

4.1 (എ) സ്വയംഭരണ AI സിസ്റ്റങ്ങളുടെ ലക്ഷ്യങ്ങളെ നമ്മുടെ സ്വന്തം ലക്ഷ്യങ്ങളുമായി എങ്ങനെ യോജിപ്പിക്കാം?

മനുഷ്യ മേൽനോട്ടമില്ലാതെ തീരുമാനങ്ങൾ എടുക്കാൻ കഴിവുള്ള കൂടുതൽ സ്വയംഭരണത്തിലേക്ക് കൃത്രിമബുദ്ധി പുരോഗമിക്കുമ്പോൾ, അതിന്റെ ലക്ഷ്യങ്ങളും പെരുമാറ്റങ്ങളും മാനുഷിക മൂല്യങ്ങളോടും ഉദ്ദേശ്യങ്ങളോടും യോജിക്കുന്നുണ്ടെന്ന് ഉറപ്പാക്കുക എന്നതാണ് പ്രധാന വെല്ലുവിളി. ഇത് AI വിന്യാസ പ്രശ്നം എന്നറിയപ്പെടുന്നു. കൃത്രിമബുദ്ധി സംവിധാനങ്ങൾ അവയുടെ പ്രോഗ്രാമിംഗിനെയോ പരിശീലനത്തെയോ അടിസ്ഥാനമാക്കി ലക്ഷ്യങ്ങൾ പിന്തുടരുന്നു. എന്നിരുന്നാലും, അവയുടെ ലക്ഷ്യങ്ങൾ അവ്യക്തമായി വ്യക്തമാക്കിയതോ, അപര്യാപ്തമോ, അല്ലെങ്കിൽ തെറ്റായി വ്യാഖ്യാനിക്കപ്പെട്ടതോ ആണെങ്കിൽ, അവ നിർദ്ദേശങ്ങൾ പാലിക്കുന്നതായി തോന്നുന്ന ഫലങ്ങൾ നൽകിയേക്കാം, എന്നാൽ അതേ സമയം മാനുഷിക മൂല്യങ്ങൾക്ക് വിരുദ്ധവുമാണ്. "കാര്യക്ഷമത പരമാവധിയാക്കാൻ" നിർദ്ദേശിക്കപ്പെട്ട ഒരു AI നിതിയെയോ സുരക്ഷയെയോ കുറിച്ചുള്ള പരിഗണനകളെ അവഗണിച്ചേക്കാം. മനുഷ്യ ലക്ഷ്യങ്ങളുമായി AI യെ വിന്യസിക്കുന്നതിന് കമ്പ്യൂട്ടറുകൾ നമ്മുടെ നിർദ്ദേശങ്ങൾ പാലിക്കുക മാത്രമല്ല, നമ്മുടെ ഉദ്ദേശ്യങ്ങൾ മനസ്സിലാക്കുകയും വേണം.

AI സംവിധാനങ്ങൾ കൂടുതൽ സങ്കീർണ്ണവും വ്യാപകവുമാകുമ്പോൾ, ഭാവിയിലെ AI പുരോഗതികൾക്കായുള്ള അഭിലാഷങ്ങൾ തീവ്രമാകുമ്പോൾ, നിരവധി ധാർമ്മിക ചോദ്യങ്ങൾ ഉയർന്നുവന്നിട്ടുണ്ട്. വൈവിധ്യമാർന്ന മേഖലകളിലെ ഉപദേശക, തീരുമാനമെടുക്കൽ പ്രക്രിയകൾ ധാർമ്മിക പ്രത്യാഘാതങ്ങൾ വർദ്ധിപ്പിച്ചിട്ടുണ്ട്. പല പ്രൊഫഷണൽ, സാമ്പത്തിക മേഖലകളിലും ധാർമ്മിക മാനദണ്ഡങ്ങൾ കൂടുതൽ പ്രാധാന്യമർഹിക്കുന്നു. വിവിധ മേഖലകളിലെ കമ്പ്യൂട്ടർ അധിഷ്ഠിത സഹായ പരിഹാരങ്ങൾ കൂടുതൽ അടിയന്തിര അന്വേഷണങ്ങൾ ഉയർത്തുന്നു. സങ്കീർണ്ണമായ ധാർമ്മിക തീരുമാനങ്ങൾ എടുക്കുന്നതിൽ ഈ സംവിധാനങ്ങൾ എത്രത്തോളം പങ്കെടുക്കുമെന്ന് നമുക്ക് പ്രതീക്ഷിക്കാം? അത്തരം സംവിധാനങ്ങൾ ഒരു പരിധിവരെ ധാർമ്മിക സ്വയംഭരണം പ്രകടിപ്പിക്കുമെന്നോ മനുഷ്യൻ സൃഷ്ടിച്ച ധാർമ്മിക വീക്ഷണകോണുകളേക്കാൾ ആധികാരികമോ മികച്ചതോ ആയി കണക്കാക്കേണ്ട ധാർമ്മിക ഉപദേശം നൽകുമെന്നോ നമുക്ക് പ്രതീക്ഷിക്കാനാകുമോ? മെഷീനുകളിൽ ധാർമ്മിക തീരുമാനമെടുക്കൽ അനുകരിക്കുന്നതിലൂടെ എന്ത് ലക്ഷ്യങ്ങൾ കൈവരിക്കാനാകും? ഈ സിമുലേഷനിൽ ഏതൊക്കെ പ്രവർത്തനങ്ങളാണ് ഉൾപ്പെടുത്തുക? ഈ പദ്ധതികൾ നടപ്പിലാക്കുന്നതിനുള്ള ഏറ്റവും ഫലപ്രദമായ രീതികൾ ഏതൊക്കെയാണ്? ഏത് സാഹചര്യത്തിലാണ് കമ്പ്യൂട്ടർ അധിഷ്ഠിത തീരുമാന സംവിധാനങ്ങൾ അവയുടെ വിലയിരുത്തലുകളുടെയോ പ്രവൃത്തികളുടെയോ ധാർമ്മിക ഉത്തരവാദിത്തമോ ഉത്തരവാദിത്തമോ ആയി കണക്കാക്കേണ്ടതെന്ന് നമുക്ക് പ്രതീക്ഷിക്കാം?

