മലയാളം റിസേർച്ച് കീ | MALAYALAM RESEARCH KEY | ONLINE EDITION
2026 APRIL-JUNE | VOLUME 02 | ISSUE 02
MULTIDISCIPLINARY PEER-REVIEWED RESEARCH JOURNAL (QUARTERLY)

Journal Home Page: www.researchkey.in | Email: mail@researchkey.in
Published on April 01, 2026 | Time: 10:00 AM IST | e-ISSN:

Article ID:MRK0027

നിർമ്മിതബുദ്ധിയും നൈതികതയും: അൽഗോരിതങ്ങളിലെ പക്ഷപാതങ്ങളും വെല്ലുവിളികളും

ഡോ. ഷിംജിത്ത് എസ്. പി.

പ്രബന്ധസംഗ്രഹം

വിവരസാങ്കേതികവിദ്യയുടെ ചരിത്രത്തിലെ ഏറ്റവും നിർണ്ണായകമായ വഴിത്തിരിവാണ് നിർമ്മിതബുദ്ധി. എന്നാൽ, മനുഷ്യബുദ്ധിയെ അനുകരിക്കാൻ ശ്രമിക്കുന്ന ഈ സാങ്കേതികവിദ്യ സങ്കീർണ്ണമായ നിരവധി നൈതികപ്രശ്നങ്ങൾ ഉയർത്തുന്നുണ്ട്. അൽഗോരിതങ്ങൾ സാമൂഹികവിവേചനത്തിന് എങ്ങനെ ആയുധമാകുന്നു എന്ന് വിശദമാക്കുന്ന കാത്തിഓനീലിന്റെ 'ഗണിതശാസ്ത്രപരമായ നശീകരണആയുധങ്ങൾ' എന്ന ആശയവും, നിർമ്മിതബുദ്ധി മനുഷ്യരാശിക്ക് ഉയർത്തുന്ന അസ്തിത്വഭീഷണിയെക്കുറിച്ചുള്ള നിക്ബോസ്ട്രോമിന്റെ 'സൂപ്പർഇന്റലിജൻസ്' സിദ്ധാന്തവും മുൻനിർത്തി, സാങ്കേതികവിദ്യയുടെ വർത്തമാനകാലദോഷങ്ങളെയും ഭാവിയിലെ വെല്ലുവിളികളെയും വിശകലനംചെയ്യുന്ന ഒരു താരതമ്യപഠനമാണ് ഈ ലേഖനം. അൽഗോരിതങ്ങൾ എങ്ങനെയാണ് സാമൂഹിക അസമത്വങ്ങളെയും വംശീയ-ലിംഗപരമായ മുൻവിധികളെയും പുനരുൽപ്പാദിപ്പിക്കുന്നതെന്നും, നിർമ്മിതബുദ്ധി മനുഷ്യന്റെ ലക്ഷ്യങ്ങൾക്കും മൂല്യങ്ങൾക്കും ധാർമ്മികതയ്ക്കും അനുസൃതമായി പ്രവർത്തിക്കുന്നുവെന്ന് ഉറപ്പുവരുത്തുന്ന എ.ഐ. അലൈൻമെന്റ് പ്രക്രിയയെയും ഈ പഠനത്തിലൂടെ വിശകലനം ചെയ്യുന്നു.

താക്കോൽവാക്കുകൾ

നിർമ്മിതബുദ്ധി, അൽഗോരിതം, എ.ഐ. നൈതികത, അൽഗോരിതമിക്പക്ഷപാതം.

.....

ആമുഖം

മനുഷ്യർ നിർമ്മിക്കുന്ന യന്ത്രങ്ങൾ മനുഷ്യരെക്കാൾ മിടുക്കരായാൽ എന്ത് സംഭവിക്കുമെന്നത് ഇരുപത്തിയൊന്നാം നൂറ്റാണ്ടിലെ തത്വശാസ്ത്രപരവും സാങ്കേതികവുമായ ഏറ്റവും വലിയ ചോദ്യമാണ്. നിർമ്മിതബുദ്ധി സമകാലികജീവിതത്തിൽ മാനവരാശിയുടെ തീരുമാനങ്ങളെ വളരെയധികം സ്വാധീനിക്കുന്നു. നാം എന്തുകാണണമെന്നും ആരെ വിവാഹംകഴിക്കണമെന്നും ആർക്ക് വായ്പ നൽകണമെന്നുമൊക്കെ തീരുമാനിക്കുന്നത് നിർമ്മിതബുദ്ധിയാൽ നിയന്ത്രിക്കപ്പെടുന്ന അൽഗോരിതങ്ങളാണ്.

എ.ഐ. സംവിധാനങ്ങൾക്ക് സ്വന്തമായി ബോധമോ ആത്മാവോ ഇല്ലെങ്കിലും, അവ മനുഷ്യന്റെ പെരുമാറ്റങ്ങളെയും മൂല്യങ്ങളെയും സമർത്ഥമായി അനുകരിക്കുന്നു. ഈ അനുകരണത്തിൽ സംഭവിക്കുന്ന പിഴവുകളും, ഡാറ്റയിൽ ഒളിഞ്ഞിരിക്കുന്ന മുൻവിധികളും എ.ഐ.യെ ഒരു 'ഡിജിറ്റൽകണ്ണാടി'യാക്കി മാറ്റുന്നുണ്ട്. ഈ ലേഖനം നിർമ്മിതബുദ്ധി ഉയർത്തുന്ന ധാർമ്മികവും സാമൂഹികവുമായ പ്രത്യാഘാതങ്ങളെക്കുറിച്ചാണ് പ്രധാനമായും ചർച്ച ചെയ്യുന്നത്.

