தமிழ் செயற்கை நுண்ணறிவு ஆய்விதழ்
மதிப்பாய்வு செய்யப்படும் இதழ்

Tamil Journal of Artificial Intelligence
Peer-Reviewed Journal

Volume 1, SPL 1 – July 2026

தமிழ்மொழியும் செயற்கை நுண்ணறிவும்

முனைவர் அ.தேன்மொழி, உதவிப் பேராசிரியர், தமிழ்த்துறை (சுயநிதிப்பிரிவு) , மன்னர் திருமலை நாயக்கர் கல்லூரி, மதுரை, தமிழ்நாடு, இந்தியா.
ORCiD: https://orcid.org/0009-0001-6446-2381

அறிமுகம்

இன்றைய உலகில் செயற்கை நுண்ணறிவு மனித வாழ்வின் பல துறைகளில் வேகமாகப் பரவி வருகிறது. மொழி, கல்வி, மருத்துவம், வணிகம், ஊடகம், ஆராய்ச்சி போன்ற துறைகளில் அதன் பயன்பாடு நாளுக்கு நாள் அதிகரிக்கிறது. இந்த மாற்றத்தின் நடுவில், தமிழ்மொழிக்கும் செயற்கை நுண்ணறிவுக்கும் இடையிலான உறவு மிகவும் முக்கியமான ஒன்றாக மாறியுள்ளது. "கல் தோன்றி மண் தோன்றாக் காலத்தே முன் தோன்றிய மூத்த குடி" என்று போற்றப்படும் தமிழ்மொழி, உலகின் மிக மூத்த உயர்தனிச்செம்மொழியாகும். மொழி என்பது வெறும் தகவல் பரிமாற்றக் கருவி மட்டுமல்ல; அது ஒரு இனத்தின் பண்பாடு, வரலாறு மற்றும் வாழ்வியலின் அடையாளம். உலகின் மூத்த மொழியான தமிழ், ஏடுகளிலும் கல்வெட்டுகளிலும் பயணித்து, இன்று கணினித் திரைகளையும் கடந்து செயற்கை நுண்ணறிவு (Artificial Intelligence – AI) என்னும் நவீன தொழில்நுட்ப உலகிற்குள் நுழைந்துள்ளது. காலம் தோறும் ஏற்படும் மாற்றங்களையும், தொழில்நுட்ப வளர்ச்சிகளையும் உள்வாங்கிக் கொண்டு தன்னைப் புதுப்பித்துக் கொள்ளும் ஆற்றல் தமிழுக்கு உண்டு. அந்த வகையில், 21-ஆம் நூற்றாண்டின் ஆகச்சிறந்த தொழில்நுட்பப் புரட்சியான செயற்கை நுண்ணறிவுடன் (Artificial Intelligence - AI).

செயற்கை நுண்ணறிவு

செயற்கை நுண்ணறிவு என்பது மனிதர்கள் சிந்தித்து முடிவெடுக்கிறதுபோல, கணினி அல்லது இயந்திரங்கள் செயல்படச் செய்யும் தொழில்நுட்பமாகும். இதன் அடிப்படை நோக்கம், ஒரு அமைப்பு தரவுகளைப் புரிந்து கொண்டு, கற்றுக் கொண்டு, அதன்படி முடிவெடுக்கச் செய்வதாகும். “இந்தப் புதிய உலகத்தில் பயன்படுபவர்கள் காணாமல் போவார்கள். கற்றுக் கொள்பவர்கள் முன்னணிக்கு செல்வார்கள்.”[1] செயற்கை நுண்ணறிவு மனித நுண்ணறிவை முழுமையாகப் பிரதிபலிக்காதபோதிலும், குறிப்பிட்ட பணிகளில் மனிதனை விட வேகமாகவும் துல்லியமாகவும் செயல்பட முடியும். மனித மூளையின் செயல்பாடுகளைப் போல, இயந்திரங்களுக்கும் பகுத்தறியும் மற்றும் கற்கும் திறனை வழங்கும் அறிவியல் புலமான செயற்கை நுண்ணறிவில் மூன்று முக்கிய அடுக்குகள் உள்ளன. அவை,

- இயந்திரக் கற்றல் (**Machine Learning – ML):** தரவுகளிலிருந்து (**Data**) கணினி தானாகவே விதிகளைக் கற்றுக்கொள்வது.
- ஆழ்ந்த கற்றல் (**Deep Learning – DL):** மனித நரம்பு மண்டலம் போன்ற செயற்கை நரம்பியல் வலைப்பின்னல்களை (**Neural Networks**) பயன்படுத்தி சிக்கலான தரவுகளைப் பகுப்பாய்வு செய்வது.
- இயற்கை மொழி செயலாக்கம் (**Natural Language Processing – NLP):** மனிதர்கள் பேசும் மற்றும் எழுதும் மொழியை இயந்திரங்கள் புரிந்து கொள்ள உதவும் தொழில்நுட்பம். இந்த மூன்றாவது அடுக்குதான் தமிழும் **AI-**உம் இணையும் மையப்புள்ளியாகும்.
- **NLP** தொழில்நுட்பம் முக்கியமாக இரண்டு பகுதிகளைக் கொண்டது:
- **NLU (Natural Language Understanding):** மனித மொழியின் சூழல் மற்றும் உட்பொருளைக் கணினி துல்லியமாகப் புரிந்துகொள்வது.

Volume 1, SPL 1 – July 2026

- **NLG (Natural Language Generation):** தரவுகளின் அடிப்படையில் கணினியே மனித மொழியில் வாக்கியங்களை உருவாக்குவது (எ.கா: **ChatGPT** தமிழில் கவிதை எழுதுவது).