കൃത്രിമ ഏജൻ്റ്മാർ നമ്മുടെ ധാർമ്മിക യുക്തിയെ എങ്ങനെ സഹായിക്കും അല്ലെങ്കിൽ ധാർമ്മികമായി ഉത്തരവാദിത്തമുള്ള പെരുമാറ്റത്തെ എങ്ങനെ പകർത്തും എന്നതിനെക്കുറിച്ചുള്ള അന്വേഷണങ്ങൾക്കൊപ്പം, ഈ ഏജൻ്റ്മാരെ യഥാർത്ഥ ധാർമ്മിക ഉത്തരവാദിത്തമുള്ള എന്റിറ്റികളായി ന്യായമായും കണക്കാക്കാൻ കഴിയുമോ എന്ന ചോദ്യങ്ങളും ഉണ്ട്. അതായത് പരമ്പരാഗത മുതിർന്ന മനുഷ്യ ധാർമ്മിക ഏജൻ്റ്മാരുടെ പ്രവർത്തനങ്ങൾക്ക് സമാനമായ ധാർമ്മിക വിലയിരുത്തൽ അർഹിക്കുന്ന ഏജൻ്റ്മാരെ സൂചിപ്പിക്കുന്നു. ധാർമ്മിക അവകാശങ്ങളുള്ളതായി നാം കാണുന്നതിന് മുമ്പ് കൃത്രിമ ജീവികളെ എത്രത്തോളം വികസിപ്പിക്കാൻ കഴിയും എന്നതിലാണ് അന്വേഷണം ശ്രദ്ധ കേന്ദ്രീകരിക്കുന്നത്. മനുഷ്യൻ്റെ ധാർമ്മിക പരിഗണനയെ ന്യായീകരിക്കുന്ന ആവശ്യങ്ങളോ താൽപ്പര്യങ്ങളോ ഉള്ളതായി കൃത്രിമ ഏജൻ്റ്മാരെ ഏതൊക്കെ സാഹചര്യങ്ങളിൽ കാണാം? കൃത്രിമവും സ്വാഭാവിക ഏജൻ്റ്മാർ ധാർമ്മികമായി വൈരുദ്ധ്യമുള്ളപ്പോൾ അവയെ എങ്ങനെ യോജിപ്പിക്കാം? കൃത്രിമ ഏജൻ്റ്മാരെ നിയമപരമായ അവകാശങ്ങളുടെ വാഹകരായോ നിയമവ്യവസ്ഥയ്ക്കുള്ളിൽ വ്യക്തിത്വമുള്ള സ്ഥാപനങ്ങളായോ അംഗീകരിക്കാനുള്ള സാധ്യതയെക്കുറിച്ച് പ്രത്യേകവും എന്നാൽ പ്രസക്തവുമായ ഒരു കൂട്ടം ചോദ്യങ്ങൾ ഉയർന്നുവരുന്നു.

ധാർമ്മിക നിലപാട് എന്ന ആശയത്തിന്റെ ദാർശനിക വിശകലനമാണിത്. ധാർമ്മിക പദവി നേടുക എന്നാൽ ധാർമ്മിക പ്രാധാന്യം അല്ലെങ്കിൽ സ്ഥാനം ഉണ്ടായിരിക്കുക എന്നാണ്. ധാർമ്മിക ഏജൻ്റ്മാർക്ക് ധാർമ്മിക ബാധ്യതകൾ ഉള്ളതോ കൈവശം വച്ചേക്കാവുന്നതോ ആയ ഒരു സ്ഥാപനമാണിത്. ഒരു സ്ഥാപനത്തിന് ധാർമ്മിക പദവി ഉണ്ടെങ്കിൽ, നാം ആവേശത്തോടെ പ്രവർത്തിക്കരുത്; നമ്മുടെ തീരുമാനമെടുക്കൽ പ്രക്രിയകളിൽ അതിന്റെ ആഗ്രഹങ്ങൾ, താൽപ്പര്യങ്ങൾ അല്ലെങ്കിൽ ക്ഷേമം പരിഗണിക്കാൻ നാം ധാർമ്മികമായി നിർബന്ധിതരാണ്. മാത്രമല്ല, നമുക്കോ മറ്റുള്ളവർക്കോ ഉണ്ടാകാവുന്ന നേട്ടങ്ങൾക്കായി മാത്രമല്ല, മറിച്ച് അതിന്റെ ആവശ്യങ്ങൾക്ക് അന്തർലീനമായ ധാർമ്മിക പ്രാധാന്യം ഉള്ളതിനാൽ ഈ പ്രവർത്തനം ഏറ്റെടുക്കാൻ നമുക്ക് ധാർമ്മിക ബാധ്യതയുണ്ട്, അതായത് അസ്തിത്വത്തിന്റെയും ആവശ്യങ്ങളുടെയും അന്തർലീനമായ മൂല്യം നാം തിരിച്ചറിയുകയും ബഹുമാനിക്കുകയും വേണം.