ഈ പഠനം പ്രധാനമായും വിശകലനാത്മകരീതിശാസ്ത്രമാണ് പിന്തുടരുന്നത്. നിർമ്മിതബുദ്ധിയുടെ നൈതികതയുമായി ബന്ധപ്പെട്ട നിലവിലുള്ള സിദ്ധാന്തങ്ങളെയും സാമൂഹികസാഹചര്യങ്ങളെയും സമന്വയിപ്പിച്ചുകൊണ്ടുള്ള ഒരു മൾട്ടി-ഡിസിപ്ലിനറി സമീപനമാണ് ഇവിടെ സ്വീകരിച്ചിരിക്കുന്നത്. സാങ്കേതികവിദ്യ മനുഷ്യന്റെ മൂല്യബോധത്തെയും സ്വത്വത്തെയും എങ്ങനെ ബാധിക്കുന്നു എന്നതിന് ഇവിടെ മുൻഗണനനൽകുന്നു. കാത്തി ഓ നീൽ(Cathy O'Neil), സ്റ്റുവർട്ട്റസ്സൽ(Stuart Russell) നിക്ബോസ്ട്രോം(Nick Bostrom) തുടങ്ങിയവരുടെ വിഖ്യാതമായ കൃതികളെയും പഠനങ്ങളെയും ആസ്പദമാക്കി ഒരു താരതമ്യവിശകലനം ഈ പ്രബന്ധത്തിൽ നടത്തുന്നു. സിദ്ധാന്തങ്ങൾ പ്രായോഗികതലത്തിൽ എങ്ങനെ പ്രവർത്തിക്കുന്നു എന്ന് പരിശോധിക്കാൻ ആമസോണിലെ റിക്രൂട്ട്മെന്റ് പക്ഷപാതം, ബ്യൂട്ടിആപ്പുകളിലെ അൽഗോരിതമിക് വിവേചനം തുടങ്ങിയ യഥാർത്ഥ സംഭവങ്ങളെ കേസ് സ്റ്റഡികളായി സ്വീകരിച്ചിരിക്കുന്നു. ആഗോളതലത്തിലുള്ള എ.ഐ. നൈതികതയെ കേരളത്തിലെയും ഇന്ത്യയിലെയും സാമൂഹികസാംസ്കാരിക പശ്ചാത്തലത്തിലേക്ക് സമന്വയിച്ചുകൊണ്ടുള്ള ഒരു വിശകലനരീതിയും ഇതിൽ ഉൾപ്പെടുത്തിയിട്ടുണ്ട്.

എ.ഐ.യുടെ അസ്തിത്വപരമായ പ്രശ്നങ്ങൾ: ബോധവും അനുകരണവും

നിർമ്മിതബുദ്ധിയെക്കുറിച്ചുള്ള പ്രശസ്തമായ ചിന്താപരീക്ഷണങ്ങളിലൊന്നാണ് 1980-ൽ തത്വശാസ്ത്രജ്ഞനായ ജോൺ സേൾ അവതരിപ്പിച്ച 'ചൈനീസ് റൂം വാദം' നിർമ്മിതബുദ്ധിക്ക് നിയമാനുസൃതമായി വിവരങ്ങൾ കൈകാര്യം ചെയ്യാനുള്ള സാങ്കേതികശേഷി മാത്രമേയുള്ളൂവെന്നും, ആശയങ്ങളെ യഥാർത്ഥമായി ഉൾക്കൊള്ളാനും മനസ്സിലാക്കാനുമുള്ള കഴിവ് ഇല്ലെന്നും ഈ വാദം സ്ഥാപിക്കുന്നു. ഇതു പ്രകാരം, കോടിക്കണക്കിന് ഡാറ്റയും അൽഗോരിതങ്ങളും ഉപയോഗിച്ച് എ.ഐ മറുപടി നൽകുന്നുണ്ടെങ്കിലും, അവയ്ക്ക് ആ വാക്കുകളുടെ അർത്ഥമോ

മലയാളം റിസേർച്ച് കീ | MALAYALAM RESEARCH KEY | ONLINE EDITION
2026 APRIL-JUNE | VOLUME 02 | ISSUE 02
MULTIDISCIPLINARY PEER-REVIEWED RESEARCH JOURNAL (QUARTERLY)

Journal Home Page: www.researchkey.in | Email: mail@researchkey.in
Published on April 01, 2026 | Time: 10:00 AM IST | e-ISSN:

അനുഭവമോ ഇല്ല. ഒരു കമ്പ്യൂട്ടറിന് ഭാഷ കൈകാര്യംചെയ്യാൻ കഴിഞ്ഞേക്കാമെങ്കിലും അതിന് ഭാഷയുടെ 'അർത്ഥം' അറിയില്ല. എ.ഐ.ക്ക് 'ബോധം' ഇല്ലാത്തതുകൊണ്ട് അതിന് മൂല്യങ്ങളും ഉണ്ടാകില്ല. മനുഷ്യൻ നൽകുന്ന ഡാറ്റയിലുള്ള മൂല്യങ്ങൾ അത് അനുകരിക്കുക മാത്രമാണ് ചെയ്യുന്നത്. അതുകൊണ്ടാണ് സമൂഹം എ.ഐ.യെ മനുഷ്യന്റെ മൂല്യങ്ങൾക്കനുസരിച്ച് കൃത്യമായി ചിട്ടപ്പെടുത്തണം എന്നുപറയുന്നത്.

അനുഭവങ്ങളുടെ അഭാവം: 'അനുഭവം' എന്ന ഘടകത്തിന്റെ തലത്തിലാണ് മനുഷ്യനും എ.ഐ.യും തമ്മിലുള്ള ഏറ്റവും വലിയ വ്യത്യാസം. മനുഷ്യൻ നന്മയും തിന്മയും തിരിച്ചറിയുന്നത് സാമൂഹിക ഇടപെടലുകളിലൂടെയും വേദന/സന്തോഷം തുടങ്ങിയ അനുഭവങ്ങളിലൂടെയുമാണ്. എന്നാൽ എ.ഐ. പഠിക്കുന്നത് 'പാറ്റേണുകളിൽ'നിന്നാണ്. ഉദാഹരണമായി, ഒരു കുട്ടി തീയിൽ തൊടുമ്പോൾ വേദന അനുഭവിക്കുന്നു. 'വേദന' എന്ന അനുഭവംവഴി "തീ പൊള്ളിക്കും, അത് അപകടമാണ്" എന്ന് കുട്ടി പഠിക്കുന്നു. പിന്നീട് മറ്റൊരാൾക്ക് പൊള്ളലേൽക്കുന്നത് കാണുമ്പോൾ, സ്വന്തം അനുഭവം വെച്ച് ആ വ്യക്തിയുടെ സങ്കടം മനസ്സിലാക്കാനും സഹായിക്കാനും കഴിയുന്നു. എന്നാൽ എ.ഐ.ക്ക് 'വേദന' അറിയില്ല. "തീ പൊള്ളിക്കും" എന്ന ലക്ഷക്കണക്കിന് ഡാറ്റയിൽനിന്ന് തീ പൊള്ളിക്കും എന്നത് ഒരു വിവരം മാത്രമായിട്ടാണ് എ.ഐ ശേഖരിക്കുന്നത്. തീയെക്കുറിച്ച് വിവരിക്കാൻ അതിന് കഴിയുമെങ്കിലും, പൊള്ളലിന്റെ സങ്കടം അതിന് ഒരിക്കലും മനസ്സിലാകില്ല.