தமிழ்மொழிக்கு NLP -இன் தேவை

இன்றைய டிஜிட்டல் யுகத்தில் பெரும்பாலான தகவல்களும் சேவைகளும் இணையம் வழியாகவே கிடைக்கின்றன. ஆங்கிலம் தெரியாத ஒரு சாமானியத் தமிழ் குடிமகனும் தொழில்நுட்பத்தின் பயன்களை முழுமையாகப் பெற வேண்டுமெனில், கணினிகள் தமிழில் உரையாட வேண்டும். வங்கிகள், அரசுத் துறைகள், கல்வி நிறுவனங்கள் போன்றவற்றில் தமிழ் வழியே சேவைகளைப் பெற தமிழ் NLP அத்தியாவசியமாகிறது.

தமிழின் தனித்தன்மை

தமிழ் உலகின் மிகப் பழமையான உயிருள்ள மொழிகளில் ஒன்று. சங்க இலக்கியம், பக்தி இலக்கியம், நீதிநூல்கள், சிற்றிலக்கியங்கள், நவீன இலக்கியம் எனப் பல்வேறு அடுக்குகளில் இதன் செழுமை காணப்படுகிறது. தமிழுக்கு தனித்தெழுத்து முறை, ஒலி அமைப்பு, இலக்கண ஒழுங்கு, சொல்வளம் ஆகியவை சிறப்பாக உள்ளன. இந்த பண்புகள் காரணமாக தமிழை கணினியில் கையாள்வது சில சமயங்களில் சவாலானதாக இருந்தாலும், அதே நேரத்தில் அது மிகவும் மதிப்புமிக்க மொழியாகவும் விளங்குகிறது. தமிழின் மற்றொரு முக்கிய அம்சம் அதன் சமூகத் தொடர்ச்சி. கல்வி, அலுவல், ஊடகம், கலை, விவசாயம், தொழில், ஆன்மீகம் என பல துறைகளில் அது தொடர்ந்து பயன்படுகிறது. இணையத் தலைமுறையிலும் தமிழ் தனது இடத்தை உறுதிப்படுத்தி வருகிறது. இதனால் தமிழ்மொழிக்கான AI கருவிகள் உருவாக்குவது வெறும் தொழில்நுட்ப முயற்சி மட்டுமல்ல; அது பண்பாட்டு பாதுகாப்பு முயற்சியுமாகும்.

கணினி மொழியியலும் தமிழின் இலக்கணச் சிறப்பும்

“பெயர்ச்சொல், வினைச்சொல், பெயரடை, வினையடை என சொற்களை அடையாளம் காணுதலும் வாக்கியத்தின் ஒவ்வொரு சொல்லையும் விளக்குதலும் ஒரு பனுவலுக்கு வடிவம் கொடுப்பதற்கு அடிப்படையாக அமைவது எதுவோ அதனைக் கண்டறிவதும் சொல்வகை விளக்கல் (**Part-of-speech Tagging**) என்று பெயர். இதனை சுருக்கமாக **POS Tagging** என்றும் சொல்வகை அறிதல் என்றும் அழைப்பார்கள்.”[2] தமிழ்மொழி ஏன் AI தொழில்நுட்பத்திற்கு மிகவும் பொருத்தமானது என்பதைப் பின்வரும் கணினி மொழியியல் (**Computational Linguistics**) காரணிகள் விளக்குகின்றன.

துல்லியமான இலக்கண விதிகள்

தொல்காப்பியமும் நன்னூலும் வகுத்துள்ள எழுத்து, சொல், பொருள் இலக்கண விதிகள் கணினி நிரலாக்கத்திற்கு (**Coding**) நெருக்கமானவை. தமிழின் புணர்ச்சி விதிகள் (நிலைமொழியும் வருமொழியும் சேரும் விதிகளுக்கு) தெளிவான கணிதத் தன்மை உடையவை. தமிழில் எழுத்துக்கள் ஒலி மாறுபாடற்றவை (**Phonetic Consistency).** ஆங்கிலத்தில் **‘Cut’** என்பதில் வரும் **‘U’** ஒலியும், **‘Put’** என்பதில் வரும் **‘U’** ஒலியும் மாறுபடும். ஆனால் தமிழில் ‘அ’ என்றால் எங்கும் ‘அ’ தான். இது AI-இன் பேச்சு அறிதல் (**Speech Recognition)** திறனுக்குப் பெரும் சாதகமாகும்.

ஒட்டுநிலை மொழி அமைப்பு (Agglutinative Language)

தமிழ் ஒரு ஒட்டுநிலை மொழியாகும். அதாவது, ஒரு வேர்ச்சொல்லுடன் பல விகுதிகளைச் சேர்த்துப் புதிய சொற்களை உருவாக்க முடியும் (எடு: வா வந்தான் வந்துகொண்டிருந்தான் வரவழைக்கப்பட்டார்). இந்த அமைப்பை கணித சூத்திரங்களாக மாற்றுவது **AI-**இன் அல்காரிதம்களுக்கு (**Algorithms**) மிகவும் எளிதானது.

Volume 1, SPL 1 – July 2026

தமிழின் ஒட்டுநிலை அமைப்பும் அதன் கணிதத் தன்மையும்

தமிழில் ஒரு வேர்ச்சொல் எவ்வாறு பல அடுக்கு ஒட்டுகளை ஏற்று, ஒரு முழு வாக்கியத்திற்கு இணையான பொருளை ஒரே சொல்லில் தருகிறது என்பதைப் பின்வரும் சான்றுகள் விளக்குகின்றன.

