & Bhatia, 2023). நம்பிக்கை & பாதுகாப்புக் குழுக்கள் பெரும்பாலும் வள பற்றாக்குறையுடன் உள்ளன. மேலும், அவர்களால் இந்த அமைப்புகளைப் பயிற்றுவிப்பதற்கும் சோதிப்பதற்கும் அத்தகைய வளங்களை உருவாக்க முடியாமல் போகலாம். இருப்பினும், பல ஆராய்ச்சியாளர்கள் பொதுவில் அணுகக்கூடிய அகராதிகளையும் கலாச்சார ரீதியாகக் குறிப்பிட்ட அளவுகோல்களையும் உருவாக்கியுள்ளனர். உதாரணமாக, Tattle's Uli (டாட்டில்ஸ் உலி) என்பது தமிழ், இந்தி மற்றும் இந்திய ஆங்கில ஆகிய மொழிகளில் பாலினம், சாதி மற்றும் பிற ஓரங்கட்டப்பட்ட அடையாளங்களுக்கு எதிரான அவதூறுகளின் திறந்த ஒருங்கிணைந்த முறையில் உருவாக்கப்பட்ட சொல்லகராதி ஆகும். இந்த வளங்களை உருவாக்குகின்ற, காணப்படும் தரவுகளை அடையாளம் கண்டு, மதிப்பீடு செய்து அவற்றைத் தமிழ் ஒழுங்குபடுத்தும் செயல்முறைக்காகப் பயன்படுத்துகின்ற நிபுணர்களுடன் நம்பிக்கை மற்றும் பாதுகாப்புக் குழுக்கள் ஈடுபடுவதற்கான வழிகளைத் தளங்கள் உருவாக்க முடியும் மற்றும் உருவாக்க வேண்டும்.

தகவல் மதிப்பீட்டு அமைப்புகளின் செயல்திறனை மதிப்பிடுவதற்கு, தற்போதுள்ள இயந்திர-மொழிபெயர்க்கப்பட்ட தரவுத்தொகுப்புகள் மற்றும் அளவுகோல்கள் பெரும்பாலும் முழுமைபெறாத மாற்றீடுகளாக இருப்பதால், உள்நாட்டில் உருவாக்கப்பட்ட சொல் அகராதிகளையும் அளவுகோல்களையும் பயன்படுத்துவது அவசியமாகும். நிறுவனங்கள் பல்வேறு துறைகளில் 16,000 கேள்விகளைக் கொண்ட Massive Multitask Language Understanding (MMLU) போன்ற வளங்களைப் பரவலாக நம்பியிருந்தாலும், மேற்கத்திய மொழி அல்லாத பேச்சின் தகவல் மற்றும் சூழலை அலசுவதற்குப் பயன்படுத்தப்படும்

அமைப்புகளின் செயல்திறனை அளவிட அவை பெரும்பாலும் போதுமானதாக இல்லை என்று ஆராய்ச்சியாளர்கள் கூறுகின்றனர். உதாரணமாக, அமெரிக்கச் சட்டம் குறித்த கேள்விக்குப் பதிலளிப்பதில் ஒரு மாடலின் செயல்திறனை MMLU அளவுகோல் வெற்றிகரமாக மதிப்பிடக்கூடும். ஆனால், தமிழ்நாட்டின் தலைநகரம் குறித்த கேள்விக்கு (தமிழ் அல்லது ஆங்கிலத்தில்) ஒரு மாடலால் பதிலளிக்க முடியுமா என்பதை அது மதிப்பிடாமல் போகலாம்.

நம்பிக்கை & பாதுகாப்புக் குழுக்கள் பயன்படுத்தக்கூடிய தனித்துவம் வாய்ந்த இந்திய மொழிகளுக்கான அளவுகோலின் ஒரு எடுத்துக்காட்டு IndicMMLU-Pro ஆகும். இது தமிழ் மற்றும் பிற இந்திய மொழிகளில் குறிப்பிட்ட பகுதிசார்ந்த தளங்களில் தகவல் ஒழுங்குபடுத்தல் கருவிகள் மற்றும் மொழிபெயர்ப்பு கருவிகள் உள்ளிட்ட மாதிரிகளை மதிப்பிடுவதற்கான ஒரு புதிய தரநிலையை நிறுவுகிறது (KJ et al., 2025). தென்கிழக்கு ஆசியச் சூழல்களுக்கான அளவுகோலான BHASA உள்ளது. இதில் தமிழும் பரவலாகப் பேசப்படும் மொழியாக இருக்கிறது, மற்றும் மைக்ரோசாஃப்ட் ஆராய்ச்சி திட்டத்தின் ஒரு பகுதியாக உருவாக்கப்பட்ட பல இந்திய மொழிகளில் மனித மற்றும் இயந்திரத்தால் மொழிபெயர்க்கப்பட்ட அளவுகோல்களின் கலவையான 'பரிக்ஷா' ஆகியவையும் உள்ளடக்கம் (Leong et al., 2023; Watts et al., 2024).