യാന്ത്രികമായി അനുകരിക്കുന്ന തത്തകൾ: നിർമ്മിതബുദ്ധിയുമായി ബന്ധപ്പെട്ട് എമിലി എം. ബെൻഡർ, ടിംനിറ്റ് ഗെബ്രു എന്നിവർ മുന്നോട്ടുവെച്ച ശ്രദ്ധേയമായ ആശയമാണ് 'യാന്ത്രികമായി അനുകരിക്കുന്ന തത്തകൾ' (Stochastic Parrots). നിർമ്മിതബുദ്ധി ഭാഷയുടെ അർത്ഥം പൂർണ്ണമായി മനസ്സിലാക്കിക്കൊണ്ടല്ല പ്രതികരിക്കുന്നത്. അത് തങ്ങൾക്ക് പരിശീലിക്കാൻ ലഭിച്ച ഡാറ്റയിലെ പാറ്റേണുകൾ വെറുതെ ആവർത്തിക്കുക മാത്രമാണ് ചെയ്യുന്നത്. അർത്ഥം എന്താണെന്നറിയാതെ 'തത്തമ്മേ പൂച്ച പൂച്ച' എന്ന് നിരന്തരം അനുകരിക്കുന്നതിന് സമാനമായ യാന്ത്രികമായ ഒരു പ്രക്രിയ മാത്രമാണിത്.

ലാർജ് ലാഗ്വേജ്മോഡലുകൾ ഭാഷയെ അർത്ഥബോധത്തോടെയല്ല കൈകാര്യം ചെയ്യുന്നത്. അവ ടില്യൺ കണക്കിന് ഡാറ്റാപോയിന്റുകളിൽനിന്ന് ഒരു വാക്കിനുശേഷം സാധ്യതയുള്ള അടുത്ത വാക്ക് ഗണിതശാസ്ത്രപരമായി പ്രവചിക്കുകയാണ് ചെയ്യുന്നത്. ഒരു തത്ത വാക്കുകളുടെ അർത്ഥമറിയാതെ ആവർത്തിക്കുന്നപോലെ, എ.ഐ.യും ഡാറ്റയിലെ 'പാറ്റേണുകൾ' മാത്രം ആവർത്തിക്കുന്നു. ഇതിനെയാണ് 'യാന്ത്രികമായി അനുകരിക്കുന്ന തത്തകൾ' എന്ന് വിളിക്കുന്നത്.

എ.ഐ.ക്ക് സ്വന്തമായി കാര്യങ്ങൾ മനസ്സിലാക്കാൻ കഴിയില്ലെന്നും, അത് ലഭിച്ച വിവരങ്ങൾ വെറുതെ ആവർത്തിക്കുന്ന തത്തകൾക്കു തുല്യമാണെന്നുമുള്ള എമിലിബെൻഡർ, ടിംനിറ്റ് ഗെബ്രു എന്നിവരുടെ അഭിപ്രായം ഇവിടെ വ്യക്തമാക്കുന്നു. അതോടൊപ്പം, അൽഗോരിതങ്ങളിൽ ഒളിഞ്ഞുകിടക്കുന്ന വിവേചനങ്ങളെക്കുറിച്ച് കാത്തി ഓ നീലും, സൂപ്പർ ഇന്റലിജൻസ് ഭാവിയിൽ സൃഷ്ടിച്ചേക്കാവുന്ന അപകടങ്ങളെക്കുറിച്ച് നിക് ബോസ്ട്രോമും നൽകുന്ന മുന്നറിയിപ്പുകൾ ഇതിൽ വിശകലനം ചെയ്യുന്നുണ്ട്. മനുഷ്യന്റെ നന്മയ്ക്കും താല്പര്യങ്ങൾക്കും അനുസരിച്ച് എ.ഐ.യെ എങ്ങനെ സുരക്ഷിതമായി ഉപയോഗിക്കാമെന്ന സ്റ്റുവർട്ട് റസലിന്റെ ആശയങ്ങൾ മുൻനിർത്തി സാങ്കേതികവിദ്യയുടെ ധാർമ്മികവശങ്ങളാണ് ഇവിടെ വിശദീകരിക്കുന്നത്.

"കള്ളം പറയരുത്" എന്ന് ഒരു മനുഷ്യൻ പറയുന്നത് അത് തെറ്റാണെന്ന ബോധ്യമുള്ളതുകൊണ്ടാണ്. അവിടെ ഒരു ധാർമ്മികമൂല്യമുണ്ട്. എന്നാൽ "കള്ളം പറയരുത്" എന്ന് എ.ഐ പറയുന്നത് അതിന് ലഭ്യമായിട്ടുള്ള ഡാറ്റയിലുള്ള കോടിക്കണക്കിന് വാചകങ്ങളിൽ 'കള്ളം' എന്ന വാക്കിനുശേഷം 'പറയരുത്' എന്ന് വരാനുള്ള സാധ്യത കൂടുതലായതുകൊണ്ടാണ്. ഇത് വെറുമൊരു ഗണിതശാസ്ത്രരൂപം മാത്രമാണ്. എ.ഐ. ഒരു തെറ്റുചെയ്താൽ അതിന് 'മനസാക്ഷി' ഇല്ലാത്തതുകൊണ്ട് അതിനെ കുറ്റപ്പെടുത്താനാവില്ല. അത് നിർമ്മിച്ചവർക്കും ഉപയോഗിക്കുന്നവർക്കുമാണ് അതിന്റെ ഉത്തരവാദിത്തം. അതുകൊണ്ടാണ് എ.ഐ. 'നൈതികത' യഥാർത്ഥ നൈതികതയല്ലെന്നും ഒരു ഗണിതശാസ്ത്രരൂപം മാത്രമാണെന്നും പറയുന്നത്.

അൽഗോരിതങ്ങളിലെ പക്ഷപാതം: കാത്തി ഓ നീലിന്റെ നിരീക്ഷണങ്ങൾ

അൽഗോരിതങ്ങൾ നമ്മുടെ ജീവിതത്തെ എങ്ങനെയൊക്കെ സ്വാധീനിക്കുന്നുവെന്നും അവയിലെ പക്ഷപാതങ്ങൾ സമൂഹത്തിൽ എന്ത് പ്രത്യാഘാതങ്ങൾ ഉണ്ടാക്കുന്നുവെന്നും ലോകത്തിനുകാട്ടിക്കൊടുത്ത പ്രമുഖ ഗണിതശാസ്ത്രജ്ഞയും എഴുത്തുകാരിയുമാണ് കാത്തി ഓ നീൽ. അവർ ഉന്നയിക്കുന്ന വാദം സാങ്കേതികവിദ്യയുടെ പിന്നിലെ 'മനുഷ്യപക്ഷപാതങ്ങളെ' കൃത്യമായി തുറന്നുകാട്ടുന്ന ഒന്നാണ്.