சொல்லாக்க அடுக்குமுறை (Morphological Layering)

உதாரணமாக, "படி" என்ற வேர்ச்சொல்லை எடுத்துக் கொள்வோம்:

படி + த் + த் + ஆன் = படித்தான் (வினைமுற்று – ஆண்பால்)

படி + த் + த் + ஆர் + கள் = படித்தார்கள் (பலர்பால் முற்று)

படி + ப்ப் + ஐ + த் + த் + அ + வர் + கள் = படித்தவர்கள் (வினையாலணையும் பெயர்)

படி + க்க + வை + க்க + ப் + பட் + டார் + கள் = படிக்கவைக்கப்பட்டார்கள் (செயப்பாட்டு வினை)

மேற்கண்ட எடுத்துக்காட்டில், "படிக்கவைக்கப்பட்டார்கள்" என்ற ஒரே சொல், ஆங்கிலத்தில் **"They were made to study"** என்ற ஐந்து சொற்கள் கொண்ட வாக்கியத்திற்கு இணையானது. தமிழின் இந்த ஒட்டுமுறை என்பது கணினியின் நிரலாக்கக் குறியீடுகளைப் **(Coding)** போன்றது. ஒரு குறிப்பிட்ட விதியின் கீழ் ஒட்டுகளைச் சேர்க்கும் போது (நிலைமொழி + வருமொழி = புணர்ச்சி விதிகள்), அங்குத் துல்லியமான கணிதத் தன்மை வெளிப்படுகிறது.

வேர்ச்சொல் + காலம் + பால்/திணை/எண் = புதிய சொல்

இந்தத் துல்லியமான சூத்திர அமைப்பானது, செயற்கை நுண்ணறிவின் அல்காரிதம்கள் **(Algorithms)** ஒரு சொல்லின் பொருளைப் பகுத்தாய்வதற்கும், புதிய சொற்களைத் தானாகவே உருவாக்குவதற்கும் **(Generative AI)** மிகச்சிறந்த அடித்தளமாக அமைகிறது.

சொற்பரிமாற்ற சுதந்திரம் (Free Word Order)

ஆங்கிலத்தில் **Subject + Verb + Object (SVO)** என்ற வரிசை முறை மாறினால் வாக்கியத்தின் பொருளே மாறிவிடும் அல்லது பிழையாகிவிடும். ஆனால், தமிழில் ஒட்டுநிலை மற்றும் வேற்றுமை உருபுகளின் துணையோடு சொற்கள் அமைவதால், சொற்களின் இடத்தை எப்படி மாற்றினாலும் பொருள் மாறாது. உதாரணமாக,

ராமன் ராவணனைக் கொன்றான்.

ராவணனை ராமன் கொன்றான்.

கொன்றான் ராமன் ராவணனை.

இந்த மூன்று வாக்கியங்களும் ஒரே பொருளைத்தான் தருகின்றன. இந்த **'Free Word Order'** தனித்துவம், செயற்கை நுண்ணறிவு ஒரு வாக்கியத்தின் ஆழமான சூழல் பொருளைப் **(Contextual Meaning)** புரிந்துகொள்வதற்கும், மனிதர்களைப் போலப் பல வடிவங்களில் உரையாடுவதற்கும் **(Generative AI Chatbots)** பெரும் சுதந்திரத்தை அளிக்கிறது.

தமிழ் NLP சந்திக்கும் முக்கியச் சவால்கள்

தங்கிலீஷ் **(Tanglish)** மற்றும் பேச்சுவழக்கு: இணையத்தில் மக்கள் தமிழை ஆங்கில எழுத்துக்களில் (எ.கா: **"Eppadi irukkinga?"**) எழுதுவதாலும், மாவட்டத்திற்கு மாவட்டம் பேச்சுவழக்கு மாறுவதாலும் AI குழப்பமடைகிறது. தரவுப் பற்றாக்குறை **(Data Scarcity):** ஆங்கில மொழியோடு ஒப்பிடும்போது, AI-க்குப் பயிற்சி அளிக்கத் தேவையான தரமான, பிழையற்ற தமிழ்த் தரவுத் தொகுப்புகள் **(Annotated Datasets)** இணையத்தில் குறைவாகவே உள்ளன.

தமிழ் NLP-இன் நிகழ்காலப் பயன்பாடுகள்

இன்று நாம் அன்றாடம் பயன்படுத்தும் பல தொழில்நுட்பங்களின் பின்னணியில் தமிழ் **NLP** இயங்குகிறது:

தமிழ் செயற்கை நுண்ணறிவு ஆய்விதழ்
மதிப்பாய்வு செய்யப்படும் இதழ்
Tamil Journal of Artificial Intelligence
Peer-Reviewed Journal
Volume 1, SPL 1 – July 2026

- கூகுள் மொழிபெயர்ப்பு (**Google Translation**): பிற மொழிகளில் உள்ள கட்டுரைகளை நொடிப்பொழுதில் தமிழில் மொழிபெயர்க்கப் பயன்படுகிறது.
- குரல் தேடல் (**Voice Assistants**): கூகுள் அசிஸ்டெண்ட் அல்லது அலெக்சாவிடம் தமிழில் பேசும்போது, நம் பேச்சை உரைவடிவமாக மாற்றிப் புரிந்துகொள்ள **NLP** உதவுகிறது.
- பிழை திருத்திகள் (**Spell Checkers**): வாணி, மென்தமிழ் போன்ற செயலிகள் தமிழில் சந்திப்பிழைகளையும் எழுத்துப்பிழைகளையும் தானாகவே திருத்துகின்றன.

டிஜிட்டல் சுவடியாக்கம் AI-இன் பங்கு

தமிழ்மொழியின் பழங்கால வரலாற்று ஆவணங்களைப் பாதுகாப்பதில் செயற்கை நுண்ணறிவு ஒரு புரட்சியை ஏற்படுத்தி வருகிறது.

