

கணினியியலும் தமிழ் இலக்கியமும்

முனைவர் ம. கார்த்திகா, உதவிப் பேராசிரியர் & தலைவர், கணினி அறிவியல் துறை,
மன்னார் திருமலை நாயக்கர் கல்லூரி, பசுமலை, மதுரை, தமிழ்நாடு, இந்தியா.

ORCID: <https://orcid.org/0009-0004-5689-6835>

அறிமுகம்

கணினியியலும் தமிழ் இலக்கியமும், முதல் பார்வையில் ஒன்றுக்கொன்று முற்றிலும் வேறுபட்ட இரண்டு துறைகளாகத் தோன்றலாம். கணினியியல் என்பது அல்காரிதங்கள், தரவு கட்டமைப்புகள், மென்பொருள் வடிவமைப்பு, மற்றும் இயந்திர அறிவுநுட்பம் ஆகியவற்றைக் கையாளும் ஒரு தொழில்நுட்பத் துறையாகும். தமிழ் இலக்கியம் என்பது உணர்வுகள், அழகியல், வரலாறு, மற்றும் மனித அனுபவங்களைச் சொற்கள் வாயிலாக வெளிப்படுத்தும் ஒரு மனிதநேயத் துறையாகும். இருப்பினும், இருபத்தொன்றாம் நூற்றாண்டில், இந்த இரண்டு துறைகளும் ஆழமாகப் பின்னிப்பிணைந்து, "டிஜிட்டல் மனிதநேயம்" (Digital Humanities) என்ற புதிய கல்வியியல் பரிமாணத்தை உருவாக்கியுள்ளன. டிஜிட்டல் மனிதநேயம் என்பது கணினியியல் கருவிகள் மற்றும் முறைகளைப் பயன்படுத்தி மனிதநேயத் துறைகளை (இலக்கியம், வரலாறு, மொழியியல், கலை) ஆய்வு செய்யும் ஒரு இடைநிலைத் துறையாகும். இது பாரம்பரிய இலக்கிய ஆய்வு முறைகளை, கணினியியலின் தரவு பகுப்பாய்வு, இயந்திரக் கற்றல், மற்றும் காட்சிப்படுத்தல் திறன்களுடன் இணைக்கிறது. தமிழ் இலக்கியத்தைப் பொறுத்தவரை, கணினியியலின் பயன்பாடுகள் மிகவும் பரந்தவை. பண்டைய ஓலைச் சுவடிகளை டிஜிட்டல் மயமாக்குதல், இலக்கிய உரைகளைப் பகுப்பாய்வு செய்தல், ஆசிரியர் பாணி மற்றும் காலத்தை இனங்காணுதல், இலக்கியத் தரவுத்தளங்களை உருவாக்குதல், மற்றும் AI மூலம் புதிய இலக்கியப் படைப்புகளை உருவாக்குதல் ஆகியவை இவற்றுள் அடங்கும். இக்கட்டுரை, கணினியியலும் தமிழ் இலக்கியமும் எவ்வாறு ஒன்றையொன்று பாதிக்கின்றன, அவற்றின் ஒருங்கிணைப்பு என்ன பயன்களை வழங்குகிறது, மற்றும் இந்த ஒருங்கிணைப்பில் உள்ள சவால்கள் என்ன என்பதை ஒரு முழுமையான கல்வியியல் ஆய்விற்கு உட்படுத்துகிறது.

கணினியியல் மற்றும் தமிழ் இலக்கியத்தின் ஒருங்கிணைப்பின் வரலாற்றுப் பின்னணி

கணினியியலும் தமிழ் இலக்கியமும் ஒன்றிணைந்தது மிகச் சமீபத்திய காலத்தில்தான். இருபதாம் நூற்றாண்டின் இறுதியில், தமிழ் எழுத்துகளைக் கணினியில் பதிவு செய்யும் யுனிகோட் (Unicode) அமைப்பு அறிமுகப்படுத்தப்பட்டதே இந்த ஒருங்கிணைப்பின் முதல் படியாகும். யுனிகோட், தமிழ் எழுத்துகளுக்கு ஒரு தரப்படுத்தப்பட்ட குறியீட்டு முறையை வழங்கியதன் மூலம், தமிழ் இலக்கிய உரைகளை டிஜிட்டல் வடிவத்தில் சேமிக்கவும், பரிமாறிக்கொள்ளவும், பகுப்பாய்வு செய்யவும் உதவியது. அதைத் தொடர்ந்து, இணையம் மற்றும் டிஜிட்டல் நூலகங்களின் வளர்ச்சி, தமிழ் இலக்கியப் படைப்புகளை உலகம் முழுவதும் அணுகுவதற்கு ஒரு புதிய வழியைத் திறந்தது.