ഒരു എ.ഐ. മോഡലിന് 'സൗന്ദര്യം' എന്താണെന്ന് സ്വയം അറിയില്ല. അതിനെ പരിശീലിപ്പിക്കാൻ ദശലക്ഷക്കണക്കിന് ചിത്രങ്ങൾ ആവശ്യമാണ്. ഈ ചിത്രങ്ങൾ തിരഞ്ഞെടുക്കുന്ന മനുഷ്യർക്ക് ചില പ്രത്യേക നിറത്തോടോ, ശരീരപ്രകൃതിയോടോ മുൻഗണനയുണ്ടെങ്കിൽ എ.ഐ.യും അതുതന്നെ പഠിക്കുന്നു. അൽഗോരിതം നിർമ്മിക്കുന്ന പ്രോഗ്രാമർ 'സൗന്ദര്യം' അളക്കാൻ ചില മാനദണ്ഡങ്ങൾ നിശ്ചയിക്കുന്നു. ചർമ്മത്തിന്റെ നിറം, മുഖത്തിന്റെ ആകൃതി, പ്രായം എന്നിങ്ങനെ പ്രോഗ്രാമറുടെ വ്യക്തിപരമായ 'അഭിപ്രായം' ആണ് മാനദണ്ഡമായി മാറുന്നത്. 2026-ൽ നടന്ന ഒരു എ.ഐ. സൗന്ദര്യമത്സരത്തിൽ വിജയിച്ചവരിൽ കൂടുതലും വെളുത്ത

മലയാളം റിസേർച്ച് കീ | MALAYALAM RESEARCH KEY | ONLINE EDITION
2026 APRIL-JUNE | VOLUME 02 | ISSUE 02
MULTIDISCIPLINARY PEER-REVIEWED RESEARCH JOURNAL (QUARTERLY)

Journal Home Page: www.researchkey.in | Email: mail@researchkey.in
Published on April 01, 2026 | Time: 10:00 AM IST | e-ISSN:

നിറമുള്ളവരായിരുന്നു. ആ അൽഗോരിതത്തെ പരിശീലിപ്പിക്കാൻ നൽകിയ വിവരങ്ങൾ പ്രധാനമായും യൂറോപ്യൻ സൗന്ദര്യസങ്കൽപ്പങ്ങൾക്ക് മുൻഗണന നൽകുന്നതായിരുന്നു എന്നതാണ് ഇതിനു കാരണം. അതിനാൽ, കറുത്തവർഗ്ഗക്കാരെയോ മറ്റ് വംശജരെയോ സൗന്ദര്യമുള്ളവരായി പരിഗണിക്കാൻ എ.ഐക്ക് കഴിഞ്ഞില്ല. മനുഷ്യർ നൽകുന്ന അപൂർണ്ണമായ വിവരങ്ങൾ എ.ഐ. സംവിധാനങ്ങളിൽ എങ്ങനെ പക്ഷപാതവും വിവേചനവും സൃഷ്ടിക്കുന്നു എന്നതിന്റെ വ്യക്തമായ ഉദാഹരണമാണിത്.

ഡിജിറ്റൽയുഗത്തിൽ മനുഷ്യർ നേരിടുന്ന വളരെ ഗൗരവകരമായ ഒരു പ്രശ്നമാണ് എ.ഐ.അധിഷ്ഠിത ഫിൽട്ടറുകൾ. ഇത് എങ്ങനെയാണ് മനുഷ്യരുടെ സ്വാഭാവിക തനിമയെ ഇല്ലാതാക്കുന്നതെന്ന് വിശദമാക്കാം. ഫിൽട്ടറുകളും എഡിറ്റിംഗ്ആപ്പുകളും ഉപയോഗിച്ച് എല്ലാവരും ഒരേപോലെയുള്ള മുഖം നേടാൻ ശ്രമിക്കുന്നു. എല്ലാവർക്കും ഒരേപോലെ കൂർത്ത മൂക്ക്, തിളങ്ങുന്ന വെളുത്ത ചർമ്മം, വലിയ കണ്ണുകൾ എന്നിവ നൽകുന്നു. ഇതിന്റെ ഫലമായി ലോകത്തിന്റെ ഏതുഭാഗത്തുള്ളവരായാലും ഫോട്ടോകളിൽ ഒരേപോലെ തോന്നിപ്പിക്കുന്നു. ഇത് പ്രാദേശികമായ സവിശേഷതകളെയും വൈവിധ്യങ്ങളെയും ഇല്ലാതാക്കുന്നു. ചർമ്മത്തിലെ പാടുകളും ചെറിയ വ്യത്യാസങ്ങളുമാണ് ഓരോ മനുഷ്യനെയും വ്യത്യസ്തനാക്കുന്നത്. എന്നാൽ എ.ഐ. ആപ്പുകൾ ഈ 'വ്യത്യസ്തത'യെ ഒരു കുറവായി അവതരിപ്പിക്കുന്നു. ഇതു കാണുന്ന യുവാക്കൾ തങ്ങളുടെ സ്വാഭാവികമുഖത്തെ വെറുക്കാനും എല്ലാവരെയുംപോലെയാകാൻ ശ്രമിക്കാനും തുടങ്ങുന്നു.

കാത്തി ഓ നീൽ നിരീക്ഷിക്കുന്നതുപോലെ അൽഗോരിതങ്ങളിൽ മനുഷ്യരുടെ മുൻവിധികളാണ് പ്രവർത്തിക്കുന്നത്. ഇതിന് ഇന്ത്യൻ പശ്ചാത്തലത്തിലുള്ള ഉദാഹരണങ്ങൾ കൂടി പരിശോധിക്കാം. ഇന്ത്യയിലെ സാമൂഹികസാഹചര്യങ്ങളിൽ നിലനിൽക്കുന്ന ജാതീയവും വംശീയവുമായ വിവേചനങ്ങൾ എ.ഐ. ഡാറ്റാസെറ്റുകളിലും പ്രതിഫലിക്കുന്നുണ്ട്. ഉദാഹരണത്തിന്, 'വിജയിയായ ഒരു ഭാരതീയൻ' എന്ന് എ.ഐ.യോട് ആവശ്യപ്പെട്ടാൽ അത് നൽകുന്ന ചിത്രങ്ങളിൽ ഭൂരിഭാഗവും സവർണ്ണ–യൂറോപ്യൻ ശൈലിയിലുള്ള രൂപഭാവങ്ങൾ ആയിരിക്കും. എന്നാൽ ശാരീരിക അധ്വാനം ചെയ്യുന്ന തൊഴിലുകളിൽ ഏർപ്പെടുന്നവരെ ചിത്രീകരിക്കുമ്പോൾ അവരെ കറുത്ത നിറമുള്ളവരായും പാരമ്പര്യവസ്ത്രങ്ങൾ ധരിച്ചവരായും എ.ഐ. അവതരിപ്പിക്കുന്നു. ഇതിൽനിന്ന് ഇന്ത്യയിലെ സാമൂഹികഅസമത്വങ്ങൾ ഡിജിറ്റൽലോകത്തും പുനരുൽപ്പാദിപ്പിക്കുന്നു എന്നു മനസിലാക്കാം.
ഇതിന്റെ കേരളീയ പശ്ചാത്തലത്തിലുള്ള ചില ഉദാഹരണങ്ങൾ കൂടി പരിശോധിക്കാം. 'മുണ്ടും ഷർട്ടും' ധരിച്ച ഒരാളെ ഒരു 'പ്രൊഫഷണൽ' ആയി അംഗീകരിക്കാൻ പല വിദേശ എ.ഐ. മോഡലുകളും തയ്യാറാകുന്നില്ല. പാശ്ചാത്യവസ്ത്രധാരണംമാത്രം മാന്യതയുടെ അടയാളമായി കാണുന്ന ഡാറ്റാസെറ്റുകളിൽനിന്നാണ് എ.ഐ. ഇത് പഠിക്കുന്നത്. തൽഫലമായി, ഒരു 'കേരളീയഅധ്യാപകൻ' അല്ലെങ്കിൽ 'കേരളീയഉദ്യോഗസ്ഥൻ' എന്ന ചിത്രത്തിനായി നാം എ.ഐ.യോട് ആവശ്യപ്പെടുമ്പോൾ, അത് പലപ്പോഴും സവർണ്ണശൈലിയിലുള്ളതോ പാശ്ചാത്യവൽക്കരിക്കപ്പെട്ടതോ ആയ രൂപങ്ങളാണ് നൽകുന്നത്. പ്രാദേശികമായ സവിശേഷതകളും വൈവിധ്യങ്ങളും ഇല്ലാതാക്കി ഒരു പ്രത്യേകതരം 'ഗ്ലോബൽസ്റ്റാൻഡേർഡ്' അടിച്ചേൽപ്പിക്കുകയാണ് ഇതിലൂടെ ചെയ്യുന്നത്.