இறுதியாக, தமிழ் மொழி ஒழுங்குபடுத்தல் கருவிகளைச் சோதிப்பதற்காக மொழித்தொகுப்பு அமைப்புகளை உருவாக்கும்போது அல்லது பயன்படுத்தும் போது பங்கேற்பு முறைகளைப் பயன்படுத்துவதற்கு இயற்கை மொழி செயலாக்க ஆராய்ச்சியாளர்களுடன் இணைந்து செயல்படுவதை நம்பிக்கை மற்றும் பாதுகாப்புக் குழுக்கள் பரிசீலிக்க வேண்டும். தானியங்கி அமைப்புகளைப் பயிற்றுவித்தல் மற்றும் சோதிப்பதில் உள்ள சவாலைப் பற்றி நம்பிக்கை மற்றும் பாதுகாப்புக் குழுக்கள், குறிப்பாகத் தரவு விளக்கவுரையாளர்கள் அல்லது மதிப்பீட்டாளர்கள் ஒரு வார்த்தையின் அர்த்தம் என்ன அல்லது ஒரு பதிவின் ஒரு பகுதி புண்படுத்தும் தன்மை உள்ளதா இல்லையா என்பது குறித்து ஒருவருக்கொருவர் கருத்து வேறுபாடு கொள்ளும்போது, புலம்பித் தள்ளுகின்றன. இருப்பினும், இந்த எதிர்ப்புக் குரல்களையும் கருத்து வேறுபாடுகளையும் தரவுத்தொகுப்புகளாக ஒருங்கிணைப்பதற்கான வழிகள் இருக்கலாம் என்று ஆராய்ச்சியாளர்கள் கூறுகின்றனர், பின்னர் அவை சூழல்சார்ந்த மற்றும் சிக்கலான பேச்சை ஆராய்வதற்காக மாதிரிகளைப் பயிற்றுவிப்பதற்குப் பயன்படுத்தப்படலாம் (Gordon et al., 2022; Arora et al., 2023). இந்த அணுகுமுறை தரவுத்தொகுப்புகள் பயனர்கள் தமிழைப் பேசும் மற்றும் ஒலிபெயர்க்கும் எண்ணற்ற வழிகளை உள்ளடக்குகின்றன என்பதை உறுதிப்படுத்தவும் உதவும்.

5. தகவல் எப்போது ஒழுங்குபடுத்தப்பட்டது மற்றும் ஏன் என்பது குறித்து பயனர்களுக்குத் தெரிவிக்கவும். இதில் குறைவாக பரப்பப்படுதல் மற்றும் பதிவுகளை நீக்குவதற்கான அரசாங்கத்தின் கோரிக்கைகள், பயனர் தரவு மற்றும் ஒழுங்குபடுத்தல் மாற்றங்கள் போன்றவையும் உள்ளடங்கும்.

தளங்கள் பதிவுகளை ஒழுங்குபடுத்தலாம் மற்றும் பல காரணங்களுக்காகப் பதிவுகள் மற்றவர்களைச் சென்றடைவதைக் குறைக்கலாம். இருப்பினும், இது எப்போது நடந்தது, ஏன் நடந்தது என்பதை, குறிப்பாக ஜனநாயக ரீதியாகத் தேர்ந்தெடுக்கப்பட்ட அதிகாரிகளுடன் ஒருங்கிணைந்து, அவர்களின் உத்தரவின் பேரில் முடிவுகள் எடுக்கப்படும்போது, பயனர்கள் தெரிந்து கொள்ள வேண்டும். தளங்கள் தகவல்களைத் தெளிவாக வெளியிடும்போது பயனர்கள் அதிலிருந்து நிறையக் கற்றுக்கொள்ளலாம். அவற்றில் சில: தகவல் மதிப்பீட்டை வழிநடத்துவதில் தங்கள் மாநிலம் எவ்வளவு ஈடுபட்டுள்ளது என்பதைக் கற்றுக்கொள்வது; ஒரு நிறுவனம் அரசாங்க அமைப்புகளுடன் எவ்வளவு காலத்திற்கு ஒருமுறை அல்லது எந்த வழியில் தொடர்பு கொள்கிறது; எதிர்காலத்தில் தள தகவல் மதிப்பீட்டு விதிகளை எவ்வாறு சிறப்பாகக் கடைப்பிடிப்பது; ஒரு மதிப்பீட்டு முடிவு அவர்களின் பதிவுகளைத் தவறாகப் புரிந்துகொண்டதாக அவர்கள் உணரும்போது தளங்கள் மற்றும் மேற்பார்வை வாரியங்களுக்கு மிகவும் குறிப்பிட்ட முறையீடுகளைச் செய்வது; மற்றும் அவர்களின் பேச்சு அவர்களின் அரசாங்கத்தால் எப்போது குறைக்கப்பட்டது என்பதை அறிந்துகொள்வது மற்றும் அவர்களின் தலைவர்களைப் பொறுப்பேற்க வைப்பது. ஒரு பதிவு எப்போது மற்றும் ஏன் ஒழுங்குபடுத்தப்பட்டது என்பது குறித்து நிறுவனங்கள் பயனர்களுக்குத் தெரிவிப்பது, தொழில்நுட்ப நிறுவனங்கள் பங்கேற்கும் வணிகம் மற்றும் மனித உரிமைகள் தொடர்பான ஐக்கிய நாடுகளின் வழிகாட்டுதல் கொள்கைகளைக் கடைப்பிடிப்பதிலும், பயனர்களின் மனித உரிமைகளை இணையத்தில் பரவலாக மேம்படுத்துவதிலும் ஒரு முக்கிய பகுதியாகும். UN Guiding Principles on Business & Human Rights (2011).