വ്യക്തിഗത ധാർമ്മികമോ മതപരമോ ആയ വിശ്വാസങ്ങളെ ആശ്രയിച്ചിരിക്കുന്ന വൈവിധ്യമാർന്ന സാങ്കേതിക വിദ്യകൾ, മനുഷ്യന്റെ സമ്പൂർണ്ണവും തുല്യവുമായ ധാർമ്മിക നിലയിലുള്ള പ്രബലമായ ബോധ്യത്തെ സാധൂകരിക്കുന്നു. ഒരു ദൈവഹിതത്തിൽ ധാർമ്മിക പ്രസ്താവനകൾ അടിസ്ഥാനപ്പെടുത്തുന്ന ഒരു ദൈവവിശ്വാസി, ഈ അസ്തിത്വം ഓരോ വ്യക്തിക്കും തുല്യമായ ധാർമ്മിക അവകാശങ്ങൾ, അന്ത്യം, മൂല്യം അല്ലെങ്കിൽ മൂല്യം നൽകിയിട്ടുണ്ടെന്ന് വാദിച്ചേക്കാം. പെരുമാറ്റത്തെക്കുറിച്ചുള്ള പഠനമാണ് ധാർമ്മികത. ഇത് മനുഷ്യന്റെ പ്രവർത്തനങ്ങളെ അവയുടെ ധാർമ്മിക കൃത്യതയോ തെറ്റോ അനുസരിച്ച് വിലയിരുത്തുന്നു. "സ്വഭാവം" എന്നർത്ഥം വരുന്ന "എത്തോസ്" എന്ന ഗ്രീക്ക് പദത്തിൽ നിന്നും "ആചാരങ്ങൾ" എന്നർത്ഥം വരുന്ന "മോർസ്" എന്ന ലാറ്റിൻ പദത്തിൽ നിന്നുമാണ് "ധാർമ്മികത" എന്ന പദം ഉത്ഭവിച്ചത്. "മനുഷ്യ പെരുമാറ്റത്തിൽ ശരിയോ മാനുഷമോ എന്താണെന്നതിന്റെ പരിശോധന" അല്ലെങ്കിൽ "പെരുമാറ്റത്തിൽ അന്തർലീനമായ ആദർശങ്ങളുടെ അന്വേഷണം" എന്നാണ് മക്കൻസി ധാർമ്മികതയെ വിശേഷിപ്പിക്കുന്നത്. പ്രവർത്തനങ്ങളുടെ ധാർമ്മികതയെ വിലയിരുത്തുന്ന ശിക്ഷണമാണ് ധാർമ്മികത. മനുഷ്യന്റെ പെരുമാറ്റങ്ങളെയും പാരമ്പര്യങ്ങളെയും, അല്ലെങ്കിൽ, അവരുടെ വ്യക്തിത്വങ്ങളെയും അവരുടെ പതിവ് പ്രവർത്തനങ്ങളെ നയിക്കുന്ന തത്വങ്ങളെയും പരിശോധിക്കുന്നു, കൂടാതെ ഈ തത്വങ്ങളുടെ ശരിയോ തെറ്റോ നിർവചിക്കുന്നതും ഈ ശീലങ്ങളുടെ നന്മയോ തിന്മയോ വിലയിരുത്തുന്നു. "ശരി", "നല്ലത്" എന്നീ വാക്യങ്ങൾക്ക് ചില വ്യക്തത ആവശ്യമാണ്.

a. വലത്: - 'വലത്' എന്ന പദം ലാറ്റിൻ 'റെക്ടസ്' എന്നതിൽ നിന്നാണ് ഉത്ഭവിച്ചത്, അതായത് 'നേരായത്' അല്ലെങ്കിൽ 'നിയമത്തിന് അനുസൃതമായി'. പ്രവർത്തനം ശരിയാണെന്ന് നമ്മൾ ഉറപ്പിക്കുമ്പോൾ, അത് അടിസ്ഥാനപരമായി സ്ഥാപിത നിയമങ്ങൾ പാലിക്കുന്നുവെന്ന് ഞങ്ങൾ സൂചിപ്പിക്കുന്നു.

b. നല്ലത്: - 'നല്ലത്' എന്ന പദം ജർമ്മൻ 'കടൽ' എന്ന പദവുമായി ബന്ധപ്പെട്ടിരിക്കുന്നു. ഒരു പ്രത്യേക ഉദ്ദേശ്യത്തിനായി ഒരു വസ്തുവിന് മൂല്യം ഉണ്ടാകുമ്പോൾ അത് സാധാരണയായി തൃപ്തികരമാണെന്ന് കണക്കാക്കപ്പെടുന്നു. തൽഫലമായി, വേദന ശമിപ്പിക്കൽ അല്ലെങ്കിൽ അണുബാധ നിയന്ത്രണം പോലുള്ള വിവിധ രോഗങ്ങൾക്ക് ഒരു പ്രത്യേക തരം ചികിത്സ ഫലപ്രദമാണെന്ന് കണക്കാക്കപ്പെടുന്നു, കൈകാര്യം ചെയ്യുന്ന അവസ്ഥയെ ആശ്രയിച്ച്. പെരുമാറ്റത്തെ സദ്ഗുണം എന്ന് വിളിക്കുമ്പോൾ, അത് നമ്മുടെ ലക്ഷ്യത്തിലെത്താൻ സഹായിക്കുമെന്ന് നമ്മൾ അർത്ഥമാക്കിയേക്കാം.

തൽഫലമായി, ധാർമ്മികതയുടെ അച്ചടക്കം മനുഷ്യ പെരുമാറ്റത്തിന്റെ നീതിയെയോ സദ്ഗുണത്തെയോ സംബന്ധിച്ചാണെന്ന് നമുക്ക് ഉറപ്പിച്ചു പറയാൻ കഴിയും.

സത്യാന്വേഷണത്തിന് ധാർമ്മികത ആവശ്യമാണ്, അറിവിനായുള്ള ആഗ്രഹം അതിന്റെ പഠനത്തെ ന്യായീകരിക്കുന്നു. പ്രവൃത്തി, പ്രചോദനം, ഉദ്ദേശ്യം, സ്വമേധയാ ഉള്ള പ്രവർത്തനം എന്നിവയുടെ കാരണങ്ങളുടെ അടിസ്ഥാന സ്വഭാവം ഇത് പരിശോധിക്കുന്നു. മനുഷ്യ പ്രവർത്തനങ്ങളുടെ ധാർമ്മികതയെ ഇത് വിലയിരുത്തുന്നു. മനുഷ്യ പെരുമാറ്റത്തിന്റെ ഉത്ഭവമോ പരിണാമമോ ഇത് പരിശോധിക്കുന്നില്ല. ധാർമ്മികതയുടെ ഒരു ശാഖ എന്ന നിലയിൽ, ധാർമ്മിക അവബോധത്തിന്റെ ഘടകങ്ങളെയും അതുമായി ബന്ധപ്പെട്ട വിവിധ പ്രശ്നങ്ങളെയും ധാർമ്മികത വിശകലനം ചെയ്യുന്നു. ആത്യന്തിക നന്മ അല്ലെങ്കിൽ സമ്പൂർണ്ണ നന്മ എന്ന ആശയത്തെയാണ് ധാർമ്മികത ബാധിക്കുന്നത്. അവകാശങ്ങൾ, കടമകൾ, നന്മ എന്നിവയുടെ അടിസ്ഥാന തത്വങ്ങളെ ഇത് വിശകലനം ചെയ്യുന്നു. നമ്മുടെ സ്വമേധയാ ഉള്ള പ്രവർത്തനങ്ങളുടെ ധാർമ്മിക വിലയിരുത്തൽ ധാർമ്മികതയുടെ