ആമസോൺ തങ്ങളുടെ കമ്പനിയിലേക്ക് ഏറ്റവും അനുയോജ്യരായ ജീവനക്കാരെ കണ്ടെത്താൻ ഒരു എ.ഐ. സോഫ്റ്റ്വെയർ നിർമ്മിച്ചു. ഇതിനെ പരിശീലിപ്പിക്കാനായി കഴിഞ്ഞ 10 വർഷത്തെ, കമ്പനിയിലെ ജീവനക്കാരുടെ റെസ്യൂമെകൾ നൽകി. ഈ 10 വർഷത്തിനിടയിൽ സാങ്കേതികമേഖലയിൽ ഭൂരിഭാഗവും പുരുഷന്മാരായിരുന്നു ജോലിചെയ്തിരുന്നത്. സ്ത്രീകളുടെ റെസ്യൂമെകൾ തള്ളിക്കളഞ്ഞതായി കണ്ടെത്തിയിട്ടുണ്ട്. കാരണം, കഴിഞ്ഞ പത്തുവർഷത്തെ ഡാറ്റയിൽ സാങ്കേതികജോലികളിൽ ഭൂരിഭാഗവും പുരുഷന്മാരായിരുന്നു. "പുരുഷന്മാരാണ് യോഗ്യർ" എന്ന തെറ്റായ പാഠം എ.ഐ. പഠിച്ചെടുത്തു. എന്നാൽ എ.ഐ.-ക്ക് ഇതിന്റെ സാമൂഹികപശ്ചാത്തലം അറിയില്ല. അത് വിചാരിച്ചത് "പുരുഷൻ എന്ന ലിംഗപദവിയാണ് ജോലിക്ക് വേണ്ട പ്രധാന യോഗ്യത" എന്നാണ്.

എ.ഐ.അലൈൻമെന്റ് : സ്റ്റുവർട്ട് റസ്സലിന്റെ കാഴ്ചപ്പാട്

നിർമ്മിതബുദ്ധിയുടെ മേഖലയിലെ ലോകപ്രശസ്തനായ കമ്പ്യൂട്ടർ ശാസ്ത്രജ്ഞനാണ് സ്റ്റുവർട്ട് റസ്സൽ. കാലിഫോർണിയ സർവകലാശാലയിലെ പ്രൊഫസറായ അദ്ദേഹം, എ.ഐ. മനുഷ്യന് ഗുണകരവും നിയന്ത്രണവിധേയവും ആയിരിക്കണം എന്ന് വാദിച്ചു. സ്റ്റുവർട്ട് റസ്സൽ തന്റെ 'Human Compatible' എന്ന വിഖ്യാത ഗ്രന്ഥത്തിലൂടെ അവതരിപ്പിക്കുന്നത് എ.ഐ.യെ എങ്ങനെ സുരക്ഷിതമായി നിലനിർത്താം എന്നതിനെക്കുറിച്ചുള്ള വിപ്ലവകരമായ ചില ആശയങ്ങളാണ്. ഇതിനെ അദ്ദേഹം 'Beneficial AI' എന്നാണ് വിളിക്കുന്നത്. എങ്ങനെയാണ് മനുഷ്യന് ദോഷകരമല്ലാത്ത രീതിയിൽ എ.ഐ.യെ നിയന്ത്രിക്കുക എന്നത് വലിയൊരു വെല്ലുവിളിയാണ്. ഇതിനെയാണ് 'എ.ഐ. അലൈൻമെന്റ് പ്രോബ്ലം' എന്ന് വിളിക്കുന്നത്. ഇത് വിശദമാക്കാനായി അദ്ദേഹം മൂന്ന് തത്വങ്ങൾ നിർദ്ദേശിക്കുന്നു. നിർമ്മിതബുദ്ധിയുടെ പ്രവർത്തനങ്ങൾ മനുഷ്യലക്ഷ്യങ്ങളുമായി ചേർന്നുപോകുന്നു എന്ന് ഉറപ്പുവരുത്തുന്ന പ്രക്രിയയാണ് 'എ.ഐ. അലൈൻമെന്റ്'. ഇതിനായി സ്റ്റുവർട്ട് റസ്സൽ മുന്നോട്ടുവെക്കുന്ന ഒന്നാമത്തെ തത്വം 'മനുഷ്യനന്മ' എന്നതാണ്. എ.ഐ.–യുടെ ഏകലക്ഷ്യം മനുഷ്യതാല്പര്യങ്ങൾ സംരക്ഷിക്കുക എന്നതായിരിക്കണം. അതായത്, എ.ഐ. അതിന്റെ സ്വന്തം നിലനിൽപ്പിനോ മറ്റ് പ്രത്യേക ലക്ഷ്യങ്ങൾക്കോ വേണ്ടി

മലയാളം റിസേർച്ച് കീ | MALAYALAM RESEARCH KEY | ONLINE EDITION
2026 APRIL-JUNE | VOLUME 02 | ISSUE 02
MULTIDISCIPLINARY PEER-REVIEWED RESEARCH JOURNAL (QUARTERLY)