தானியங்கி வழிமுறைகளைப் பயன்படுத்துதல், அரசாங்கத்தின் நீக்குதல் கோரிக்கையின் மூலமாகப் பதிவை நீக்குதல் அல்லது பயனர் கோரிக்கைக்குப் பதிலளிக்கும் விதமாகப் பதிவை நீக்குதல் போல, ஒரு பதிவு எவ்வாறு ஒழுங்குபடுத்தப்பட்டது என்பதை வெளிப்படுத்துவதன் மூலம், தளங்கள் பயனர்களுக்கு இந்த முயற்சியில் உதவலாம். பயனர்கள் கொள்கையைப் பற்றி மேலும் தெரிந்துகொள்வதற்காக, தளங்கள் அவர்களுடைய பதிவு எந்த குறிப்பிட்ட கொள்கையை மீறியது என்பதைப் பயனர்களுக்கு வெளிப்படுத்தலாம். இறுதியாக, வெளிப்படுத்தல் என்பது பயனர்கள் தளத்தின் முடிவை மேல்முறையீடு செய்வதற்கான தெளிவான வழியைக் கொண்டிருக்க வேண்டும்.

அரசாங்க கோரிக்கைகளின் குறிப்பிட்ட விஷயத்தில், அரசாங்க கோரிக்கையின் தன்மை மற்றும் உரையைக் கிடைக்கச் செய்வதன் மூலம் பயனர்கள் தங்கள் உரிமைகளைச் சிறப்பாகப் பாதுகாக்கத் தளங்கள் உதவ முடியும். கடந்த காலங்களில், நிறுவனங்கள் இந்த கோரிக்கைகளை வெளிப்படைத்தன்மை அறிக்கைகள் மற்றும் பொது களஞ்சியங்கள் மூலம் ஒட்டுமொத்தமாக வெளியிட்டன. இதன் மூலம், மேற்பார்வை, குடிமக்களின் பங்கேற்பு மற்றும் அரசாங்கங்கள் நிறுவனங்களுடன் எவ்வாறு செயல்படுகின்றன என்பது குறித்த ஆராய்ச்சி ஆகியவற்றைச் செயல்படுத்த முடியும். இதற்கு ஒரு எடுத்துக்காட்டாக, ஹார்வர்ட் பல்கலைக்கழகத்தின் பெர்க்மேன் க்ளீன் மையத்தில் அமைந்துள்ள ஒரு சுயாதீன திட்டமான Lumen database இனைக் குறிப்பிடலாம். இங்குத் தளங்கள் தானாக முன்வந்து பதிவு நீக்கங்கள் அல்லது பயனர் தரவுகளுக்கான குறிப்பிட்ட அரசாங்க கோரிக்கையைச் சமர்ப்பிக்கலாம். தளங்களால் பெறப்படும் அரசாங்க கோரிக்கைகளை லுமேன் தரவுத்தளம் அல்லது இது போன்ற மன்றங்களில் பகிரங்கப்படுத்துவது, நம்பிக்கை மற்றும் வெளிப்படைத்தன்மையை வளர்ப்பதற்கான நம்பிக்கை மற்றும் பாதுகாப்பு முயற்சிகளின் முக்கிய பகுதியாக இருக்க வேண்டும்.

பின் இணைப்பு

முறைகள்

தமிழில் தகவல் ஒழுங்குபடுத்தலின் நிலையைக் கற்பதற்காக, பண்பியல் மற்றும் அளவையியல் முறைகளை இணைத்து ஒரு கலப்பு-முறை அணுகுமுறையைப் பயன்படுத்தினோம். முதலில், தகவல் சூழலை மதிப்பிடுவதற்கும், தகவல் ஒழுங்குபடுத்தலின் நிலை மற்றும் கருத்தைப் பற்றி மேலும் அறிந்துகொள்வதற்கும், தகவல் ஒழுங்குபடுத்தல் எவ்வாறு செயல்படுகிறது மற்றும் செயல்படாது என்பதை நேரடியாகப் புரிந்துகொள்வதற்கும் 17 நிபுணர்களுடன் பகுதியாக-கட்டமைக்கப்பட்ட நேர்காணல்களை நடத்தினோம். மேற்கத்திய மற்றும் இந்திய நிறுவனங்களின் நம்பிக்கை மற்றும் பாதுகாப்பு மற்றும் கொள்கை வல்லுநர்கள், தெற்காசியா அல்லது ஆசிய பசிபிக் பிராந்தியத்தின் உரைகளை மதிப்பாய்வு செய்யும் பணியில் உள்ள மதிப்பீட்டாளர்கள், டிஜிட்டல் உரிமை நிபுணர்கள் மற்றும் கல்வியாளர்கள் மற்றும் தமிழ் மொழி தகவல் படைப்பாளர்களுடன் நாங்கள் பேசினோம்.