മണ്ഡലത്തിലാണ്. ധാർമ്മിക വിധിന്യായത്തെക്കുറിച്ചുള്ള ചർച്ച അതിന്റെ സത്ത, വിഷയം, ശേഷി, നിലവാരം എന്നിവയും പരിഗണിക്കണം. ധാർമ്മിക വിധിന്യായം പരിഗണിക്കുമ്പോൾ നമ്മുടെ മനസ്സിൽ ധാർമ്മിക വികാരങ്ങളും വികാരങ്ങളും ഉടലെടുക്കുന്നു; അതിനാൽ, ധാർമ്മികത ധാർമ്മിക വികാരങ്ങളുടെയും ധാർമ്മിക വിധിന്യായത്തിന്റെയും സ്വഭാവം അന്വേഷിക്കണം.

മനുഷ്യസ്വാതന്ത്ര്യത്തിന്റെ സ്വഭാവത്തെയാണ് ധാർമ്മികത അന്വേഷിക്കുന്നത്. നിതി, അനിതി എന്നിവയ്ക്കൊപ്പം നന്മയുടെയും തിന്മയുടെയും സത്തയെയും ധാർമ്മികത അന്വേഷിക്കുന്നു. സദ്ഗുണം, നിയമം, മനസ്സാക്ഷി, കടമ എന്നിവയുടെ സത്തയെയും ഇത് പരിശോധിക്കുന്നു. ഏതൊക്കെ കടമകളാണ് സാർവത്രികമായി അംഗീകരിക്കപ്പെട്ടിരിക്കുന്നത്? എല്ലാ പ്രശംസനീയമായ പ്രവൃത്തികളിലും അന്തർലീനമായ നന്മ എന്താണ്? ഈ അന്വേഷണങ്ങൾ ധാർമ്മികതയുടെ മേഖലയുമായി ബന്ധപ്പെട്ടിരിക്കുന്നു. കടമബോധം, അല്ലെങ്കിൽ ധാർമ്മിക ബാധ്യത, പ്രവൃത്തികളുടെ ഉത്തരവാദിത്തത്തോടൊപ്പം, ധാർമ്മികതയുമായി ബന്ധപ്പെട്ടിരിക്കുന്നു. സ്വമേധയാ ഉള്ള പെരുമാറ്റങ്ങളെ അവയുടെ ധാർമ്മിക മൂല്യത്തിനോ അത്തരം മനുഷ്യ പെരുമാറ്റങ്ങളുടെ അന്തർലീനമായ നീതിക്കോ അനുസൃതമായി ധാർമ്മികത വിലയിരുത്തുന്നു. ഒരു വ്യക്തി ഒരു കലാകാരനോ പ്രഭാഷകനോ ആയി അഭിവൃദ്ധി പ്രാപിക്കുകയും ധാർമ്മിക പോരായ്മകൾ പ്രകടിപ്പിക്കുകയും ചെയ്യാം, അല്ലെങ്കിൽ നേരെമറിച്ച്, ധാർമ്മികമായി സദ്ഗുണമുള്ള ഒരു വ്യക്തിക്ക് ഒരു കലാകാരനോ സാങ്കേതിക വിദഗ്ദ്ധനോ എന്ന നിലയിൽ പ്രാവീണ്യം ഇല്ലായിരിക്കാം. മനുഷ്യന്റെ പെരുമാറ്റത്തെ നിയന്ത്രിക്കുന്ന തത്വങ്ങളെ മാത്രം ധാർമ്മികത ബാധിക്കുന്നു, ഒരു സദ്ഗുണമുള്ള വ്യക്തിയുടെ സംസ്കാരം വളർത്തുന്നു, ഇത് സൂചിപ്പിക്കുന്നത് ഒരാൾക്ക് കലാപരമായ അല്ലെങ്കിൽ സാങ്കേതിക മേഖലകളിൽ മികവ് പുലർത്താൻ കഴിയുമെങ്കിലും, വ്യക്തിപരമായ സമഗ്രതയും സാമൂഹിക സംഭാവനയും യഥാർത്ഥ ധാർമ്മിക പെരുമാറ്റം അത്യന്താപേക്ഷിതമാണ് എന്നാണ്.

1. ജീവിക്കാനുള്ള അവകാശം - ആദ്യത്തെ മനുഷ്യാവകാശം ജീവിക്കാനുള്ള അവകാശമാണ്. ആത്മസാക്ഷാത്കാരം ആത്യന്തിക നന്മയാണ്, അത് നേടിയെടുക്കുന്നതിന് ജീവിതത്തിന്റെ തുടർച്ച ആവശ്യമാണ്. ജീവിക്കാനുള്ള അവകാശം അടിസ്ഥാന അവകാശമാണ്. ജീവിതത്തിന്റെ പവിത്രത അംഗീകരിക്കപ്പെടണം. ജീവിക്കാനുള്ള അവകാശം നമ്മുടെയും മറ്റുള്ളവരുടെയും നിലനിൽപ്പിനെ പവിത്രമായി കണക്കാക്കാനുള്ള ധാർമ്മിക കടമയെ ഉൾക്കൊള്ളുന്നു.