ஓலைச்சுவடிகளை வாசித்தல் (Manuscript Recognition)

ஆயிரக்கணக்கான தமிழ் ஓலைச்சுவடிகள் இன்னும் வாசிக்கப்படாமலும், சிதைந்த நிலையிலும் உள்ளன. கணினிப் பார்வை (**Computer Vision**) மற்றும் AI சார்ந்த **OCR (Optical Character Recognition)** தொழில்நுட்பங்கள் மூலம், சிதைந்த ஓலைச்சுவடிகளில் உள்ள எழுத்து வடிவங்களை பகுப்பாய்வு செய்ய முடியும். எழுத்துக்கள் அழிந்த இடங்களில், முந்தைய மற்றும் பிந்தைய வரிகளின் சூழலை வைத்து, அங்கு என்ன சொல் இருந்திருக்கும் என்பதை **AI** துல்லியமாக ஊகித்துக் கூறுகிறது.

எழுத்து வடிவ மாற்றமும் சிக்கல்களும்

தமிழ் பிராமி (தமிழி), வட்டெழுத்து, கிரந்தம் போன்ற பழங்கால எழுத்து வடிவங்களை இன்றைய தலைமுறையினரால் எளிதில் வாசிக்க முடியாமை. வெவ்வேறு கால கட்டங்களில் வாழ்ந்த புலவர்கள் மற்றும் சுவடி எழுதுவோரின் (**Scribes**) தனித்துவமான கையெழுத்து முறைகள் போன்றவை ஓலைச்சுவடிகளை வாசிப்பதில் சிக்கலாக அமைகின்றன.

கணினிப் பார்வையும் (Computer Vision) தமிழ் ஓலைச்சுவடிகளை மீட்டெடுத்தலும்

நவீன செயற்கை நுண்ணறிவின் ஒரு பிரிவான கணினிப் பார்வை (**Computer Vision**) மற்றும் ஆழ்ந்த கற்றல் (**Deep Learning**) மாதிரிகள், ஓலைச்சுவடிகளைத் துல்லியமாக டிஜிட்டல் உரையாக (**Digital Text**) மாற்றுகின்றன.

பிம்ப மேம்பாடு (Image Preprocessing)

சிதைந்த ஓலைச்சுவடிகளின் புகைப்படங்களை AI முதலில் செப்பனிடுகிறது. **'Sauvola Adaptive Thresholding'** மற்றும் **'Denoising'** போன்ற நுட்பங்கள் மூலம், ஓலையின் பின்புறம் உள்ள நிழல்கள், கறைகள் மற்றும் பூஞ்சை அடையாளங்கள் நீக்கப்பட்டு, எழுத்துக்கள் மட்டும் தனியாகத் துலங்குமாறு (**Contrast Enhancement**) செய்யப்படுகிறது. ஓலையின் நடுவே உள்ள துளைகளை (**Punch Holes**) உரைப்பகுதியாகக் கருதி AI குழப்பமடையாமல் இருக்க, பிரத்யேக அல்காரிதம்கள் பயன்படுத்தப்படுகின்றன.

'கல்வெட்டு AI' (Kalveetu AI) மற்றும் எழுத்து மாற்றம்

தமிழ்நாட்டின் கோயில் சுவர்களிலும், நடுகற்களிலும் உள்ள ஆயிரக்கணக்கான கல்வெட்டுகளை ஆவணப்படுத்துவதில் AI ஒரு புதிய புரட்சியை ஏற்படுத்தியுள்ளது. சமீபத்திய ஆராய்ச்சிகளின் மூலம் உருவாக்கப்பட்ட **'கல்வெட்டு AI' (Kalveetu AI)** போன்ற ஆழ்ந்த கற்றல் அமைப்புகள், பழங்காலத் தமிழ் பிராமி மற்றும் வட்டெழுத்து வடிவங்களை நவீன தமிழ் எழுத்துக்களாக (**Transliteration**) மாற்றுகின்றன. பல கல்வெட்டுகள் தேய்மானம் அடைந்து, எழுத்துக்கள் முழுமையாகத் தெரியாத நிலையில் இருக்கும். இந்தத் தரவுப் பற்றாக்குறையைச் சரிசெய்ய, AI-இன் **GAN (Generative Adversarial Networks)** தொழில்நுட்பம் பயன்படுத்தப்படுகிறது. இது விடுபட்ட அல்லது உடைந்த எழுத்துக்களின் வடிவங்களை அதன்

Volume 1, SPL 1 – July 2026

வரலாற்றுச் சூழலைக் கொண்டு தானாகவே முழுமையாக்கி, கல்வெட்டறிஞர்கள் அதனை எளிதாக வாசிக்க வழிவகை செய்கிறது. இதன் மூலம் 'பாரத்ஸ்ரீ' **(BharatSHRI)** போன்ற டிஜிட்டல் களஞ்சியத் திட்டங்களில் தமிழ்க் கல்வெட்டுகள் வேகமாகப் பாதுகாக்கப்பட்டு வருகின்றன. சுவடிகளிலோ கல்வெட்டுகளிலோ ஒரு செய்யுளின் சில சொற்கள் பூச்சிகளால் அரிக்கப்பட்டு அழிந்திருந்தால், இயற்கை மொழி செயலாக்கம் **(NLP)** மூலம் இயங்கும் பெரிய மொழி மாதிரிகள் **(LLMs)** அதனை மீட்டெடுக்கின்றன. யாப்பு இலக்கண விதிகள் (எதுகை, மோனை, தளை, சீர்) மற்றும் அகராதித் தரவுகளைக் கொண்டு, "இந்த இடத்தில் இந்தச் சொல்தான் வந்திருக்க முடியும்" என்பதை AI கணித்துக் கூறுகிறது. உதாரணமாக, ஒரு சங்கப் பாடலின் ஒரு வரியில் ‘மயிலாடு’ என்று அழிந்திருந்தால், அதன் திணை, துறை மற்றும் செய்யுள் இலக்கணத்தை ஆராய்ந்து அது மயிலாடும் என்ற சொல்தான் என்பதை AI பரிந்துரைக்கும் (**Predictive Decipherment).**