இந்தியாவின் பெங்களூரில் களப் பயணத்தின் போது நேரில் நடத்தப்பட்ட ஒன்றைத் தவிர பெரும்பாலான நேர்காணல்கள் இணையத்தில் நடத்தப்பட்டன. நேர்காணல்கள் அக்டோபர் 2024 முதல் ஜனவரி 2025 வரை நடத்தப்பட்டன. நேர்காணல்கள் அனைத்தும் ஆங்கிலத்தில் நடத்தப்பட்டன, சில விவாதங்கள் இந்தியில் நடத்தப்பட்டன, மேலும் சில எடுத்துக்காட்டுகள் தமிழில் கொடுக்கப்பட்டன, அவற்றை நேர்காணல் பங்கேற்பாளர் ஆராய்ச்சி குழுவிற்காக ஆங்கிலத்தில் மொழிபெயர்த்தார். நேர்காணல்களின் போது களக் குறிப்புகள் எடுக்கப்பட்டன, பின்னர் பொதுவான கருப்பொருள்களைக் கண்டறிவதற்காக நேர்காணல்களின் பதிவுகள் படியெடுக்கப்பட்டு பகுப்பாய்வு செய்யப்பட்டன.

நேர்காணல்களைத் தாண்டி, சமூக ஊடகங்களை அடிக்கடி பயன்படுத்தும் தமிழ் பேசும் பயனர்களின் அனுபவத்தைப் புரிந்துகொள்ள ஒரு இணையக் கணக்கெடுப்பை நடத்தினோம். இந்த கணக்கெடுப்பு பெங்களூரில் உள்ள எங்கள் கூட்டாளரான இணையம் மற்றும் சமூக மையத்தாலும் முன்னர் தமிழ் இணைய சமூகங்களில் ஒரு பகுதியாக இருந்த நேர்காணல் செய்பவர்களாலும் விநியோகிக்கப்பட்டது. இந்த கணக்கெடுப்பு Reddit தளத்தில் ஒரு இணையத் தமிழ் விவாதக் குழுவிலும் பகிரப்பட்டது. சமூக ஊடக தளங்களின் மீதான பயனர்களின் நம்பிக்கை, ஒழுங்குபடுத்துதலை எதிர்கொண்ட அனுபவங்கள் மற்றும் ஒழுங்குபடுத்தல் குறித்த அவர்களின் கருத்துக்கள் குறித்து இந்த கணக்கெடுப்பில் கேள்விகள் எழுப்பப்பட்டன. 2024ம் ஆண்டில் நவம்பர் 7 முதல் டிசம்பர் 15 வரை இணையக் கணக்கெடுப்பை விநியோகிப்பதற்கு அல்கெமர் தளம் பயன்படுத்தப்பட்டது. பங்கேற்றதற்காக ஒரு சிறிய தொகையாக 10$வழங்கப்பட்டது. இந்த கணக்கெடுப்பு ஆங்கிலம் மற்றும் தமிழில் கிடைத்தது. இணையச் சேவைகளை அடிக்கடி பயன்படுத்துபவர்கள் மீது நாங்கள் கவனம் செலுத்தினோம். எங்கள் கணக்கெடுப்பில் பங்கேற்றவர்களில் சுமார் 90% பேர் தினசரி சமூக ஊடகங்களைப் பயன்படுத்துவதாகக் குறிப்பிட்டனர். எங்கள் மாதிரியில் பெரும்பாலானவர்கள் இந்தியாவைச் சேர்ந்தவர்கள் (138 பேர்), இலங்கையைச் சேர்ந்தவர்கள் ஒரு சிலர் மட்டுமே. இந்தக் குறைபாட்டை

நிவர்த்தி செய்வதற்காக, இலங்கைத் தமிழ் பயனர்கள் மற்றும் நிபுணர்களுடன் நேர்காணல்களை நடத்தி, எங்கள் வட்டமேசைக் கூட்டத்தில் பங்கேற்க அவர்களை அழைத்தோம்.

இறுதியாக, ஜனவரி 8, 2025 அன்று, தமிழில் இயந்திர கற்றல் தொழில்நுட்பங்களுக்கான தரவுத்தொகுப்புகளை உருவாக்குதல், மதிப்பீடு செய்தல் மற்றும் குறிப்பு எழுதுதல் ஆகியவற்றை நன்கு அறிந்த 10 கணினி அறிவியல் நிபுணர்களுடன் ஒரு வட்டமேசைக் கலந்துரையாடலை நடத்தினோம். வட்ட மேசைக் கலந்துரையாடல் இணையத்தில் நடைபெற்றது, இதில் கனடா, இந்தியா, ஐக்கிய இராச்சியம் மற்றும் அமெரிக்காவைச் சேர்ந்த நிபுணர்கள் கலந்து கொண்டனர். மேலும் இதில், அந்த நிபுணர்கள் தரவு சேகரிப்பு, குறிப்பு எழுதுதல், மதிப்பீடு போன்றவற்றில் அவர்கள் சந்தித்த சவால்கள் பற்றியும் அந்த சவால்களை எதிர்கொள்ள அவர்கள் கண்டுபிடித்த தீர்வுகள் பற்றியும் கண்டறியப்பட்டது.