2. സ്വാതന്ത്ര്യത്തിനുള്ള അവകാശം - തുടർന്നുള്ള അവകാശം സ്വാതന്ത്ര്യമാണ്. ആത്മസാക്ഷാത്കാരം പരമമായ പുണ്യമാണ്. അത് ഒരു വ്യക്തിയുടെ ഇച്ഛാശക്തിയാൽ സാക്ഷാത്കരിക്കപ്പെടുന്നു. തന്റെ ആത്യന്തിക ലക്ഷ്യം നേടുന്നതിനായി സ്വയംഭരണം പ്രയോഗിക്കാൻ അയാൾക്ക് സ്വാതന്ത്ര്യമുണ്ടായിരിക്കണം. ഒരു ഘടനാപരമായ സമൂഹത്തിൽ, ഒരു വ്യക്തിക്ക് തന്റെ ഇച്ഛാശക്തിയുടെ അനിയന്ത്രിതമായ പ്രയോഗത്തിലൂടെ തന്റെ ആത്യന്തിക ലക്ഷ്യം നേടാൻ അനുവദിക്കപ്പെടണം,

മറ്റുള്ളവരുടെ അവകാശങ്ങളെ ലംഘിക്കാതെ കരിയർ പുരോഗതി അല്ലെങ്കിൽ സൃഷ്ടിപരമായ ആവിഷ്കാരം പോലുള്ള വ്യക്തിപരമായ ലക്ഷ്യങ്ങൾ പിന്തുടരുന്നത് പോലുള്ള സാമൂഹിക ക്രമത്തിന്റെ സംരക്ഷണവുമായി ഇത് യോജിക്കുന്നു. 3. ജോലി ചെയ്യാനുള്ള അവകാശം ജീവിക്കാനുള്ള മാലികാവകാശത്തിൽ നിന്നാണ് ഉരുത്തിരിഞ്ഞത്. ഒരു വ്യക്തിക്ക് തൊഴിൽ ഉറപ്പാക്കിയില്ലെങ്കിൽ, അയാൾക്ക് ഉപജീവനമാർഗ്ഗം നിലനിർത്താൻ കഴിയില്ല. തൊഴിലില്ലായ്മയോ തൊഴിലില്ലായ്മയോ വ്യക്തികൾക്ക് സ്വയം സാക്ഷാത്കരിക്കാനുള്ള അവസരം നിഷേധിക്കുന്നതിനാൽ, ഒരു സമകാലിക ക്ഷേമരാഷ്ട്രം എല്ലാ പൗരന്മാർക്കും പൂർണ്ണമായ തൊഴിൽ നൽകണം. തൊഴിൽ അവകാശം എല്ലാ ക്ഷേമരാഷ്ട്രങ്ങളും അംഗീകരിക്കേണ്ടതുണ്ട്.

4. തുടർന്നുള്ള അവകാശം വിദ്യാഭ്യാസത്തിനുള്ള അവകാശമാണ്. അവകാശങ്ങളും കടമകളും പരസ്പരം സങ്കീർണ്ണമായി ബന്ധപ്പെട്ടിരിക്കുന്നു. ഓരോ വ്യക്തിക്കും അവർക്ക് കഴിയുന്നത്ര ഉയർന്ന നിലവാരമുള്ള വിദ്യാഭ്യാസം നേടാനുള്ള അവകാശമുണ്ട്. തന്റെ കഴിവുകൾക്ക് അനുസൃതമായി ഉയർന്ന വിദ്യാഭ്യാസം നേടാൻ അയാൾ ധാർമ്മികമായി നിർബന്ധിതനാണ്. ഒരു വികസിത സമൂഹം വ്യക്തികൾക്ക് അവരുടെ കഴിവുകൾ തിരിച്ചറിയുന്നതിനും കൂട്ടായ ക്ഷേമത്തിന് സംഭാവന ചെയ്യുന്നതിനും പരമാവധി അവസരങ്ങൾ നൽകണം.

ധാർമ്മികതയുടെ പ്രായോഗിക വശത്തെ "പ്രയോഗിത ധാർമ്മികത" എന്ന് വിളിക്കുന്നു, ഇത് ധാർമ്മികതയുടെ സമഗ്രമായ അച്ചടക്കത്തിന്റെ ഒരു ഉപവിഭാഗമാണ്. ഇവ സാഹചര്യപരമായ ധാർമ്മികതയുടെ മേഖലയെ ഉൾക്കൊള്ളുന്നു, അതേസമയം മറ്റൊ-എത്തിക്സ് "സ്വാതന്ത്ര്യം", "നിർണ്ണയവാദം" എന്നിവയാൽ നമ്മൾ എന്താണ് അർത്ഥമാക്കുന്നത് തുടങ്ങിയ യുക്തിപരവും അർത്ഥപരവുമായ അന്വേഷണങ്ങളെ അടിസ്ഥാനമാക്കി ചെയ്യുന്നു. സാമൂഹ്യശാസ്ത്രം, രാഷ്ട്രീയ ശാസ്ത്രം, നിയമശാസ്ത്രം, നിയമം, മനുശാസ്ത്രം, നരവംശശാസ്ത്രം, സാംസ്കാരിക പഠനങ്ങൾ, പരിസ്ഥിതി ശാസ്ത്രം, പരിസ്ഥിതി പഠനങ്ങൾ, സാമ്പത്തിക ശാസ്ത്രം, മതം, സൗന്ദര്യശാസ്ത്രം, മറ്റ് അനുബന്ധ മേഖലകൾ എന്നിവയുൾപ്പെടെ മറ്റ് വിഷയങ്ങളുമായി ധാർമ്മികത അടിസ്ഥാനപരമായി പരസ്പരം

ബന്ധപ്പെട്ടിരിക്കുന്നു. ആത്യന്തിക നന്മയായുള്ള അന്വേഷണത്തിൽ രാഷ്ട്രീയ, സാമൂഹികശാസ്ത്ര, സാംസ്കാരിക, മനുഷ്യാശ്വ, സാമ്പത്തിക, പാരിസ്ഥിതിക, മതപരമായ വിഷയങ്ങളെ ധർമ്മികത അഭിസംബോധന ചെയ്യുന്നു. തൽഫലമായി, ധർമ്മികതയുടെ മേഖല ഈ പ്രശ്നങ്ങളെ ഉൾക്കൊള്ളുന്നു. പുതിയ സാങ്കേതികവിദ്യയുടെ ആവിർഭാവം ഉയർന്നുവരുന്ന ആശങ്കകളെ നേരിടുന്നതിന് ധർമ്മികതയുടെ മേഖല വികസിപ്പിക്കാനുള്ള അവസരം നൽകുന്നു.