கல்வித் துறையில் பங்கு

கல்வித்துறையில் செயற்கை நுண்ணறிவு மிகப்பெரிய மாற்றத்தை ஏற்படுத்தி வருகிறது. தமிழில் AI பயன்படுத்தப்படும்போது மாணவர்களின் கற்றல் திறன் மேம்படுகிறது. உதாரணமாக, ஒரு மாணவர் இலக்கணம், சொல்லாட்சி, இலக்கியக் கருத்து, கட்டுரை அமைப்பு ஆகியவற்றை AI மூலம் பயிற்சி செய்து கொள்ளலாம். இது தனிப்பட்ட ஆசிரியரின் வழிகாட்டுதலை ஒத்த அனுபவத்தைத் தரக்கூடும். ஆசிரியர்களுக்கும் AI உதவிக்கரமாக உள்ளது. பாடத்திட்டத் தயாரிப்பு, தேர்வுக் கேள்விகள் உருவாக்கம், விடைத்தாள் திருத்தத்துக்கு முன்னுரிமை கொடுத்தல், மாணவர் முன்னேற்றம் பகுப்பாய்வு போன்ற பணிகளில் அது உதவுகிறது. தமிழ்மொழிக் கல்வியில் இது முக்கியமான முன்னேற்றம். குறிப்பாகக் கிராமப்புறம் மற்றும் குறைந்த வளமுள்ள பகுதிகளில் தமிழில் செயல்படும் AI கல்விச் சேவைகள் சம வாய்ப்பை உருவாக்கக்கூடும்.

மின்வெளி கற்றலின் சிறப்பியல்புகள்

1. **அனைத்து பாடங்களையும் மின்னணு கற்றல் முறைக்கு மாற்றியமைக்க இயலும்.**
2. **சுயகற்றல், சுய வேகத்தின் படி பாடங்கள் கற்பிக்கப்படுகிறது**
3. **கல்விசார் வளங்களை பெறுவதற்கான வழிமுறைகளைக் காணலாம்.**
4. **பாடங்களை எளிதில், விரைவாக உருவாக்கவும், நவீனபடுத்தவும்,**
5. **திருத்தியமைக்கவும் இயலும். இவற்றை மலிவான விலைவில் நிறைவேற்ற இயலும்**
6. **துடிப்பான கற்றல் அனுபவங்களை ஆன்லைன் மதிப்பீட்டு கருவிகள், நகரும் பாடங்கள் மூலம் பெறலாம்.**”[3]

இலக்கியம் மற்றும் படைப்பாற்றல்

தமிழ் இலக்கியம் **AI**க்கு மிகப் பெரிய வளமாகவும் சவாலாகவும் உள்ளது. சங்க இலக்கியம் முதல் நவீன கவிதை வரை, தமிழ் படைப்புகள் தனித்துவமான உருவம், இசை, குறியீடு, உவமை, மரபு ஆகியவற்றைக் கொண்டவை. AI கருவிகள் இவ்வுலகை புரிந்து கொண்டு, இலக்கிய ஆய்வுக்கு உதவ முடியும். உதாரணமாக, குறிப்பிட்ட கவிஞரின் சொல்வழக்கம், ஒரு காலகட்டத்தின் கருத்தியல் போக்கு, ஒரு நூலின் கருப்பொருள் மாற்றங்கள் போன்றவற்றை AI பகுப்பாய்வு செய்யலாம். ஆனால் இலக்கியத்தை உருவாக்கும் படைப்பாற்றல் என்பது வெறும் தரவு மாதிரியல்ல. ஒரு கவிதையின் உணர்வு, சொல் தேர்வு, அடிக்குரல், பண்பாட்டு ஆழம் ஆகியவை மனித அனுபவத்துடன் நெருக்கமாக இணைந்தவை. AI இங்கு உதவியாளராக இருக்க முடியும்; ஆனால் மனித எழுத்தாளரின் உள்ளுணர்வு, சமூகப் புரிதல், அழகியல்திறன் ஆகியவற்றை முழுமையாக மாற்ற முடியாது.

Volume 1, SPL 1 – July 2026

மொழிபெயர்ப்பு மற்றும் அணுகல்

தமிழ் பேசும் மக்களுக்கு உலக அறிவை அணுகுவதில் மொழிபெயர்ப்பு மிக முக்கியமானது. AI அடிப்படையிலான மொழிபெயர்ப்பு இயந்திரங்கள் தமிழ்-ஆங்கிலம், தமிழ்-இந்தி, தமிழ்-மலையாளம், தமிழ்-சிங்களம் போன்ற பல திசைகளில் பயன்படலாம். இது வணிகம், கல்வி, அரசுத் தொடர்பு, மருத்துவ விளக்கம், சட்ட ஆவணங்கள் போன்ற துறைகளில் பெரும் பயன்தரும். அணுகல் வசதிகளும் முக்கியம். குரல் வழிக் கையேடுகள், வாசிப்பு வசதி, தானியங்கி வசனம், துணை எழுத்து, ஒலி விளக்கம் போன்றவை அனைவருக்கும் தமிழ் தகவலை எளிதாக அணுக உதவும். குறிப்பாகப் பார்வைக் குறைபாடுள்ளோர், எழுதப் படிக்கச் சிரமப்படுவோர், முதியோர் ஆகியோருக்குப் பெரிய ஆதரவாக இது அமையும். **AI** இங்கு சமத்துவத்தின் கருவியாக மாறும்.