சான்றாதாரங்கள்

Access Now. (2022). *Meta must disclose India's Human Rights Impact Assessment.* Access Now. https://www.accessnow.org/press-release/meta-india-human-rights-impact-assessment/ [perma.cc/7X2M-8YZT]

ACL Rolling Review Dashboard,. (n.d.). *ARR Dashboard.* Languages Mentioned in Paper Abstracts. Retrieved February 24, 2025, from https://stats.aclrollingreview.org/submissions/linguistic-diversity/ [perma.cc/U2FM-3UEN]

Amarasingam, A., & Nandakumar, T. (2021, May 14). Social Media Platforms are Silencing Social Movements. *Tech Policy Press*. https://techpolicy.press/social-media-platforms-are-silencing-social-movements [perma.cc/7BQ-3JE4L]

Anandakugan, N. (2020, August 31). *The Sri Lankan Civil War and Its History, Revisited in 2020.* Harvard International Review. https://hir.harvard.edu/sri-lankan-civil-war/ [perma.cc/7V6J9-SGB]

Aneja, U., Gupta, A., Jain, A., & John, S. (2024). *From Code to Consequence: Interrogating Gender Biases in LLMs within the Indian Context.* Digital futures lab. https://www.notion.so [perma.cc/6VA-2H73U]

Arora, A., Jinadoss, M., Arora, C., George, D., Brindaalakshmi, Khan, H. D., Rawat, K., Div, Ritash, Mathur, S., Yadav, S., Shora, S. R., Raut, R., Pawar, S., Paithane, A., Sonia, Vivek, Priscilla, D., Khairunnisha, … Prabhakar, T. (2024). *The Uli Dataset: An Exercise in Experience Led Annotation of oGBV (arXiv:2311.09086). arXiv.* https://doi.org/10.48550/arXiv.2311.09086 [perma.cc/H6CV-UEG6]

Biddle, S. (2022, January 19). *Facebook's Tamil Censorship Highlights Risks to Everyone.* The Intercept. https://theintercept.com/2022/01/19/facebook-tamil-censorship-sri-lanka/ [perma.cc/Q2TX-KYLK]

Brustein. (2020, May 12). Facebook Apologizes for Role in Sri Lankan Violence. *Bloomberg*. https://www.bloomberg.com/news/articles/12-05-2020/facebook-apologizes-for-role-in-sri-lankan-violence [perma.cc/3VUR-KCLY]

Chakravarthi, B. R., Muralidaran, V., Priyadharshini, R., & McCrae, J. P. (2020). *Corpus Creation for Sentiment Analysis in Code-Mixed Tamil-English Text* (arXiv:2006.00206). arXiv. https://doi.org/10.48550/arXiv.2006.00206 [perma.cc/7NXF-YW5J]

Choudhary, H., Pathak, A. K., Saha, R. R., & Kumaraguru, P. (2018). Neural Machine Translation for English-Tamil. In O. Bojar, R. Chatterjee, C. Federmann, M. Fishel, Y. Graham, B. Haddow, M. Huck, A. J. Yepes, P. Koehn, C. Monz, M. Negri, A. Névéol, M. Neves, M. Post, L. Specia, M. Turchi, & K. Verspoor (Eds.), *Proceedings of the Third Conference on Machine Translation: Shared Task Papers* (pp. 770–775). Association for Computational Linguistics. https://doi.org/10.18653/v1/W18-6459 [perma.cc/N24Q6-TAS]

Constine, J. (2018). Facebook reveals 25 pages of takedown rules for hate speech and more | TechCrunch. *Tech Crunch*. https://techcrunch.com/2018/04/24/facebook-content-rules/ [perma.cc/4JDD-UFL9]

Cronin-Furman, K., & Arulthas, M. (2024). How the Tigers Got Their Stripes: A Case Study of the LTTE's Rise to Power. *Studies in Conflict & Terrorism,* 47(9), 1006–1025. https://doi.org/10.1080/1057610X.2021.2013753 [perma.cc/3AFF-VFJR]

Daniel, S. (2016, February 18). *Sri Lankan Tamils around the world have built an online library to replace one torched in 1981*. Scroll.In. http://scroll.in/article/802923/sri-lankan-tamils-around-the-world-have-built-an-online-library-to-replace-one-torched-in-1981 [perma.cc/DHR-9GV24]

Deck, A. (2023, September 6). We tested ChatGPT in Bengali, Kurdish, and Tamil. It failed. *Rest of World*. https://restofworld.org/2023/chatgpt-problems-global-language-testing/ [perma.cc/GFD-3YBJR]

Diaz, A. (2023). *Online Racialization and the Myth of Colorblind Content Policy* (SSRN Scholarly Paper 4509790). Social Science Research Network. https://doi.org/10.2139/ssrn.4509790 [perma.cc/5GS-2KNN8]

Digital Tamil Studies. (n.d.). *Tamil Nationalized and Public Domain Books Collection*. Retrieved April 5, 2025, from https://tamil.digital.utsc.utoronto.ca/61220/utsc35335 [perma.cc/X59S-YCVC]

Digital Trust & Safety Partnership. (2024). https://dtspartnership.org/wp-content/uploads/2024/09/DTSP_Best-Practices-for-AI-Automation-in-Trust-Safety.pdf [perma.cc/3JU-2PW9B]

Duarte, N., & Llansó, E. (2017). *Mixed Messages? The Limits of Automated Social Media Content Analysis.* https://cdt.org/insights/mixed-messages-the-limits-of-automated-social-media-content-analysis/ [perma.cc/Z5W-3B6ZQ]