Journal Home Page: www.researchkey.in | Email: mail@researchkey.in
Published on April 01, 2026 | Time: 10:00 AM IST | e-ISSN:

പ്രവർത്തിക്കാൻ പാടില്ല. മനുഷ്യൻ ആഗ്രഹിക്കുന്നത് കൃത്യമായി സാധിച്ചുകൊടുക്കുന്ന വിശ്വസ്തസേവകനായി എ.ഐ. പ്രവർത്തിക്കണം.
രണ്ടാമതായി എ.ഐ.-ക്ക് മനുഷ്യന്റെ താല്പര്യങ്ങൾ എന്താണെന്ന് കൃത്യമായി അറിയില്ല എന്ന ബോധ്യം ഉണ്ടാകണം. ഉദാഹരണമായി "എനിക്ക് ചായ വേണം" എന്ന് നമ്മൾ പറഞ്ഞാൽ, എ.ഐ. ഉടനെ ചായ ഉണ്ടാക്കാൻ ഓടരുത്. അതിനുപകരം, ഏതുതരം ചായ, പഞ്ചസാരയുടെ അളവ് തുടങ്ങിയ കാര്യങ്ങൾ ചോദിച്ച മനസ്സിലാക്കാൻ ശ്രമിക്കണം. തനിക്ക് എല്ലാം അറിയാം എന്ന ധാരണ എ.ഐ.-ക്ക് ഉണ്ടാകുന്നത് വലിയ അപകടമാണ്. മൂന്നാമത്തേത് മനുഷ്യന്റെ പെരുമാറ്റം നിരീക്ഷിച്ച പഠിക്കുക എന്നതാണ്. മനുഷ്യന്റെ യഥാർത്ഥ താല്പര്യങ്ങൾ എന്താണെന്ന് അറിയാൻ അവരുടെ പെരുമാറ്റത്തെ എ.ഐ. നിരീക്ഷിക്കണം. മനുഷ്യർ എപ്പോഴും അവരുടെ ആഗ്രഹങ്ങൾ വാക്കുകളിലൂടെ കൃത്യമായി പ്രകടിപ്പിക്കാറില്ല. മനുഷ്യർ ചെയ്യുന്ന കാര്യങ്ങൾ കണ്ടും, മനുഷ്യരുടെ തീരുമാനങ്ങൾ വിശകലനംചെയ്തുമാണ് എ.ഐ. മാനവരാശിയുടെ മൂല്യങ്ങൾ മനസ്സിലാക്കേണ്ടത്. ഇതാണ് 'മൂല്യസംസ്കാരം' എന്ന് സ്റ്റുവർട്ട് റസ്സൽ നിരീക്ഷിക്കുന്നത്.

എ.ഐ. അലൈൻമെന്റിലെ പ്രായോഗിക വെല്ലുവിളികൾ

സ്റ്റുവർട്ട് റസ്സൽ മുന്നോട്ടുവെക്കുന്ന തത്വങ്ങൾ സൈദ്ധാന്തികമായി വളരെ മികച്ചതാണെങ്കിലും, അവ പ്രായോഗികമായി നടപ്പിലാക്കുന്നതിൽ സങ്കീർണ്ണമായ ചില സാങ്കേതിക-നൈതിക തടസ്സങ്ങളുണ്ട്. ലോകമെമ്പാടുമുള്ള മനുഷ്യർ ഒരേ മൂല്യങ്ങളല്ല പിന്തുടരുന്നത്. ഒരു സംസ്കാരത്തിൽ ശരിയെന്ന് കരുതുന്നത് മറ്റൊരു സംസ്കാരത്തിൽ തെറ്റായേക്കാം. ഇത്തരം സാഹചര്യങ്ങളിൽ എ.ഐ. ഏതു വിഭാഗത്തിന്റെ മൂല്യങ്ങൾക്കാണ് മുൻഗണന നൽകേണ്ടത് എന്നത് വലിയൊരു ചോദ്യമാണ്. ഉദാഹരണത്തിന്, സ്വകാര്യതയെക്കാൾ സാമൂഹികസുരക്ഷയ്ക്ക് പ്രാധാന്യംനൽകുന്ന ഒരു സമൂഹത്തിലെ എ.ഐ.യുടെ പ്രവർത്തനം. മാത്രമല്ല, മനുഷ്യർ എപ്പോഴും യുക്തിസഹമായ തീരുമാനങ്ങളല്ല എടുക്കുന്നത്. പലപ്പോഴും വികാരങ്ങൾക്കോ മുൻവിധികൾക്കോ അടിമപ്പെട്ടാണ് പ്രവർത്തിക്കുന്നത്. എ.ഐ. മനുഷ്യന്റെ പെരുമാറ്റം നിരീക്ഷിച്ച് മൂല്യങ്ങൾ പഠിക്കാൻ ശ്രമിച്ചാൽ, മനുഷ്യരിലെ തെറ്റായ ശീലങ്ങളും വിവേചനബുദ്ധിയും കൂടി അത് സ്വാഭാവികമായി പഠിച്ചെടുക്കാൻ സാധ്യതയുണ്ട്. മനുഷ്യർ പലപ്പോഴും തങ്ങളുടെ ആവശ്യങ്ങൾ കൃത്യമായി പ്രകടിപ്പിക്കാറില്ല. "എന്റെ സൗകര്യം വർദ്ധിപ്പിക്കുക" എന്ന് ഒരാൾ ആവശ്യപ്പെടുമ്പോൾ, മറ്റൊരാളുടെ അവകാശങ്ങൾ ലംഘിച്ചുകൊണ്ട് എ.ഐ. അത് ചെയ്യരുത്. ലക്ഷ്യങ്ങൾ നൽകുന്നതിലെ ഈ അവ്യക്തത എ.ഐ. തെറ്റായ രീതിയിൽ വ്യാഖ്യാനിക്കാൻ ഇടയാക്കിയേക്കാം. എ.ഐ.-യെ ആര് നിയന്ത്രിക്കുന്നു എന്നതും പ്രധാനമാണ്. വൻകിട കോർപ്പറേറ്റുകളുടെ ലാഭത്തിനാണോ അതോ സാധാരണ മനുഷ്യരുടെ നന്മയ്ക്കാണോ എ.ഐ. അലൈൻമെന്റ് നടത്തേണ്ടത് എന്നതിൽ സാമ്പത്തികവും രാഷ്ട്രീയവുമായ താല്പര്യങ്ങൾ ഒളിഞ്ഞിരിപ്പുണ്ട്.