அரசுத்துறையில் AI

மத்திய அரசிடமிருந்தோ அல்லது பிற மாநிலங்களிடமிருந்தோ வரும் ஆங்கில மற்றும் இந்தி ஆவணங்களை, மாநில அரசுத் துறைகள் உடனுக்குடன் தமிழில் மொழிபெயர்க்க வேண்டியுள்ளது. மனித மொழிபெயர்ப்பாளர்கள் செய்யும் போது இதற்குப் பல நாட்கள் ஆகலாம். ஆனால், அரசாங்கத் தரவுகளைக் கொண்டு பிரத்யேகமாகப் பயிற்றுவிக்கப்பட்ட **AI** இயந்திர மொழிபெயர்ப்பான்கள் (**Machine Translation),** ஆவணத்தின் அசல் பொருள் மாறாமல், நொடிப்பொழுதில் சட்ட விதிகளுக்கு உட்பட்ட துல்லியமான தமிழ் மொழிபெயர்ப்பை வழங்குகின்றன.

மொழியியல் பன்முகத்தன்மையும் வட்டார வழக்குச் சவால்களும்

தமிழ்மொழி உலகளாவிய பரவலைக் கொண்டது. தமிழ்நாடு, இலங்கை, சிங்கப்பூர், மலேசியா எனப் பல்வேறு நாடுகளில் தமிழ் பேசப்படுகிறது. தமிழ்நாட்டிற்குள்ளேயே நெல்லைத் தமிழ், கொங்குத் தமிழ், சென்னைத் தமிழ், மத்திய மாவட்டத் தமிழ் எனப் பல வட்டார வழக்குகள் உள்ளன. இலங்கையின் யாழ்ப்பாணத் தமிழ்நடை முற்றிலும் தனித்துவமானது. தற்போதைய **AI** கருவிகள் பொதுவாக அச்சு ஊடகங்களில் உள்ள “முறையான எழுத்துத் தமிழை” **(Formal Written Tamil)** மட்டுமே கொண்டு பயிற்றுவிக்கப்படுகின்றன. இதன் காரணமாக, ஒரு கிராமப்புற மனிதரோ அல்லது இலங்கைத் தமிழரோ தங்களின் இயல்பான பேச்சுவழக்கில் **AI** குரல் தேடலிலோ (**Voice Search)** அல்லது சாட்பாட்களிலோ பேசும்போது, அவர்களின் உச்சரிப்பையும் சொற்களின் தேர்வையும் **AI** துல்லியமாகப் புரிந்து கொள்ளத் தவறுகிறது.

பேச்சுத்தமிழும் எழுத்துத்தமிழும்

தமிழில் பேச்சு மொழிக்கும் எழுத்து மொழிக்கும் இடையே உள்ள வேறுபாடு மிக அதிகம். “வந்துகொண்டிருக்கிறேன்” என்று எழுதுவதை “வந்துட்டு இருக்கேன்” என்று பேசுவது மரபு. **AI** மாதிரிகள் இந்த இரண்டு வடிவங்களுக்கும் இடையே உள்ள செமண்டிக் தொடர்பை (**Semantic Relationship)** முழுமையாகக் கற்க இன்னும் சிரமப்படுகின்றன.

தங்கிலீஷ் (Tanglish) **கலாச்சாரமும் தரவுச் சிதைவுகளும்**

இணையத்தில், குறிப்பாக முகநூல், எக்ஸ் **(X),** வாட்ஸ்அப் போன்ற சமூக ஊடகங்களில் பெரும்பாலான தமிழர்கள் தமிழ் எழுத்துருக்களைப் **(Unicode)** பயன்படுத்துவதில்லை. மாறாக, தமிழ் வார்த்தைகளை ஆங்கில எழுத்துக்களில் (தங்கிலீஷ் – **Tanglish)** எழுதும் வழக்கம் பரவலாக உள்ளது.

“Eppadi irukkinga?” (**எப்படி இருக்கீங்க**?)

“Naan nallaa irukken” (நான் நல்லா இருக்கேன்)

Volume 1, SPL 1 – July 2026

AI மாதிரிகள் இணையத்தில் இருந்து தரவுகளைச் சேகரிக்கும் போது இந்தத் தங்கிலீஷ் சொற்கள் தூய ஆங்கிலச் சொற்களோடு கலந்து, தரவுச் சிதைவுகள் ஏற்படுத்துகின்றன.

மொழிச் சிதைவுகள்

இந்தச் சிதைந்த தரவுகளைக் கொண்டு பயிற்றுவிக்கப்படும் **AI** மாதிரிகள், முறையான தமிழ் இலக்கண விதிகளைத் தவறவிடுகின்றன. மேலும், ஆங்கிலம் மற்றும் தமிழ் சொற்கள் கலந்து எழுதப்படும் வாக்கியங்களின் உணர்வுகளை (**Sentiment Analysis**) பகுப்பாய்வு செய்வதில் AI-க்குத் துல்லியம் பாதியாகக் குறைகிறது. தங்கிலீஷை முறையான தமிழுக்கு மாற்றும் வலுவான இடைநிலை அல்காரிதம்கள் (**Transliteration Models**) இன்னும் முழுமையான நிலையை எட்டவில்லை.