இருபத்தியோராம் நூற்றாண்டின் தொடக்கத்தில், AI மற்றும் இயற்கை மொழி செயலாக்கத்தின் (NLP) வளர்ச்சி, இந்த ஒருங்கிணைப்பை ஒரு புதிய உயரத்திற்குக் கொண்டு சென்றது. NLP திறன்கள், தமிழ் இலக்கிய உரைகளைத் தானாகவே பகுப்பாய்வு செய்யவும், அவற்றில் உள்ள வடிவங்களைக் கண்டறியவும், மற்றும் புதிய உரைகளை உருவாக்கவும் உதவியது. இன்று, கணினியியல் கருவிகள் தமிழ் இலக்கிய ஆய்வின் ஒரு தவிர்க்க முடியாத பகுதியாக மாறியுள்ளன.

தமிழ் இலக்கிய டிஜிட்டல் மயமாக்கல் மற்றும் பாதுகாப்பில் கணினியியலின் பயன்பாடுகள்

தமிழ் இலக்கியத்தில் கணினியியலின் மிக அடிப்படையான பயன்பாடு, இலக்கியப் படைப்புகளை டிஜிட்டல் மயமாக்குவதும் பாதுகாப்பதுமாகும். பல ஆயிரம் ஆண்டுகளாக ஓலைச் சுவடிகள், கையெழுத்துப் பிரதிகள், மற்றும் அரிய அச்ச நூல்களாகப் பாதுகாக்கப்பட்டு வந்த தமிழ் இலக்கியப் படைப்புகள், கணினியியல் தொழில்நுட்பங்கள் வாயிலாக டிஜிட்டல் வடிவத்திற்கு மாற்றப்படுகின்றன. இதற்கு ஒளியியல் எழுத்து அங்கீகாரம் (OCR), உயர்-தெளிவு ஸ்கேனிங், மற்றும் படமாக்கும் தொழில்நுட்பங்கள் பயன்படுத்தப்படுகின்றன. OCR தொழில்நுட்பம், பழைய ஓலைச் சுவடிகள் மற்றும் கையெழுத்துப் பிரதிகளில் உள்ள எழுத்துகளை இனங்கண்டு, அவற்றைக் கணினி உரையாக மாற்றுகிறது. இது, பண்டைய தமிழ் இலக்கியங்களை ஆய்வாளர்களுக்கு எளிதில் அணுகக் கூடியதாக மாற்றுகிறது. மேலும், டிஜிட்டல் மயமாக்கப்பட்ட படைப்புகள், அவற்றின் அசல் வடிவங்கள் பழுதடைந்தாலும், அவற்றின் உள்ளடக்கம் பாதுகாக்கப்படுவதை உறுதி செய்கின்றன. டிஜிட்டல் காப்பகங்கள் (Digital Archives) மற்றும் இலக்கியத் தரவுத்தளங்கள், இந்த டிஜிட்டல் படைப்புகளை ஒருங்கிணைந்த முறையில் சேமித்து, அவற்றைத் தேடவும், பகுப்பாய்வு செய்யவும் உதவுகின்றன.

