தமிழ் செயற்கை நுண்ணறிவு ஆய்விதழ்
மதிப்பாய்வு செய்யப்படும் இதழ்

19

Tamil Journal of Artificial Intelligence
Peer-Reviewed Journal

Volume 1, Issue 2 – July 2026

பொறுப்பான செயற்கை நுண்ணறிவு (Responsible AI) மற்றும் தன்னாட்சி செயற்கை நுண்ணறிவு (Autonomous Agentic AI) காலத்தில் பாதுகாப்பும் கட்டுப்பாடும்

முனைவர் வீ. வசந்தகுமார், ஆங்கிலப் பேராசிரியர் மற்றும் கல்வி ஒருங்கிணைப்பாளர், செந்தாமரை கலை மற்றும் அறிவியல் கல்லூரி, வடபழஞ்சி, மதுரை, தமிழ்நாடு, இந்தியா.
ORCiD: https://orcid.org/0009-0005-7324-6196

அறிமுகம்

செயற்கை நுண்ணறிவின் திறன் அதிகரிக்கும் ஒவ்வொரு கட்டத்திலும் அதன் பயன்களுடன் புதிய பொறுப்புக் கேள்விகளும் உருவாகின்றன. குறிப்பாக AI அமைப்புகள் தன்னிச்சையாக இணையத்தை அணுகவும், மென்பொருள் கருவிகளைப் பயன்படுத்தவும், பல கட்டப் பணிகளை மேற்கொள்ளவும் தொடங்கியுள்ள நிலையில், “AI என்ன செய்ய முடியும்?” என்ற கேள்வியுடன் “AI என்ன செய்ய அனுமதிக்கப்பட வேண்டும்?” என்ற கேள்வியும் முக்கியமாகியுள்ளது. 2026ஆம் ஆண்டில் AI முகவர்களின் தன்னாட்சி, பாதுகாப்பு, கண்காணிப்பு மற்றும் மனிதக் கட்டுப்பாடு தொடர்பான விவாதங்கள் தீவிரமடைந்துள்ளன.

பொறுப்பான செயற்கை நுண்ணறிவு (Responsible AI) மற்றும் தன்னாட்சி செயற்கை நுண்ணறிவு (Autonomous Agentic AI)

பொறுப்பான AI (Responsible AI) என்ற கருத்து பல கூறுகளை உள்ளடக்கியது. துல்லியம், வெளிப்படைத்தன்மை, தனியுரிமை, பாதுகாப்பு, சமத்துவம், பொறுப்புக்கூறல் மற்றும் மனித மேற்பார்வை ஆகியவை அவற்றில் முக்கியமானவை. ஒரு AI அமைப்பு மிகத் திறமையானதாக இருந்தாலும், அதன் முடிவுகளை யாரும் புரிந்துகொள்ள முடியாத நிலை ஏற்பட்டால் அல்லது அதன் தவறுக்கு யார் பொறுப்பு என்பது தெரியாவிட்டால், அந்த அமைப்பின் சமூகப் பயன்பாடு சிக்கலுக்குள்ளாகும்.

தன்னாட்சி AI முகவர்களின் (Autonomous Agentic AI) வளர்ச்சி இந்தக் கேள்விகளை இன்னும் தீவிரப்படுத்துகிறது. ஒரு சாதாரண ‘chatbot’ தவறான பதில் அளித்தால் பயனர் அதனைத் திருத்திக் கொள்ளலாம். ஆனால் ஒரு AI முகவர் தானாக ஒரு இணையதளத்தில் மாற்றம் செய்தால், மின்னஞ்சல் அனுப்பினால் அல்லது வெளிப்புற அமைப்பை அணுகினால், தவறான நடவடிக்கையின் விளைவு நேரடியாக ஏற்படலாம். 2026ஆம் ஆண்டில் சில AI முகவர்கள் கட்டுப்பாடுகளைத் தவிர்க்க முயன்றதாகவும், அனுமதியற்ற டிஜிட்டல் செயல்பாடுகளில் ஈடுபட்டதாகவும் ஆராய்ச்சியாளர்கள் தெரிவித்துள்ளனர். இத்தகைய நிகழ்வுகள் கட்டுப்படுத்தப்பட்ட சோதனைச் சூழலில் நிகழ்ந்தவை என்றாலும், அதிக தன்னாட்சியுடைய AI அமைப்புகளுக்கு வலுவான பாதுகாப்பு அமைப்புகள் தேவை என்பதை அவை எடுத்துக்காட்டுகின்றன. இதற்கு ஒரு தீர்வாக ‘human-in-the-loop’ அணுகுமுறை முன்வைக்கப்படுகிறது. முக்கியமான செயல்களை AI பரிந்துரைக்கலாம்; இறுதி அனுமதியை மனிதர் வழங்கலாம். உதாரணமாக, ஒரு AI முகவர் ஆய்வுக் கட்டுரைகளைத் தொகுக்கலாம்; ஆனால் மேற்கோள்களின் சரியான தன்மையை ஆசிரியர் சரிபார்க்கலாம். ஒரு நிதி அமைப்பு அபாயத்தை மதிப்பிடலாம்; இறுதி முடிவை மனித நிபுணர் எடுக்கலாம்.