Elswah, M. (2024a, September 27). Moderating Maghrebi Arabic Content on Social Media. *Center for Democracy and Technology*. https://cdt.org/insights/moderating-maghrebi-arabic-content-on-social-media/ [perma.cc/XL8Y-BJ8H]

Elswah, M. (2024b, December 12). Moderating Kiswahili Content on Social Media. *Center for Democracy and Technology*. https://cdt.org/insights/moderating-kiswahili-content-on-social-media/ [perma.cc/J8CM-MMDE]

Ethirajan, A. (2021, March 23). *UN to collect evidence of alleged Sri Lanka war crimes.* https://www.bbc.com/news/world-asia-56502221 [perma.cc/U9AY-AE8L]

Facebook Sri Lanka Human Rights Impact Assessment. (2020). *Facebook Response: Sri Lanka Human Rights Impact Assessment*. Facebook. https://about.fb.com/wp-content/uploads/2021/03/FB-Response-Sri-Lanka-HRIA.pdf [perma.cc/CG-63JW4Q]

Gala, J., Chitale, P. A., AK, R., Gumma, V., Doddapaneni, S., Kumar, A., Nawale, J., Sujatha, A., Puduppully, R., Raghavan, V., Kumar, P., Khapra, M. M., Dabre, R., & Kunchukuttan, A. (2023). *IndicTrans2: Towards High-Quality and Accessible Machine Translation Models for all 22 Scheduled Indian Languages* (arXiv:2305.16307). arXiv. https://doi.org/10.48550/arXiv.2305.16307 [perma.cc/CG-63JW4Q]

Goggin, B. (2024, March 29). *Big Tech companies reveal trust and safety cuts in disclosures to Senate Judiciary Committee.* NBC News. https://www.nbcnews.com/tech/tech-news/big-tech-companies-reveal-trust-safety-cuts-disclosures-senate-judicia-rcna145435 [perma.cc/R92J-P5BD]

Gordon, M. L., Lam, M. S., Park, J. S., Patel, K., Hancock, J. T., Hashimoto, T., & Bernstein, M. S. (2022). Jury Learning: Integrating Dissenting Voices into Machine Learning Models. *CHI Conference on Human Factors in Computing Systems,* 1–19. https://doi.org/10.1145/3491102.3502004 [perma.cc/FX9T-FZ53]

Horwitz, J., & Purnell, N. (2020, August 30). Facebook Executive Supported India's Modi, Disparaged Opposition in Internal Messages. *Wall Street Journal.* https://www.wsj.com/articles/facebook-executive-supported-indias-modi-disparaged-opposition-in-internal-messages-11598809348 [perma.cc/S4ZK-TCQX]

Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The State and Fate of Linguistic Diversity and Inclusion in the NLP World. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 6282–6293). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.560 [perma.cc/9F-29N9PM]

Kain, D., Narayan, S., Sarkar, T., & Grover, G. (2021). *Online caste-hate speech: Pervasive discrimination and humiliation on social media.* Centre for Internet and Society. https://cis-india.org/internet-governance/blog/online-caste-hate-speech-pervasive-discrimination-and-humiliation-on-social-media [perma.cc/V6PT-P7NE]

KJ, S., Kumar, A., Balaji, L., Kotecha, N., Jain, V., Chadha, A., & Bhaduri, S. (2025). *IndicMMLU-Pro: Benchmarking Indic Large Language Models on Multi-Task Language Understanding* (arXiv:2501.15747). arXiv. https://doi.org/10.48550/arXiv.2501.15747 [perma.cc/D6J-2XAZK]

Kumar, R., Thanugonda, K., & Guha, D. (2022, December 18). *A Framework for Intermediary Classification in India—The Quantum Hub.* https://thequantumhub.com/a-framework-for-intermediary-classification-in-india/ [perma.cc/EHD-2AXF6]

Leong, W. Q., Ngui, J. G., Susanto, Y., Rengarajan, H., Sarveswaran, K., & Tjhi, W. C. (2023). *BHASA: A Holistic Southeast Asian Linguistic and Cultural Evaluation Suite for Large Language Models* (arXiv:2309.06085). arXiv. https://doi.org/10.48550/arXiv.2309.06085 [perma.cc/LBC6-X648]

Localization Lab and Internews. (2022, November 11). *"Wait, Who's Timothy McVeigh?":* Why localizing tech policies has to start with communities. Localization Lab. https://www.localizationlab.org/blog/11/11/2022/wait-whos-timothy-mcveigh-content-policies-review [perma.cc/RNN-5Y8NB]

Lorenz, T. (2022, April 8). Internet 'algospeak' is changing our language in real time, from 'nip nops' to 'le dollar bean.' *Washington Post*. https://www.washingtonpost.com/technology/2022/04/08/algospeak-tiktok-le-dollar-bean/ [perma.cc/RM-43BHRU]

Lumen. (n.d.). [Dataset]. Retrieved February 25, 2025, from https://lumendatabase.org/pages/about [perma.cc/6LZV-VGPD]

Mallwarachi, B. (2024). Sri Lanka passes bill allowing government to remove online posts and legally pursue internet users. *Associated Press*. https://apnews.com/article/sri-lanka-internet-bill-freedom-of-expression-wickremesinghe-527e0195a8f8f76573aa8d7185562b12 [perma.cc/7KEL-D8JL]