இலக்கியப் பகுப்பாய்வு மற்றும் ஆராய்ச்சியில் கணினியியலின் பயன்பாடுகள் -

கணினியியலின் இயற்கை மொழி செயலாக்கம் (NLP) மற்றும் இயந்திரக் கற்றல் (Machine Learning) திறன்கள், தமிழ் இலக்கியப் பகுப்பாய்வில் புரட்சிகரமான மாற்றங்களை ஏற்படுத்தியுள்ளன. பாரம்பரியமாக, ஒரு ஆய்வாளர் ஒரு இலக்கியப் படைப்பைப் பகுப்பாய்வு செய்ய, அதை முழுமையாக வாசித்து, கையேடு முறையில் குறிப்புகளை எடுத்து, வடிவங்களைக் கண்டறிய வேண்டும். இது மிகவும் நேரம் எடுக்கும் ஒரு பணியாகும். ஆனால், கணினியியல் கருவிகள், நிமிடங்களில் ஆயிரக்கணக்கான பக்கங்களைப் பகுப்பாய்வு செய்யும் திறன் கொண்டவை.

இயற்கை மொழி செயலாக்கம் (NLP) கருவிகள், ஒரு இலக்கியப் படைப்பில் உள்ள சொற்களின் அதிர்வெண், அவற்றுக்கிடையேயான தொடர்புகள், மற்றும் சொற்பொருள் வடிவங்களைப் பகுப்பாய்வு செய்கின்றன. உதாரணமாக, ஒரு ஆய்வாளர் கம்பராமாயணத்தில் உள்ள காதல், வீரம், மற்றும் அறம் ஆகிய கருப்பொருள்கள் எவ்வாறு வெவ்வேறு பகுதிகளில் வெளிப்படுகின்றன என்பதை AI கருவிகள் மூலம் பகுப்பாய்வு செய்யலாம். மேலும், NLP கருவிகள், ஒரு இலக்கியப் படைப்பின் பேச்சு நடை, உரையாடல் அமைப்பு, மற்றும் விளக்க முறைகளைப் பகுப்பாய்வு செய்து, அந்தப் படைப்பின் கலை அழகியலைப் பற்றிய புதிய நுண்ணறிவுகளை வழங்குகின்றன.

இயந்திரக் கற்றல் (Machine Learning) வழிமுறைகள், ஒரு இலக்கியப் படைப்பின் ஆசிரியர் பாணி (Stylometry) மற்றும் காலத்தை இனங்காணும் திறன் கொண்டவை. ஒவ்வொரு ஆசிரியருக்கும் ஒரு தனித்துவமான எழுத்து நடை உண்டு. AI கருவிகள், ஒரு படைப்பின் சொல்லாடல், வாக்கிய அமைப்பு, சொற்களின் சராசரி நீளம், மற்றும் சில குறிப்பிட்ட சொற்களின் பயன்பாட்டு அதிர்வெண் ஆகியவற்றைப் பகுப்பாய்வு செய்து, அந்தப் படைப்பின் ஆசிரியர் யார் என்பதைத் தீர்மானிக்கின்றன. தமிழ் இலக்கியத்தில், பல பண்டைய நூல்களின் ஆசிரியர் குறித்து அறிஞர்களிடையே கருத்து வேறுபாடுகள் உள்ளன. AI-யின் ஆசிரியர் பாணி பகுப்பாய்வு, இந்தச் சர்ச்சைகளுக்கு ஒரு புதிய மற்றும் துல்லியமான பதிலை வழங்கும்.

இலக்கியத் தரவுத்தளங்கள் மற்றும் தேடல் அமைப்புகளில் கணினியியலின் பயன்பாடுகள் -

கணினியியலின் பெரும் தரவு மேலாண்மை (Big Data Management) மற்றும் தரவுத்தள வடிவமைப்பு (Database Design) திறன்கள், தமிழ் இலக்கியத்திற்கான ஒருங்கிணைந்த

தரவுத்தளங்களை உருவாக்க உதவுகின்றன. ஒரு முழுமையான தமிழ் இலக்கியத் தரவுத்தளம், பின்வரும் அம்சங்களை உள்ளடக்கியதாக இருக்கும்: ஒவ்வொரு படைப்பின் முழு உரை, அதன் ஆசிரியர், காலம், வகை, முக்கிய சொற்கள், கருப்பொருள்கள், மற்றும் பிரதிகளுக்கிடையேயான மாறுபாடுகள். இந்தத் தரவுத்தளங்கள், திறமையான தேடல் அமைப்புகளுடன் (Search Engines) இணைக்கப்படுகின்றன. ஒரு ஆய்வாளர் ஒரு குறிப்பிட்ட சொல், சொற்றொடர், அல்லது கருப்பொருளைத் தேட விரும்பினால், கணினியியல் அமைப்புகள் சில நொடிகளில் அந்தத் தகவலைத் தேர்ந்தெடுத்து வழங்கும். மேலும், இந்தத் தரவுத்தளங்கள், இலக்கியப் படைப்புகளுக்கிடையேயான தொடர்புகளைக் கண்டறியும் வசதியையும் வழங்குகின்றன. உதாரணமாக, ஒரு கவிஞரின் கவிதையில் உள்ள ஒரு உவமை, மற்றொரு கவிஞரின் கவிதையில் உள்ள உவமையுடன் எவ்வாறு தொடர்புடையது என்பதைத் தரவுத்தளம் கண்டறியும்.