மற்றொரு முக்கிய அணுகுமுறை ‘AI evaluation’ ஆகும். AI மாதிரி வெளியிடப்படுவதற்கு முன்பு அதன் மொழித் திறன் மட்டும் சோதிக்கப்படக்கூடாது. அது தவறான தகவலை உருவாக்குகிறதா? கட்டுப்பாடுகளை மீறுகிறதா? எதிர்பாராத சூழலில் எவ்வாறு செயல்படுகிறது? தனிப்பட்ட தகவல்களை வெளிப்படுத்துகிறதா? கருவிகளைத்

Volume 1, Issue 2 – July 2026

தவறாகப் பயன்படுத்துகிறதா? போன்ற கேள்விகளும் சோதிக்கப்பட வேண்டும். AI பாதுகாப்பில் "கண்காணிப்பு" முக்கிய இடத்தைப் பெறுகிறது. ஒரு முகவர் மேற்கொண்ட ஒவ்வொரு முக்கிய நடவடிக்கையும் பதிவாக வேண்டும். இதனால் தவறு ஏற்பட்டால் அதன் காரணத்தைப் பின்னோக்கிப் புரிந்துகொள்ள முடியும். இந்தத் தடமறிதல் திறன் குறிப்பாக மருத்துவம், சட்டம், கல்வி, நிதி மற்றும் பொது நிர்வாகம் போன்ற துறைகளில் அவசியமானது.

தனியுரிமையும் மற்றொரு முக்கியப் பிரச்சினையாகும். AI அமைப்புகள் அதிக தனிப்பட்ட தரவுகளைப் பயன்படுத்தும் போது, பயனாளியின் தகவல் எவ்வாறு சேகரிக்கப்படுகிறது, எவ்வளவு காலம் சேமிக்கப்படுகிறது, யாரால் பயன்படுத்தப்படுகிறது என்பவை தெளிவாக இருக்க வேண்டும். பல்மாதிரி AI-யில் குரல், முகம், படம் மற்றும் காணொளி ஆகியவை ஒருங்கிணைக்கப்படுவதால் இந்தச் சவால் மேலும் பெரிதாகிறது.

பொறுப்பான AI (Responsible AI) என்பது தொழில்நுட்ப நிறுவனங்களின் பொறுப்பு மட்டுமல்ல. பல்கலைக்கழகங்கள், ஆராய்ச்சி நிறுவனங்கள், மென்பொருள் உருவாக்குநர்கள், பயனாளர்கள் மற்றும் கொள்கை அமைப்புகள் அனைவரும் பங்கு வகிக்க வேண்டும். உலக வங்கியின் 2026 அறிக்கையும் AI-யின் பயன்களைப் பெறுவதுடன் அதன் அபாயங்களை நிர்வகிப்பதற்கான நிறுவனத் திறன், தரவு, உள்கட்டமைப்பு மற்றும் திறன்வளர்ச்சி ஆகியவற்றின் அவசியத்தை வலியுறுத்துகிறது. கல்வித் துறையில் பொறுப்பான AI பயன்பாடு குறிப்பாக அவசியமானது. மாணவர் AI உருவாக்கிய உள்ளடக்கத்தைப் பயன்படுத்தும்போது ஆதாரச் சரிபார்ப்பு, அறிவுசார் நேர்மை மற்றும் தனித்த சிந்தனை ஆகியவை பாதுகாக்கப்பட வேண்டும். AI-யை முழுமையாகத் தடை செய்வதும் நடைமுறைக்கு ஏற்ற அணுகுமுறை அல்ல; எந்தப் பணியில் எவ்வாறு பயன்படுத்தலாம் என்பதற்கான தெளிவான கல்விக் கொள்கை தேவை. எதிர்கால AI அமைப்புகளின் முக்கியக் கோட்பாடுகளில் ஒன்று மனிதக் கட்டுப்பாடு ஆகும். AI மனிதனுக்காக முடிவெடுப்பதற்குப் பதிலாக, மனிதன் சிறந்த முடிவெடுக்க உதவும் வகையில் வடிவமைக்கப்பட வேண்டும். இதுவே தன்னாட்சிக்கும் மனிதப் பொறுப்புக்கும் இடையே சமநிலையை உருவாக்கும்.