'ബ്ലാക്ക്ബോക്സ്' പ്രതിഭാസവും ഉത്തരവാദിത്തവും

'ബ്ലാക്ക്ബോക്സ്' പ്രതിഭാസം എന്നത് ആധുനിക എ.ഐ. സാങ്കേതികവിദ്യയിലെ ഏറ്റവും ദുരൂഹമായ പ്രതിഭാസമാണ്. ഒരു യന്ത്രം തീരുമാനമെടുത്തത് എങ്ങനെയെന്ന് അത് നിർമ്മിച്ച പ്രോഗ്രാമർക്കു പോലും കൃത്യമായി വിശദീകരിക്കാൻ കഴിയാത്ത അവസ്ഥയാണുള്ളത്. ആധുനിക എ.ഐ. നേരിടുന്ന ഏറ്റവും വലിയ വെല്ലുവിളി 'ബ്ലാക്ക്ബോക്സ്' പ്രശ്നമാണ്. സാധാരണ കമ്പ്യൂട്ടർ പ്രോഗ്രാമുകളിൽ മനുഷ്യർ നൽകുന്ന നിർദ്ദേശങ്ങൾ അനുസരിച്ചാണ് എ.ഐ. പ്രവർത്തിക്കുന്നത്. എന്നാൽ ആധുനിക എ.ഐ. പ്രവർത്തിക്കുന്നത് കോടിക്കണക്കിന് ഡാറ്റയിൽ നിന്ന് സ്വയം പാറ്റേണുകൾ കണ്ടെത്തിയാണ്.
എ.ഐ.യുടെ ഉള്ളിൽ 'ന്യൂറൽ നെറ്റ്വർക്കുകൾ' എന്നറിയപ്പെടുന്ന ദശലക്ഷക്കണക്കിന് പാളികളുണ്ട്. ഇവയിലൂടെ ഡാറ്റ കടന്നുപോകുമ്പോൾ ഓരോ ഘട്ടത്തിലും വളരെ സങ്കീർണ്ണമായ മാറ്റങ്ങൾ സംഭവിക്കുന്നു. ഈ മാറ്റങ്ങൾ മനുഷ്യന്റെ ബുദ്ധിക്ക് പെട്ടെന്ന് വിശകലനം ചെയ്യാൻ കഴിയുന്നതിലും അപ്പുറമാണ്.

'ബ്ലാക്ക്ബോക്സ്' പ്രതിഭാസവും സുതാര്യതയും

ആധുനിക എ.ഐ. നേരിടുന്ന ഏറ്റവും വലിയ വെല്ലുവിളിയായ 'ബ്ലാക്ക് ബോക്സ്' പ്രശ്നത്തെക്കുറിച്ച് നാം ചർച്ച ചെയ്തു. ഒരു യന്ത്രം എങ്ങനെയാണ് ഒരു പ്രത്യേക തീരുമാനമെടുത്തത് എന്ന് വിശദീകരിക്കാൻ കഴിയാത്തത് നൈതികമായ വലിയ പ്രതിസന്ധിയാണ്. സാധാരണ എ.ഐ. സംവിധാനങ്ങൾ ഒരു തീരുമാനം പുറത്തുവിടുക മാത്രമാണ് ചെയ്യുന്നത്. എന്നാൽ Explainable AI ആ തീരുമാനത്തിലേക്ക് എത്താൻ ഉപയോഗിച്ച കാരണങ്ങൾ കൂടി മനുഷ്യർക്ക് വിശദീകരിക്കുന്നു. ഉദാഹരണത്തിന്, ഒരാളുടെ വായ്പാ അപേക്ഷ എ.ഐ. നിരസിക്കുകയാണെങ്കിൽ, അത് 'നിരസിച്ചു' എന്ന് പറയുന്നതിനുപകരം, "താങ്കളുടെ വരുമാനവും മുൻകാല ഇടപാടുകളും തമ്മിലുള്ള പൊരുത്തക്കേടാണ് ഇതിന് കാരണം" എന്ന് വ്യക്തമാക്കാൻ Explainable AI -ക്ക് കഴിയും. അൽഗോരിതങ്ങൾ എങ്ങനെ പ്രവർത്തിക്കുന്നു എന്ന് ബോധ്യപ്പെടുത്താനും തീരുമാനങ്ങളുടെ പിന്നിലെ യുക്തി മനസ്സിലാക്കാനും ഒരു തെറ്റായ തീരുമാനമുണ്ടായാൽ അത് എവിടെ സംഭവിച്ചു എന്ന് കണ്ടെത്താനും തിരുത്താനും ഇത് പ്രോഗ്രാമർമാരെ സഹായിക്കുന്നു.

മലയാളം റിസേർച്ച് കീ | MALAYALAM RESEARCH KEY | ONLINE EDITION
2026 APRIL-JUNE | VOLUME 02 | ISSUE 02
MULTIDISCIPLINARY PEER-REVIEWED RESEARCH JOURNAL (QUARTERLY)

Journal Home Page: www.researchkey.in | Email: mail@researchkey.in
Published on April 01, 2026 | Time: 10:00 AM IST | e-ISSN:

ഭാവിയിലെ വെല്ലുവിളികൾ: സൂപ്പർ ഇന്റലിജൻസ്

നിക് ബോസ്ട്രോമിന്റെ 'സൂപ്പർഇന്റലിജൻസ്' എന്ന ആശയം സാങ്കേതികവിദ്യയുടെ വളർച്ച എവിടെച്ചെന്ന് അവസാനിക്കും എന്നതിനെക്കുറിച്ചുള്ള ഗൗരവകരമായ മുന്നറിയിപ്പാണ്. മനുഷ്യബുദ്ധിയെക്കാൾ ദശലക്ഷക്കണക്കിന് മടങ്ങ് വേഗത്തിലും ആഴത്തിലും ചിന്തിക്കാൻ കഴിയുന്ന ഒരു എ.ഐ. വരുമ്പോൾ സംഭവിക്കാവുന്ന അപകടങ്ങളെ അദ്ദേഹം ലളിതമായ ഉദാഹരണങ്ങളിലൂടെ വിവരിക്കുന്നു. മനുഷ്യന്റെ ബൗദ്ധികമായ എല്ലാ കഴിവുകളെയും (ശാസ്ത്രം, തന്ത്രം, സാമൂഹിക വൈദഗ്ധ്യം) പിന്നിലാക്കുന്ന ഒരു കൃത്രിമബുദ്ധിയെയാണ് സൂപ്പർഇന്റലിജൻസ് എന്നുവിളിക്കുന്നത്. സൂപ്പർ ഇന്റലിജൻസിന്റെ നീക്കങ്ങൾ മനസ്സിലാക്കാൻ മനുഷ്യനുകഴിയില്ല എന്നതാണ് ഇതിന്റെ പ്രധാന വെല്ലുവിളി. മനുഷ്യനേക്കാൾ ബുദ്ധിയുള്ള 'സൂപ്പർഇന്റലിജൻസ്' വരുമ്പോൾ, അതിനെ നിയന്ത്രിക്കാൻ നമുക്ക് കഴിഞ്ഞെന്നു വരില്ല.