AI மாதிரிகளின் தவறான கற்பிதங்களும் தமிழ்க் கலாச்சார முரண்பாடுகளும் (Hallucination in Tamil)

AI தொழில்நுட்பத்தில் ‘ஹாலுசினேஷன்’ (**Hallucination**) என்பது ஒரு பெரிய குறைபாடாகும். அதாவது, தன்னிடம் ஒரு கேள்விக்கான தரவு இல்லாதபோது, AI தனக்குத் தெரிந்த தரவுகளைக் கொண்டு உண்மையாகத் தோற்றமளிக்கும் ஒரு பொய் பதிலை உருவாக்கிவிடும். ஆங்கிலத்தில் தரவுகள் அதிகம் என்பதால் அங்கு ஹாலுசினேஷன் விகிதம் குறைவு. ஆனால், தமிழில் வரலாற்றுத் தரவுகளோ அல்லது இலக்கியத் தரவுகளோ குறைவாக இருக்கும் பட்சத்தில், ஒரு குறிப்பிட்ட புலவரைப் பற்றியோ அல்லது சங்கப் பாடலைப் பற்றியோ கேட்கும்போது, AI முற்றிலும் தவறான தகவல்களையோ அல்லது இல்லாத ஒரு பாடலையோ தானே இயற்றி உண்மையான பாடம் போலக் காட்டும் ஆபத்து நேரிடுகிறது.

கலாச்சார மற்றும் சமூகச் சார்புகள் (Cultural and Social Biases)

உலகளாவிய AI மாதிரிகள் மேலைநாட்டுக் கலாச்சாரத் தரவுகளைக் கொண்டு வடிவமைக்கப்பட்டவை. தமிழ் சமூகத்தின் தனித்துவமான அறக்கோட்பாடுகள், அகப்பொருள்-புறப்பொருள் மரபுகள், மெய்ப்பாட்டியல் உணர்வுகள் போன்றவற்றை மேலைநாட்டு விழுமியங்களைக் கொண்டு AI அணுகும்போது, அது தமிழ்மொழி மற்றும் கலாச்சாரத்தின் ஆன்மாவைப் பிரதிபலிக்கத் தவறுகிறது.

பதிப்புரிமைத் தடைகள்

தமிழில் பல தரமான இலக்கியப் படைப்புகள், அகராதிகள் மற்றும் ஆராய்ச்சி நூல்கள் இன்னும் பதிப்புரிமைச் சட்டங்களுக்கு உட்பட்டே உள்ளன. இவை வணிகரீதியான புத்தகங்களாக மட்டுமே புழக்கத்தில் இருப்பதால், AI ஆராய்ச்சியாளர்கள் இந்தத் தரவுகளைத் தங்களின் மாதிரிகளுக்குப் பயிற்சியளிக்கப் பயன்படுத்த முடிவதில்லை. மறுபுறம், இணையத்தில் இலவசமாகக் கிடைக்கும் பல தமிழ் உள்ளடக்கங்கள் பிழைகள் நிறைந்தவையாகவும், தரம் குறைந்தவையாகவும் உள்ளன. தரமான புத்தகங்கள் டிஜிட்டல் வடிவில் ‘திறந்தமூலத் தரவுகளாக’ (**Open-source Corpora**) கிடைக்காத வரை, தமிழ் **AI**-இன் மொழியியல் துல்லியம் உலகத் தரத்தை எட்டுவது கடினம்.

வேலைவாய்ப்பு மற்றும் பொருளாதாரம்

செயற்கை நுண்ணறிவு வேலைவாய்ப்பில் இரு முகத்தைக் கொண்டது. ஒருபுறம் சில வழக்கமான பணிகளை தானியக்கமாக மாற்றுவதால் சில வேலைகள் குறைய வாய்ப்பு உள்ளது. மறுபுறம் புதிய வேலைகள், புதிய திறன்கள், புதிய தொழில்துறைகள் உருவாகின்றன. தமிழ்மொழிச் சார்ந்த உள்ளடக்கத் தயாரிப்பு, தரவு குறிச்சொல்லிடல், மொழித் திருத்தம், குரல் மாதிரி உருவாக்கம், AI பயிற்சி தரவுத் தயாரிப்பு போன்ற புதிய வாய்ப்புகள் உருவாகும். தமிழ்நாடு மற்றும் பிற தமிழ் பேசும் பகுதிகளில் உள்ள இளைஞர்கள் AI, இயந்திரக் கற்றல், மொழி தொழில்நுட்பம், தரவு அறிவியல் போன்ற துறைகளில் திறன் பெறுவது

Volume 1, SPL 1 – July 2026

அவசியமாகிறது. தமிழ்மொழிக்கான AI வளர்ச்சி, உள்ளூர் தொழில்நுட்ப நிறுவனங்கள், கல்வி நிறுவனங்கள், தொடக்கநிறுவனங்கள் ஆகியவற்றுக்குப் புதிய சந்தையை உருவாக்க முடியும். இது மொழிப் பாதுகாப்புடன் பொருளாதார வளர்ச்சியையும் இணைக்கும்.