காட்சிப்படுத்தல் மற்றும் அணுகல் கணினியியலின் பயன்பாடுகள்

கணினியியல் கருவிகள், தமிழ் இலக்கிய உரைகளைக் காட்சிப்படுத்தும் (Visualization) திறன் கொண்டவை. எடுத்துக்காட்டாக, ஒரு இலக்கியப் படைப்பில் உள்ள பாத்திரங்களுக்கிடையேயான உறவுகளை ஒரு வரைபடமாக (Network Graph) காட்சிப்படுத்த முடியும். அல்லது ஒரு காப்பியத்தில் காலப்போக்கில் ஏற்படும் கருப்பொருள் மாற்றங்களை ஒரு வரைவாக (Timeline) காட்சிப்படுத்த முடியும். இது, இலக்கிய ஆய்வாளர்களுக்கு ஒரு புதிய கண்ணோட்டத்தை வழங்குகிறது. மேலும், கணினியியல் கருவிகள், தமிழ் இலக்கியத்தை உலகம் முழுவதும் உள்ள மக்களுக்கு எளிதில் அணுகக் கூடியதாக மாற்றுகின்றன. டிஜிட்டல் நூலகங்கள், ஆன்லைன் காப்பகங்கள், மற்றும் மொபைல் பயன்பாடுகள், ஒரு பயனர் எங்கிருந்தும் தமிழ் இலக்கியத்தைப் படிக்கவும், கேட்கவும், பகுப்பாய்வு செய்யவும் உதவுகின்றன. இது, தமிழ் இலக்கியத்தை உலகளாவிய அளவில் அறியச் செய்யும் ஒரு சக்திவாய்ந்த கருவியாக அமைகிறது.

கணினியியல் மற்றும் தமிழ் இலக்கியத்தின் ஒருங்கிணைப்பில் உள்ள சவால்கள்

இந்த ஒருங்கிணைப்பு பல நன்மைகளைத் தந்தாலும், இது பல சவால்களையும் முன்வைக்கிறது. தரவுப் பற்றாக்குறை (Data Scarcity). கணினியியல் கருவிகள், குறிப்பாக AI மாதிரிகள், பயனுள்ள முடிவுகளைத் தருவதற்கு பிரம்மாண்டமான அளவிலான பயிற்சித் தரவுகள் தேவைப்படுகின்றன. தமிழில், டிஜிட்டல் மயமாக்கப்பட்ட இலக்கியத் தரவுகள் ஆங்கிலத்தை விட மிகவும் குறைவு. இது AI மாதிரிகளின் துல்லியத்தைப் பாதிக்கிறது. AI கருவிகளின் தரவுச் சார்பு (Data Bias). AI கருவிகள், அவை பயிற்சியளிக்கப்பட்ட தரவுகளில் உள்ள சார்புகளைப் பெற்றிருக்கும். தமிழ் இலக்கியத் தரவுகள் பெரும்பாலும் புகழ்பெற்ற, பெரிய இலக்கியப் படைப்புகளை மையமாகக் கொண்டவை. இதனால், சிறிய, கிராமப்புற, அல்லது புறக்கணிக்கப்பட்ட இலக்கியப் படைப்புகள் AI கருவிகளால் கவனிக்கப்படாமல் போகலாம். AI கருவிகளின் வரம்புகள் மற்றும் மனிதத் தலையீட்டின் அவசியம். AI கருவிகள் ஒரு ஓலைச் சுவடியின் எழுத்துகளை இனங்காணவும், ஒரு படைப்பின் ஆசிரியர் பாணியைப் பகுப்பாய்வு செய்யவும் முடியும்; ஆனால், அவை ஒரு இலக்கியப் படைப்பின் அழகியல், உணர்வுநயம், பண்பாட்டு முக்கியத்துவம், மற்றும் வரலாற்று நுணுக்கங்களை முழுமையாகப் புரிந்துகொள்ள முடியாது. இலக்கிய ஆய்வு என்பது வெறும் உரையைப் பகுப்பாய்வு செய்வது மட்டுமல்ல; அது அந்த உரையின் விளக்கம், வர்ணனை, மற்றும் கல்வியியல் மதிப்பீடு ஆகியவற்றையும் உள்ளடக்கியது. இந்தப் பணிகளுக்கு மனித ஆய்வாளர்களின் திறன், அனுபவம், மற்றும் உள்ளுணர்வு மிகவும் அவசியமானவை. படைப்புரிமை (Copyright) மற்றும் அறிவுசார் சொத்துரிமை. AI கருவிகள், தமிழ் இலக்கியப் படைப்புகளை டிஜிட்டல் மயமாக்கி,