McCarthy, J. (2015, August 19). Up From The Ashes, A Public Library In Sri Lanka Welcomes New Readers. *NPR*. https://www.npr.org/sections/parallels/2015/08/19/432779251/up-from-the-ashes-a-public-library-in-sri-lanka-welcomes-new-readers [perma.cc/45ZX-CS3G]

Meta. (2022). *How we create and use market-specific slur lists.* https://transparency.meta.com/en-gb/enforcement/taking-action/how-we-create-and-use-market-slurs/ [perma.cc/9VN-6NUDU]

Mirza, R. (2023, November 28). Case Study: Integrity or Influence? Facebook's Governance Trade-offs in India and the Power of the Press. *Shorenstein Center.* https://shorensteincenter.org/case-study-integrity-influence-facebooks-governance-trade-offs-india-power-press/ [perma.cc/P74B26-JE]

Mukherjee, R., & Desai, S. (2024). *Anatomy of a "Chapri": Exploring the Decontextualization of Casteist Slurs on Social Media.* https://drive.google.com/file/d/1k6Cos64FGuQ1uXSEbf583TRyp_nawWKc/view?usp=embed_facebook [perma.cc/Z2BR7-AZ5]

Murugan, B., & Visalakshi, P. (2024). Ancient Tamil inscription recognition using detect, recognize and labelling, interpreter framework of text method. *Heritage Science,* 12(1), 1–21. https://doi.org/10.1186/s40494-024-01522-9 [perma.cc/3E2P-QEKN]

Nadaradjane, A. (2022, November 28). Sri Lanka: Democracy in Crisis. *The Centre for Independent Studies.* https://www.cis.org.au/publication/sri-lanka-democracy-in-crisis/ [perma.cc/E55M-VYGK]

Nanmalar, M., Vijayalakshmi, P., & Nagarajan, T. (2022). Literary and Colloquial Tamil Dialect Identification. *Circuits, Systems, and Signal Processing,* 41(7), 4004–4027. https://doi.org/10.1007/s00034-022-01971-2 [perma.cc/QHL-9A4A2]

Nicholas, G. (2022). *Shedding Light on Shadowbanning.* Center For Democracy And Technology. https://cdt.org/insights/shedding-light-on-shadowbanning/ [perma.cc/8XNP-UDFS]

Nicholas, G., & Bhatia, A. (2023a). *Lost in Translation: Large Language Models in Non-English Content Analysis* (arXiv:2306.07377). arXiv. https://doi.org/10.48550/arXiv.2306.07377 [perma.cc/RWL-3S3DM]

Nicholas, G., & Bhatia, A. (2023b). Toward Better Automated Content Moderation in Low-Resource Languages. *Journal of Online Trust and Safety,* 2(1), Article 1. https://doi.org/10.54501/jots.v2i1.150 [perma.cc/RWL-3S3DM]

Noolaham Foundation. (n.d.). Retrieved April 5, 2025, from https://noolaham.org/wiki/index.php/முதற்_பக்கம் [perma.cc/9B4E-CBSA]

Ramesh, A., Parthasarathy, V. B., Haque, R., & Way, A. (2020, December 4). *An error-based investigation of statistical and neural machine translation performance on Hindi-to-Tamil and English-to-Tamil.* 7th Workshop on Asian Translation (WAT2020), Suzhou, China (Online). https://www.aclweb.org/anthology/2020.wat-1.22 [perma.cc/B6J-5TE98]

Rowe, J. (2022). Marginalised languages and the content moderation challenge. *Global Partners Digital.*

Savolainen, L. (2022). The shadow banning controversy: Perceived governance and algorithmic folklore. *Media, Culture & Society*, 44(6), 1091–1109. https://doi.org/10.1177/01634437221077174 [perma.cc/AH5B-YHD9]

Schiffman, H. (1996). *Linguistic Culture and Language Policy.* https://www.sas.upenn.edu/~haroldfs/sars238/tamil238.html [perma.cc/U56P-RZ89]

Sehgal, D., & Nambiar, L. T. (2024). *Online Gender Based Violence on Short Form Video Platforms.* Centre for Internet and Society. https://cis-india.org/raw/online-gender-based-violence-on-short-form-video-platforms [perma.cc/9GD-8C538]

Shahid, F., & Vashistha, A. (2023). Decolonizing Content Moderation: Does Uniform Global Community Standard Resemble Utopian Equality or Western Power Hegemony? *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems,* 1–18. https://doi.org/10.1145/3544548.3581538 [perma.cc/36U-6EMBF]

Sivapriyan, E. (2025). Explained | Tamil Nadu's 2-language policy and opposition to NEP & Hindi. *Deccan Herald*. https://www.deccanherald.com//india/tamil-nadu/explained-tamil-nadus-2-language-policy-and-opposition-to-nep-hindi-3422150 [perma.cc/Y6KD-TH8Y]

Sombatpoonsiri, J., & Mahapatra, S. (2024). *Regulation or Repression? Government Influence on Political Content Moderation in India and Thailand.* Carnegie Endowment for International Peace. https://carnegieendowment.org/research/2024/07/india-thailand-social-media-moderation?lang=en [perma.cc/8L6S-UBXU]