ഉപദർശനങ്ങൾ

നിർമ്മിതബുദ്ധി എന്നത് മനുഷ്യൻ ഇതുവരെ നിർമ്മിച്ചതിൽവെച്ച് ഏറ്റവും ശക്തമായ കണ്ടെത്തലാണ്. എന്നാൽ അത് പ്രവർത്തിക്കുന്നത് മനുഷ്യൻ നൽകുന്ന വിവരങ്ങളുടെ അടിസ്ഥാനത്തിലാണ്. അതിനാൽ, എ.ഐ.യിലെ പിഴവുകൾ യഥാർത്ഥത്തിൽ മനുഷ്യസമൂഹത്തിലെ പിഴവുകളുടെ പ്രതിഫലനമാണ്. സാങ്കേതികവിദ്യയുടെ വളർച്ചയ്ക്കൊപ്പംതന്നെ നൈതികതയുടെ വളർച്ചയും അനിവാര്യമാണ്.

സൗന്ദര്യസങ്കൽപ്പങ്ങളെ ഏകീകരിക്കുന്ന, വംശീയമായ വേർതിരിവുകൾ കാണിക്കുന്ന ഒരു എ.ഐ. മാനവരാശിക്ക് ആവശ്യമില്ല. വൈവിധ്യങ്ങളെ ഉൾക്കൊള്ളുന്ന, മാനുഷികമൂല്യങ്ങളോട് ചേർന്നുനിൽക്കുന്ന സാങ്കേതികതയാണ് നമുക്കുവേണ്ടത്. അതിനായി സാങ്കേതികവിദഗ്ധരും, സാമൂഹികശാസ്ത്രജ്ഞരും, നിയമവിദഗ്ധരും ഒത്തുചേർന്നുള്ള ഒരു പ്രവർത്തനം അനിവാര്യമാണ്. നിർമ്മിതബുദ്ധിയുടെ കാലത്തെ നൈതികപ്രശ്നങ്ങൾ പരിഹരിക്കാൻ 'എ.ഐ അലൈൻമെന്റ്' പോലെയുള്ള സുരക്ഷാമാനദണ്ഡങ്ങൾ കർശനമാക്കേണ്ടതുണ്ട്. മനുഷ്യന്റെ ധാർമ്മികമൂല്യങ്ങൾക്കു വിധേയമായി പ്രവർത്തിക്കാൻ കഴിയുന്ന രീതിയിൽ അൽഗോരിതങ്ങളെ പരിശീലിപ്പിക്കണം. നിക്ക് ബോസ്ട്രോം സൂചിപ്പിക്കുന്നതുപോലെ ഭാവിയിൽ ഒരു 'സൂപ്പർ ഇന്റലിജൻസ്' രൂപപ്പെട്ടാൽ, അത് മനുഷ്യരാശിയുടെ നിലനിൽപ്പിന് ഭീഷണിയാകാതിരിക്കാൻ ആഗോളതലത്തിൽ നിയമനിർമ്മാണങ്ങളും നൈതികചട്ടക്കൂടുകളും അടിയന്തരമായി രൂപീകരിക്കേണ്ടതുണ്ട്.

ഗ്രന്ഥസൂചി

Bender, E. M., et al. (2021). "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?". FAccT '21 Proceedings.

Bostrom, Nick. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford University Press.

Crawford, Kate. (2021). Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press.

O'Neil, Cathy. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Books.

Russell, Stuart. (2019). Human Compatible: Artificial Intelligence and the Problem of Control. Viking.

ഇ-സൂചിക

. ഏഷ്യാനെറ്റ് ന്യൂസ്. (2024). ഡീപ് ഫേക്കും എ.ഐ പക്ഷപാതവും: ഡിജിറ്റൽ ലോകത്തെ പുതിയ വെല്ലുവിളികൾ. [Online]. ലഭ്യമായ ഇടം: asianetnews.com
. ദേശാഭിമാനി. (2023). ആർട്ടിഫിഷ്യൽ ഇന്റലിജൻസ്: തൊഴിലും മൂലധനവും തമ്മിലുള്ള പുതിയ പോരാട്ടം. [Online]. ലഭ്യമായ ഇടം: deshabhimani.com
. മനോരമ ഓൺലൈൻ. (2023). എ.ഐ കാലത്തെ തൊഴിൽ ഭീതി: ആർക്കൊക്കെ ജോലി നഷ്ടപ്പെടും? [Online]. ലഭ്യമായ ഇടം: manoramaonline.com
. മനോരമ ഓൺലൈൻ. (2024). നിങ്ങളെ ജോലിക്ക് എടുക്കുന്നത് ഒരു എ.ഐ ആണോ? റിക്രൂട്ട്മെന്റിലെ പക്ഷപാതങ്ങൾ. [Online]. ലഭ്യമായ ഇടം: manoramaonline.com
. മനോരമ ഓൺലൈൻ. (2026). ആർട്ടിഫിഷ്യൽ ഇന്റലിജൻസിലെ പക്ഷപാതങ്ങളും വെല്ലുവിളികളും. [Online]. ലഭ്യമായ ഇടം: manoramaonline.com
. മാതൃഭൂമി. (2023). ചാറ്റ് ജിപിറ്റിയും എ.ഐ വിപ്ലവവും: തൊഴിൽ വിപണിയിൽ സംഭവിക്കുന്ന മാറ്റങ്ങൾ. [Online]. ലഭ്യമായ ഇടം: mathrubhumi.com
. മാതൃഭൂമി. (2024). അൽഗോരിതങ്ങളിലെ ചതിക്കുഴികൾ: എ.ഐ പക്ഷപാതപരമാകുന്നത് എന്തുകൊണ്ട്? [Online]. ലഭ്യമായ ഇടം: mathrubhumi.com
. മാതൃഭൂമി. (2026). എ.ഐ.യും ഭാവിയിലെ തൊഴിൽ വിപണിയും: ഒരു വിശകലനം. [Online]. ലഭ്യമായ ഇടം: mathrubhumi.com

മലയാളം റിസേർച്ച് കീ | MALAYALAM RESEARCH KEY | ONLINE EDITION
2026 APRIL-JUNE | VOLUME 02 | ISSUE 02
MULTIDISCIPLINARY PEER-REVIEWED RESEARCH JOURNAL (QUARTERLY)

Journal Home Page: www.researchkey.in | Email: mail@researchkey.in
Published on April 01, 2026 | Time: 10:00 AM IST | e-ISSN:

About the Researcher
ഡോ.ഷിംജിത്ത് എസ്. പി
അസിസ്റ്റന്റ് പ്രൊഫസർ(On Contract)
സെന്റ് മേരീസ് കോളേജ് (Autonomous) തൃശ്ശൂർ
Email: shimjithsp@gmail.com
Mob: 9020868197