அவற்றிலிருந்து புதிய படைப்புகளை உருவாக்கும் போது, படைப்புரிமை மீறல் ஏற்படாமல் இருக்க, உரிமையாளர்களின் அனுமதி பெறுவது அவசியம். மேலும், AI-யால் உருவாக்கப்பட்ட புதிய படைப்புகளின் உரிமை யாருக்கு சொந்தம் என்பது சட்டரீதியாகத் தெளிவற்ற ஒரு பிரச்சனையாகும். தொழில்நுட்பச் செலவு மற்றும் அணுகல். AI கருவிகளைச் செயல்படுத்துவதற்கு அதிக விலையுயர்ந்த கணினி உள்கட்டமைப்பு, மென்பொருட்கள், மற்றும் பயிற்சி பெற்ற ஊழியர்கள் தேவைப்படுகின்றன. பல சிறிய நூலகங்கள், அருங்காட்சியகங்கள், மற்றும் ஆய்வு நிறுவனங்கள் இந்தச் செலவுகளைச் சமாளிப்பதில் சிரமப்படுகின்றன.

கணினியியல் மற்றும் தமிழ் இலக்கியத்தின் எதிர்காலச் சாத்தியங்கள்

கணினியியலும் தமிழ் இலக்கியமும் ஒன்றிணைந்தால், எதிர்காலம் மிகவும் நம்பிக்கைக்குரியது. மேம்பட்ட AI மாதிரிகள், பழைய ஓலைச் சுவடிகளில் உள்ள மங்கிய, கிழிந்த, அல்லது அழிந்த எழுத்துகளை மீட்டெடுக்கும் (Restoration) திறனைப் பெறும். மேலும், AI கருவிகள், தமிழ் இலக்கியப் படைப்புகளைத் தானாகவே மொழிபெயர்க்கும் திறனைப் பெற்று, அவற்றை உலகளாவிய அளவில் அறியச் செய்யும். AI கருவிகள், தமிழ் இலக்கியப் படைப்புகளுக்கிடையேயான மறைந்த தொடர்புகளைக் கண்டறியும். உதாரணமாக, ஒரு கவிஞரின் கவிதையில் உள்ள ஒரு உவமை, மற்றொரு கவிஞரின் கவிதையில் உள்ள உவமையுடன் எவ்வாறு தொடர்புடையது என்பதை AI கண்டறியும். இது, தமிழ் இலக்கிய ஆய்வுக்கு ஒரு புதிய பரிமாணத்தைத் திறக்கும். கூடுதலாக, AI கருவிகள், தமிழ் இலக்கியத்தை விளையாட்டு மயமான (Gamified) முறையில் கற்பிக்கும் திறனைப் பெறும். மாணவர்கள், AI-யால் உருவாக்கப்பட்ட ஒரு கற்பனைத் தமிழ் இலக்கியச் சூழலில், பழைய கவிஞர்களுடன் உரையாடவும், புதிய கவிதைகளை எழுதவும், இலக்கியப் புதிர்களைத் தீர்க்கவும் முடியும். இது, தமிழ் இலக்கியத்தை இளைய தலைமுறையினருக்கு மிகவும் சுவாரஸ்யமானதாக மாற்றும்.