അതിനാൽ, AI യുടെ വികസനത്തിൽ, മനുഷ്യരാശിയുടെ താൽപ്പര്യങ്ങൾ സംരക്ഷിക്കുന്നതിനുള്ള ഒരു പ്രാഥമിക മാർഗമായി സ്ഥാപിതമായ ധർമ്മിക തത്വങ്ങൾ നടപ്പിലാക്കുന്നതിന് മുൻഗണന നൽകേണ്ടത് അത്യന്താപേക്ഷിതമാണ്. കൃത്രിമബുദ്ധി (AI) പ്രാധാന്യം നേടിക്കൊണ്ടിരിക്കുന്നു. സാങ്കേതികവിദ്യ നമ്മുടെ നിലനിൽപ്പിന്റെ നിരവധി വശങ്ങളിലേക്ക് വ്യാപിക്കുന്നു, അതിന്റെ അനന്തരഫലങ്ങൾ കൈകാര്യം ചെയ്യാൻ നമ്മെ നിർബന്ധിക്കുന്നു. AI യുടെ വികസനവും നടപ്പാക്കലും നമ്മുടെ ജീവിതത്തെയും സമൂഹങ്ങളെയും പല തരത്തിൽ പരിവർത്തനം ചെയ്യുന്നു.

AI വ്യക്തമായ നേട്ടങ്ങൾ നൽകുമെങ്കിലും, പരിശീലനത്തിനായി ഉപയോഗിക്കുന്ന ഡാറ്റയുടെ ഏറ്റെടുക്കൽ, ഉപയോഗം, ദുരുപയോഗം, അതുപോലെ തന്നെ അൽഗോരിതം എന്നിവ വ്യക്തികളെ ഇതുവരെ തിരിച്ചറിയാത്ത അപകടങ്ങൾക്ക് വിധേയമാക്കിയേക്കാം. തൊഴിലുടമകൾക്ക് ഹാസ്യമായും അപ്രതീക്ഷിതമായും ജോലിസ്ഥലത്തെ പ്രകടനവും പെരുമാറ്റവും നിരീക്ഷിക്കാൻ കഴിയും, ഇത് ജീവനക്കാർക്കിടയിൽ സ്വകാര്യതയുടെയും വിശ്വാസത്തിന്റെയും ലംഘനത്തിലേക്ക് നയിച്ചേക്കാം. വിവരസാങ്കേതികവിദ്യ, അൽഗോരിതങ്ങൾ, ഡാറ്റാ രീതികൾ, പ്രത്യേകിച്ച് കൃത്രിമം എന്നിവ ഉയർത്തുന്ന ധർമ്മിക വെല്ലുവിളികളെ അഭിസംബോധന ചെയ്യുന്ന പ്രായോഗിക ധർമ്മികതയുടെ ശാഖയാണ് AI ധർമ്മികത.

ഇന്റലിജൻസ് സംവിധാനങ്ങൾ. സാങ്കേതികവിദ്യയുടെ ധർമ്മികവും പ്രയോജനകരവുമായ പുരോഗതി, നടപ്പാക്കൽ, ഉപയോഗം എന്നിവയ്ക്കുള്ള നിയന്ത്രണങ്ങളും മാനദണ്ഡങ്ങളും സ്ഥാപിക്കുക എന്നതാണ് ലക്ഷ്യം. AI സാങ്കേതികവിദ്യകൾ വികസിപ്പിക്കുകയും ഉത്തരവാദിത്തത്തോടെ ഉപയോഗിക്കുകയും ചെയ്യുന്നുവെന്ന് ഉറപ്പാക്കുന്നതിന് അത്യാവശ്യമായ സുതാര്യത, ഉത്തരവാദിത്തം, ന്യായബോധം, സ്വകാര്യത സംരക്ഷണം തുടങ്ങിയ വിശ്വസനീയമായ AI സൃഷ്ടിക്കുന്നതിന് വഴികാട്ടുന്നതിനായി രൂപകൽപ്പന ചെയ്തിരിക്കുന്ന അടിസ്ഥാന തത്വങ്ങളുടെ ഒരു കൂട്ടത്തിന് ചുറ്റും AI ധർമ്മികതയ്ക്കുള്ള വിവിധ ചട്ടങ്ങളുകൾ സംയോജിക്കുന്നു.

മനുഷ്യ മൂല്യങ്ങൾ, ഉദ്ദേശ്യങ്ങൾ, താൽപ്പര്യങ്ങൾ എന്നിവയുമായി സ്വയംഭരണ AI സിസ്റ്റങ്ങളുടെ ലക്ഷ്യങ്ങൾ സമന്വയിപ്പിക്കുന്നതിലൂടെയാണ് AI അലൈൻമെന്റ് സാധ്യമാകുന്നത്. അങ്ങനെ, ഈ സംവിധാനങ്ങൾ സ്വതന്ത്രമായി പ്രവർത്തിക്കുന്നുണ്ടെന്ന് ഉറപ്പുവരുത്താൻ, മനുഷ്യ മുൻഗണനകളെ പ്രതിഫലിപ്പിക്കുന്നു. വിവിധ അടിസ്ഥാന രീതികളിലൂടെ ഇത് സാധ്യമാകും.