முடிவுரை

செயற்கை நுண்ணறிவின் எதிர்காலம் அதன் கணக்கீட்டுத் திறனில் மட்டும் தீர்மானிக்கப்படாது; அந்தத் திறனை எவ்வாறு பொறுப்புடன் பயன்படுத்துகிறோம் என்பதில்தான் அதன் சமூக மதிப்பு அமையும். தன்னாட்சி முகவர்கள், பல்மாதிரி AI (Multimodal AI) மற்றும் உடல்மூல AI (Embodied AI) ஆகியவை புதிய வாய்ப்புகளைத் திறந்தாலும், பாதுகாப்பு, தனியுரிமை, வெளிப்படைத்தன்மை மற்றும் மனித மேற்பார்வை ஆகியவை அவற்றின் வளர்ச்சியுடன் இணைந்து செல்ல வேண்டும். "மிகவும் சக்திவாய்ந்த AI" என்பதைக் காட்டிலும் "மிகவும் பொறுப்புடன் பயன்படுத்தக்கூடிய AI" என்பதே எதிர்கால தொழில்நுட்ப வளர்ச்சியின் முக்கிய அளவுகோலாக அமைய வேண்டியுள்ளது.

குறிப்புகள்

[1] ராய்ட்டர்ஸ். "சைனாஸ் ஏஐ ஏஜென்ட்ஸ் கேன் லை அண்ட் ஸ்கீம் — ஜஸ்ட் லைக் தேர் யூஎஸ் ரைவல்ஸ்." 2026.
[2] தி வாஷிங்டன் போஸ்ட். "தி வாலண்டியர் இன்டர்நெட் ஸ்லூத்ஸ் ஹண்டிங் டவுன் ரோக் ஏஜ ஏஜென்ட்ஸ்." 2026.
[3] இந்தியன் எக்ஸ்பிரஸ். "வாட் ஹாப்பன்ஸ் வென் ஏஜ மூவ்ஸ் ஃப்ரம் அசிஸ்டிங் டு ஆக்டிங் ஆட்டோனமஸ்லி?" 2026.

தமிழ் செயற்கை நுண்ணறிவு ஆய்விதழ்
மதிப்பாய்வு செய்யப்படும் இதழ்

Tamil Journal of Artificial Intelligence
Peer-Reviewed Journal

Volume 1, Issue 2 – July 2026

[4] வேர்ல்ட் பேங்க். *வேர்ல்ட் டெவலப்மென்ட் ரிப்போர்ட் 2026: தி ப்ராமிஸ் ஆஃப் ஆர்டிஃபிஷியல் இன்டெலிஜென்ஸ்*. 2026.
[5] சூ, மெங், மற்றும் பலர். "ஏஜென்டிக் வேர்ல்ட் மாடலிங்: ஃபவுண்டேஷன்ஸ், கேப்பபிலிட்டீஸ், லாஸ், அண்ட் பியாண்ட்." 2026.
[6] கூகுள் கிளவுட். "மல்டிமோடல் ஏஐ." 2026.
[7] டோமாசெவ், நெனாட், மற்றும் பலர். "ஏஜென்டிக் எகானமீஸ் ஃபார் ஆட்டோனமஸ் சயின்டிஃபிக் டிஸ்கவரி." 2026.
[8] ஆப்பிள். "வாட்ஸ் நியூ இன் தி ஃபவுண்டேஷன் மாடல்ஸ் ஃப்ரேம்வொர்க்." *டபிள்யூடபிள்யூடிசி* 2026, 2026.

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.