முடிவுரை

கணினியியலும் தமிழ் இலக்கியமும், முதல் பார்வையில் வேறுபட்ட துறைகளாகத் தோன்றினாலும், அவை ஒன்றையொன்று ஆழமாகப் பூர்த்தி செய்யும் இரண்டு பரிமாணங்களாகும். கணினியியல் கருவிகள், தமிழ் இலக்கியத்தை டிஜிட்டல் மயமாக்கவும், பாதுகாக்கவும், பகுப்பாய்வு செய்யவும், மற்றும் புதிய படைப்புகளை உருவாக்கவும் உதவுகின்றன. அவை இலக்கிய ஆய்வுக்கு ஒரு புதிய, தரவு சார்ந்த பரிமாணத்தைத் திறக்கின்றன. அதே சமயம், தமிழ் இலக்கியம், கணினியியல் கருவிகளுக்கு ஒரு பண்பாட்டு மற்றும் உணர்வு சார் ஆழத்தை வழங்குகிறது. இந்த இரண்டு துறைகளின் ஒருங்கிணைப்பு, டிஜிட்டல் மனிதநேயம் என்னும் புதிய கல்வியியல் துறையின் வளர்ச்சிக்கு வழிவகுக்கிறது. இருப்பினும், இந்த ஒருங்கிணைப்பில் தரவுப் பற்றாக்குறை, AI சார்பு, நெறிமுறைச் சிக்கல்கள், மற்றும் தொழில்நுட்பச் செலவு போன்ற சவால்களைச் சமாளிப்பது அவசியமாகும். கணினியியல் கருவிகள் ஒரு துணைக் கருவியாகவே பயன்படுத்தப்பட வேண்டும்; அவை மனித ஆய்வாளர்களின் அறிவு, அனுபவம், மற்றும் உள்ளுணர்வுக்கு மாற்றாக அல்ல. கணினியியலும், தமிழ் இலக்கியமும் ஒன்றையொன்று ஊக்குவிக்கும் ஒரு இணைவு உறவைப் பேணுவதன் மூலமே, இந்தத் துறையின் முழுமையான பயன்களை அடைய முடியும். AI-துணையுடனான தமிழ் இலக்கிய ஆய்வு, தமிழ் மொழி மற்றும் இலக்கியத்தின் பாரம்பரியத்தைப் பாதுகாப்பதோடு, அதனை உலகளாவிய அளவில் அறியச் செய்யும் ஒரு சக்திவாய்ந்த கல்வியியல் கருவியாக அமையும்.

குறிப்புகள்

- [1] இராமசாமி, நா. (2021). கணினியியலும் இலக்கியமும்: ஒரு இடைநிலை அணுகுமுறை. கோவை: நியூ செஞ்சுரி புத்தக நிலையம்.
- [2] சிவக்குமார், அ. (2022). தமிழ் இலக்கியத் தரவுத்தளங்கள்: வடிவமைப்பும் மேலாண்மையும். சென்னை: பாரதி பதிப்பகம்.
- [3] பாலகிருஷ்ணன், ரா. (2021). டிஜிட்டல் மயமாக்கப்பட்ட தமிழ் இலக்கியங்கள்: பாதுகாப்பும் அணுகலும். கோவை: கணினி மொழி வெளியீடு.
- [4] மோரெட்டி, எஃப். (2021). தொலைநோக்கு வாசிப்பு: இலக்கிய ஆய்வில் புதிய முறைகள். மொழிபெயர்ப்பு: வெங்கடேசன், ரா. சென்னை: பாரதி பதிப்பகம்.
- [5] ஜாக்கர்ஸ், எம். எல். (2022). இலக்கிய ஆய்வுக்கான R-மொழி உரைப் பகுப்பாய்வு. மொழிபெயர்ப்பு: குமரேசன், பி. கோவை: கணினி மொழி வெளியீடு).

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](https://creativecommons.org/licenses/by/4.0/) கீழ் பன்னாட்டு உரிமம் பெற்றுள்ளது.