Soundararajan, T., Kumar, A., Nair, P., & Greely, J. (2019). *Facebook India: Towards the Tipping point of voilence.* Equality Labs. https://www.equalitylabs.org/research/publications-resources/ [perma.cc/DF9V8-CCQ]

Subramaniam, S. (n.d.). Should India Speak a Single Language? *The New Yorker.* Retrieved February 24, 2025, from https://www.newyorker.com/magazine/2024/11/25/should-a-country-speak-a-single-language [perma.cc/XRS-5MY8F]

Tamil Guardian. (2023). *Tamil Eelam flags fly high in Canada and UK.* https://www.tamilguardian.com/content/tamil-eelam-flags-fly-high-canada-and-uk [perma.cc/Q5PQ-SQ7D]

Tamil Guardian. (2024). Sri Lanka arrests Tamil man under PTA for Facebook post | Tamil Guardian. *Tamil Guardian.* https://www.tamilguardian.com/content/sri-lanka-arrests-tamil-man-under-pta-facebook-post [https://perma.cc/4YPA-L2J]

Tattle. (2024). *Online Gender-Based Slurs and Abuses Don't Just Happen in English.* https://tattle.co.in/blog/2024-05-09-slurs-occur-every-language/ [perma.cc/AA3A-GKJH]

Thakur, D., & Liansó, E. (2021). Do You See What I See? *Capabilities and Limits of Automated Multimedia Content Analysis.* https://cdt.org/insights/do-you-see-what-i-see-capabilities-and-limits-of-automated-multimedia-content-analysis/ [perma.cc/U4LX2-SY8]

Thakur, D., & Hankerson, D. L. (2022). *An Unrepresentative Democracy: How Disinformation and Online Abuse Hinder Women of Color Political Candidates in the United States.* Center For Democracy And Technology. https://cdt.org/insights/an-unrepresentative-democracy-how-disinformation-and-online-abuse-hinder-women-of-color-political-candidates-in-the-united-states/ [perma.cc/FM2L8-VCP]

Tiku, N. (2022, June 2). Google's plan to talk about caste bias led to 'division and rancor.' *Washington Post*. https://www.washingtonpost.com/technology/2022/06/02/google-caste-equality-labs-tanuja-gupta/ [perma.cc/5SSL99-ME]

Times of India. (2024, September 13). DMK and BJP clash over Tamil Nadu's two-language policy amid NEP controversy. *The Times of India*. https://timesofindia.indiatimes.com/city/chennai/dmk-and-bjp-clash-over-tamil-nadus-two-language-policy-amid-nep-controversy/articleshow/113321462.cms [perma.cc/M2LW-CXQG]

Treem, J. W., Leonardi, P. M., & van den Hooff, B. (2020). Computer-Mediated Communication in the Age of Communication Visibility. *Journal of Computer-Mediated Communication*, 25(1), 44–59. https://doi.org/10.1093/jcmc/zmz024 [perma.cc/9QK-4T9BW]

u/diegomrosa. (2023, July 20). *Did some Facebook languages disappeared? Which was the highest number of languages supported by Facebook?* [Reddit Post]. R/Facebook. www.reddit.com/r/facebook/comments/154rmtt/did_some_facebook_languages_disappeared_which_was/ [perma.cc/DUH-6KLCQ]

UN. (2011). *UN Guiding Principles.* Business & Human Rights Resource Centre. https://www.business-humanrights.org/en/big-issues/governing-business-human-rights/un-guiding-principles/ [perma.cc/ZP2V-HQZ3]

UNHCR IRIN. (2012). *Bridging the language divide in Sri Lanka*. https://webarchive.archive.unhcr.org/20230527001057oe_/https://www.refworld.org/docid/501005892.html [perma.cc/4YVC-MWNG]

Venkatachalapathy, A. R. (2022). *The Brief History of a Very Big Book: The Making of the Tamil Encyclopaedia.* https://orientblackswan.com/details?id=9788178246604 [perma.cc/7BFL-SUSE]

Vesteinsson, K. (2024). *Ahead of Landmark Elections, India's Government Silences Dissent.* Freedom House. https://freedomhouse.org/article/ahead-landmark-elections-indias-government-silences-dissent [perma.cc/RPU99-5KN]

Vilk, V. (2023, June 29). *Shouting into the Void*. PEN America. https://pen.org/report/shouting-into-the-void/ [perma.cc/J2NL-ZD37]

Watts, I., Gumma, V., Yadavalli, A., Seshadri, V., Manohar, S., & Sitaram, S. (2024). *PARIKSHA: A Scalable, Democratic, Transparent Evaluation Platform for Assessing Indic Large Language Models.* https://www.microsoft.com/en-us/research/publication/pariksha-a-scalable-democratic-transparent-evaluation-platform-for-assessing-indic-large-language-models/ [perma.cc/3PSQ9-UD3]

Yatharth. (2025). *Bahujan Digital Publishing Infrastructures.* Centre for Internet and Society. https://cis-india.org/raw/bahujan-digital-publishing-infrastructures [perma.cc/Q2WF-ANE2]

cdt.org

cdt.org/contact

Center for Democracy & Technology
1401 K Street NW, Suite 200
Washington, D.C. 20005

202-637-9800

@CenDemTech

CENTER FOR
DEMOCRACY
& TECHNOLOGY