எதிர்காலத் தொலைநோக்கும் சிக்கல்களுக்கான தீர்வுகளும்

தமிழ் AI எதிர்கொள்ளும் சவால்கள் என்பது வெறும் தொழில்நுட்பச் சிக்கல்கள் மட்டுமல்ல; அவை மொழியியல், பொருளாதாரம் மற்றும் கலாச்சாரம் சார்ந்த பன்முகத் தடைகளாகும். இத்தடைகளைத் தகர்க்கப் பின்வரும் கூட்டு முயற்சிகள் அவசியமாகின்றன. தமிழ்நாடு அரசு மற்றும் தமிழ்ப் பல்கலைக்கழகங்கள் இணைந்து பதிப்புரிமை முடிந்த ஆவணங்களையும், அகராதிகளையும் பொதுத் தரவுத்தளங்களாக (**Open Datasets**) மாற்ற வேண்டும். பொதுமக்கள் பங்களிப்பு - பல்வேறு வட்டார வழக்கிலமைந்த குரல் பதிவுகளைப் பொதுவெளியில் சேகரித்துப் பேச்சுத் தொழில்நுட்பத்தை (**Speech AI**) மேம்படுத்த வேண்டும். செயற்கை நுண்ணறிவு யுகத்தில் தமிழ்மொழி பின்தங்கிவிடாமல் இருக்க, இத்தகைய சவால்களைக் கண்டறிந்து அவற்றை அறிவியல் பூர்வமாக எதிர்கொள்வதே தமிழ் அறிவுச் சமூகத்தின் தற்போதைய முதன்மைக் கடமையாகும். தமிழ் AI எதிர்கொள்ளும் சவால்கள் எவ்வளவு பெரியவையாக இருந்தாலும், அவற்றைத் தகுந்த கணினியியல் உத்திகள், மொழியியல் சீரமைப்பு மற்றும் கூட்டுத் திட்டங்கள் மூலம் வெற்றிகரமாக எதிர்கொள்ள முடியும்.

நிறைவுரை

"யாமறிந்த மொழிகளிலே தமிழ்மொழி போல் இனிதாவது எங்கும் காணோம்" என்றார் பாரதியார். கல்வெட்டிலும் செப்பேட்டிலும் காகிதத்திலும் நிலைத்து நின்ற அந்த இனிமையான தமிழ்மொழி இனி கணினித் திரைகளிலும், சிலிக்கான் சிப்புகளிலும் செயற்கை நுண்ணறிவு உலகிலும் கம்பீரமாக நடைபோடத் தொடங்கியுள்ளது. மொழியின் கட்டமைப்பு என்பது ஒரு மென்பொருளின் வடிவமைப்பு போன்றது. தொல்காப்பியர் காலம் தொட்டே தமிழ்மொழி பெற்றுள்ள இந்த ஒட்டுநிலை இலக்கணக் கட்டமைப்பு, இன்று 21-ஆம் நூற்றாண்டின் செயற்கை நுண்ணறிவுத் தொழில்நுட்பத்திற்கு மிக இயல்பாகப் பொருந்திப் போகிறது. செயற்கை நுண்ணறிவு யுகத்தில் தமிழ்மொழி தொடர்ந்து தன் ஆதிக்கத்தைச் செலுத்த வேண்டுமெனில், தமிழ் NLP ஆராய்ச்சிகள் மேலும் பல மடங்கு பெருக வேண்டும். தமிழ் இணையக் கல்விக்கழகம் மற்றும் பல்வேறு தொழில்நுட்ப நிறுவனங்களின் கூட்டு முயற்சியால், தமிழ் NLP வேகமாக வளர்ந்து வருகிறது. நம்முடைய சங்க இலக்கியங்கள் முதல் நவீனச் செய்திகள் வரை அனைத்தையும் கணினியின் நரம்பியல் வலைப்பின்னல்களுக்குக் கற்றுக்கொடுப்பதன் மூலம், எதிர்காலத் தலைமுறைக்குத் தமிழை டிஜிட்டல் வடிவில் கொண்டு சேர்க்க முடியும். AI தமிழை உலகளாவிய டிஜிட்டல் வெளியிடத்திற்கு கொண்டு செல்லும்; தமிழ் AIக்கு மனிதநேயம், பண்பாடு, ஆழம் ஆகியவற்றைக் கொடுக்கும். செயற்கை நுண்ணறிவு என்பது தமிழுக்குக் கிடைத்த ஒரு வரப்பிரசாதம். இதன் மூலம் கால எல்லையைக் கடந்து நம் மொழியை அடுத்த தலைமுறைக்குக் கொண்டு சேர்க்க முடியும். தொழில்நுட்ப வளர்ச்சியால் தமிழ் அழியாது, மாறாக, செயற்கை நுண்ணறிவின் துணையோடு தமிழ் உலக அரங்கில் இன்னும் பல்லாயிரமாண்டுகள் மிக வீரியமாகத் தழைத்தோங்கும் என்பதில் ஐயமில்லை.

குறிப்புகள்

[1] சாமுவேல் மாத்யூ, கார்த்திகா குமாரி (மொ.பெ.ஆ.) Unnai அறிந்தால் ஏற்றம் தரும் AI, சிக்ஸ்த்சென்ஸ் பப்ளிகேஷன்ஸ், தியாகராயநகர், சென்னை- 600 017.

Volume 1, SPL 1 – July 2026

[2] முனைவர் இல.சுந்தரம், கணினித் தமிழ், விகடன் பிரசுரம், 757, அண்ணா சாலை, சென்னை – 600 002.

[3] த.சம்பத்குமார், கல்வியில் கணினி, பாவை பப்ளிகேஷன்ஸ் வெளியீடு, சென்னை – 600 014.

நிதிசார் கட்டுரையாளர் உறுதிமொழி: இல்லை

கட்டுரையாளர் நன்றியுரை: இல்லை

கட்டுரையாளர் உறுதிமொழி: இக்கட்டுரையில் எவ்வித முரண்பாடும் இல்லை என்று உறுதிமொழி அளிக்கிறேன்.

இக்கட்டுரை கிரியேட்டிவ் காமன்சு ஆட்ரிபியூசன் 4.0வின் **Creative Commons Attribution4.0** கீழ் பன்னாட்டு உரிமம் பெற்றுள்ளது.