വ്യക്തമായ ലക്ഷ്യവും മൂല്യ വിവരണവും

വ്യക്തമായ ലക്ഷ്യവും മൂല്യ സ്പെസിഫിക്കേഷനും AI അലൈൻമെന്റിൽ, സ്വയംഭരണ AI സിസ്റ്റങ്ങളുടെ ലക്ഷ്യങ്ങളെ മാനുഷിക മൂല്യങ്ങൾ, ഉദ്ദേശ്യങ്ങൾ, താൽപ്പര്യങ്ങൾ എന്നിവയുമായി സമന്വയിപ്പിക്കുകയും, ഈ സംവിധാനങ്ങൾ സ്വതന്ത്രമായി പ്രവർത്തിക്കുന്നുണ്ടെന്ന് ഉറപ്പാക്കുകയും, അതോടൊപ്പം നമ്മുടെ മുൻഗണനകളെയും പ്രതിഫലിപ്പിക്കുകയും ചെയ്യുന്നു. വിവിധ അടിസ്ഥാന രീതികളിലൂടെ ഇത് നേടിയെടുക്കാൻ കഴിയും.

കൃത്യമായ ലക്ഷ്യ നിർവ്വചനം ഒരു AI സിസ്റ്റത്തിന്റെ പ്രതീക്ഷിക്കുന്ന പ്രവർത്തനങ്ങൾ, അതിന്റെ പ്രവർത്തനങ്ങളെ നിയന്ത്രിക്കുന്ന നിയന്ത്രണങ്ങൾ, അത് ഒഴിവാക്കേണ്ട ഫലങ്ങൾ എന്നിവയുടെ വ്യക്തവും അസന്ദിഗ്ധവുമായ ആവിഷ്കാരത്തെ സൂചിപ്പിക്കുന്നു. നൽകിയിരിക്കുന്ന ഇൻപുട്ടുകളെ അടിസ്ഥാനമാക്കി സ്വയംഭരണ AI സിസ്റ്റങ്ങൾ ഒപ്റ്റിമൈസ് ചെയ്യപ്പെടുന്നതിനാൽ, മനുഷ്യന്റെ ലക്ഷ്യങ്ങളുമായി പൊരുത്തപ്പെടുന്നതിന് കൃത്യത നിർണായകമാണ്.

മാനുഷിക മൂല്യങ്ങൾ ഉൾച്ചേർക്കുന്നത്, സാങ്കേതിക കാര്യക്ഷമതയിലോ പ്രകടനത്തിലോ മാത്രം ശ്രദ്ധ കേന്ദ്രീകരിക്കുന്നതിനപകരം, അടിസ്ഥാന മാനുഷിക മാനദണ്ഡങ്ങൾ, ധർമ്മികത, സാമൂഹിക താൽപ്പര്യങ്ങൾ എന്നിവ തുടർച്ചയായി മാനിക്കുന്ന AI സംവിധാനങ്ങളുടെ സൃഷ്ടിയെ അർത്ഥമാക്കുന്നു.

അടിസ്ഥാന മാനുഷിക മൂല്യങ്ങൾ തിരിച്ചറിയൽ

സാധാരണയായി ഊന്നിപ്പറയുന്ന മൂല്യങ്ങളിൽ ഇവ ഉൾപ്പെടുന്നു:

•സുരക്ഷയും ദോഷരഹിതതയും

സുരക്ഷയും ദോഷരഹിതതയും വ്യക്തികൾക്കും സമൂഹത്തിനും ശാരീരികമോ മാനസികമോ സാമൂഹികമോ സാമ്പത്തികമോ ആയ ദോഷങ്ങൾ വരുത്തുന്നതിൽ നിന്ന് AI സംവിധാനങ്ങൾ വിട്ടുനിൽക്കണമെന്ന് ആവശ്യപ്പെടുന്ന ഒരു അടിസ്ഥാന മാനുഷിക മൂല്യമാണ് ഇത്. AI വിന്യാസത്തിൽ, ഈ മൂല്യം സ്ഥിരമായി കാര്യക്ഷമത, വേഗത അല്ലെങ്കിൽ ലാഭം എന്നിവയെ മറികടക്കുന്നു. AI വിന്യാസത്തിലെ ഭൗതിക സുരക്ഷ എന്നത് AI സംവിധാനങ്ങൾ മൂലമുണ്ടാകുന്ന പരിക്കുകൾ, ജീവഹാനി അല്ലെങ്കിൽ മൂർത്തമായ ആസ്തികൾക്ക് കേടുപാടുകൾ എന്നിവ തടയുന്നതിനെ സൂചിപ്പിക്കുന്നു. ഇത് പരമപ്രധാനമായ ആവശ്യകതയാണ്, പ്രത്യേകിച്ച് സ്വയംഭരണാധികാരമുള്ളതും ഉയർന്ന അപകടസാധ്യതയുള്ളതുമായ AI സംവിധാനങ്ങൾ. AI സംവിധാനങ്ങൾ മാനസിക ക്ലേശങ്ങൾ, വൈകാരിക അസ്വസ്ഥതകൾ അല്ലെങ്കിൽ സാമൂഹിക ഒഴിവാക്കൽ എന്നിവ ഉണ്ടാക്കുന്നില്ലെന്നും സാമൂഹിക ബന്ധങ്ങളെയും സ്ഥാപനങ്ങളെയും ദുർബ്ബലപ്പെടുത്തുന്നില്ലെന്നും ഉറപ്പുനൽകുന്നതാണ് മനുഷാസ്ത്രപരവും സാമൂഹികവുമായ സുരക്ഷ. ശാരീരിക പരിക്കിന്റെ അഭാവത്തിൽ പോലും, ദോഷകരമായ സ്റ്റീരിയോടൈപ്പുകൾ നിലനിർത്തുക, തെറ്റായ വിവരങ്ങൾ പ്രചരിപ്പിക്കുക, അല്ലെങ്കിൽ ഉപയോക്താക്കൾക്കിടയിൽ ഉത്കണ്ഠയും അവിശ്വാസവും വളർത്തുന്ന അന്തരീക്ഷം സൃഷ്ടിക്കുക തുടങ്ങിയ വഴികളിലൂടെ വ്യക്തികൾക്കും സമൂഹത്തിനും AI കാര്യമായ നാശനഷ്ടങ്ങൾ വരുത്തിവയ്ക്കും. AI സംവിധാനങ്ങൾ ഇവയിൽ നിന്ന് വിട്ടുനിൽക്കണം:

•വൈകാരിക കൃത്രിമത്വം

•വഞ്ചനാപരമായ പ്രേരണ

•മനുഷ്യന്റെ ദുർബ്ബലതകളെ ചൂഷണം ചെയ്യൽ

•നിതിയും സമത്വവും

•മനുഷ്യന്റെ അന്തസ്സം സ്വയംഭരണവും

•സ്വകാര്യതയും സമ്മതവും

2. മാനുഷിക മൂല്യങ്ങൾ പഠിക്കുക

മനുഷ്യന്റെ പെരുമാറ്റം, തിരുമാനമെടുക്കൽ, സാമൂഹിക ഇടപെടലുകൾ എന്നിവയെ നയിക്കുന്ന അവശ്യ തത്വങ്ങളും ധാർമ്മിക മാനദണ്ഡങ്ങളുമാണ് മാനുഷിക മൂല്യങ്ങൾ. AI വിന്യാസത്തിന്റെ മേഖലയിൽ, AI സിസ്റ്റങ്ങളിൽ നിന്ന് പ്രതീക്ഷിക്കുന്ന പെരുമാറ്റവും ലക്ഷ്യങ്ങൾ കൈവരിക്കാൻ ശ്രമിക്കുമ്പോൾ അവ പാലിക്കേണ്ട തത്വങ്ങളും അവ നിർവ്വചിക്കുന്നു. മാനുഷിക മൂല്യങ്ങൾ നേടുന്നത്, വഴക്കമില്ലാത്തതും മുൻകൂട്ടി സ്ഥാപിതമായതുമായ നിയമങ്ങളെ മാത്രം ആശ്രയിക്കുന്നതിനുപകരം, മനുഷ്യന്റെ പെരുമാറ്റത്തെയും ഇൻപുട്ടിനെയും അടിസ്ഥാനമാക്കിയുള്ള മനുഷ്യന്റെ മുൻഗണനകൾ, കൺവെൻഷനുകൾ, ധാർമ്മിക വിധിന്യായങ്ങൾ എന്നിവ അനുമാനിക്കാനും ക്രമീകരിക്കാനും ബഹുമാനിക്കാനും AI സിസ്റ്റങ്ങളെ സജ്ജമാക്കുന്നതിനെയാണ് അർത്ഥമാക്കുന്നത്. ഈ രീതി യഥാർത്ഥ മനുഷ്യ ലക്ഷ്യങ്ങളുമായി AI വിന്യാസം മെച്ചപ്പെടുത്തുന്നു.

മനുഷ്യ അന്തസ്സിനോടുള്ള ബഹുമാനം ഓരോ വ്യക്തിയുടെയും പ്രയോജനം, പദവി അല്ലെങ്കിൽ കഴിവ് എന്നിവ പരിഗണിക്കാതെ, അവരുടെ അന്തർലീനമായ മൂല്യത്തെ അംഗീകരിക്കുന്ന ഒരു അടിസ്ഥാന നൈതിക തത്വമാണ്. നൈതികതയിൽ, പ്രത്യേകിച്ച് AI നൈതികതയിൽ, വ്യക്തികളെ ഒരിക്കലും ഒരു ലക്ഷ്യത്തിലേക്കുള്ള ഉപകരണങ്ങളായോ, ഡാറ്റാ പോയിന്റുകളായോ, മാർഗങ്ങളായോ മാത്രം കണക്കാക്കരുതെന്നാണ് ഇത് സൂചിപ്പിക്കുന്നത്.

സ്വാതന്ത്ര്യവും സ്വയംഭരണവും ഒരു വ്യക്തിയുടെ സ്വാതന്ത്ര്യത്തെയും അവരുടെ ജീവിതത്തെക്കുറിച്ചുള്ള സ്വയംഭരണപരവും, അറിവുള്ളതും,

സ്വമേധയാ ഉള്ളതുമായ തീരുമാനങ്ങൾ എടുക്കാനുള്ള കഴിവിനെയും അംഗീകരിക്കുന്ന അടിസ്ഥാന ധാർമ്മിക തത്വങ്ങൾ രൂപപ്പെടുത്തുന്നു. ധാർമ്മികതയിൽ, പ്രത്യേകിച്ച് AI ധാർമ്മികതയിൽ, മനുഷ്യ ഏജൻസിയോടുള്ള സംരക്ഷണവും ബഹുമാനവും പരമപ്രധാനമാണ്, കാരണം ഇത് വ്യക്തികൾ അവരുടെ തിരഞ്ഞെടുപ്പുകളിൽ നിയന്ത്രണം നിലനിർത്തുന്നുവെന്നും ഓട്ടോമേറ്റഡ് സിസ്റ്റങ്ങളുടെ കൃത്രിമ സ്വാധീനങ്ങൾക്ക് വിധേയരാകുന്നില്ലെന്നും ഉറപ്പാക്കുന്നു.

തുല്യതയും നീതിയും വ്യക്തികൾക്ക് ന്യായവും പക്ഷപാതരഹിതവും വിവേചനരഹിതവുമായ പരിഗണന ലഭിക്കേണ്ട അടിസ്ഥാന നൈതിക തത്വങ്ങളാണ്. നൈതികതയിലും AI നൈതികതയിലും, ഗുണങ്ങളും ദോഷങ്ങളും ധാർമ്മികമായി സ്വീകാര്യമായ രീതിയിൽ വിതരണം ചെയ്യുന്നുവെന്ന് അവ ഉറപ്പുനൽകുന്നു, ഇത് ഉറപ്പാക്കുന്നു

സാങ്കേതിക പുരോഗതി അരികുവൽക്കരിക്കപ്പെട്ട വിഭാഗങ്ങളെ അനുപാതമില്ലാതെ ബാധിക്കുന്നില്ലെന്നും എല്ലാ വ്യക്തികൾക്കും AI യുടെ നേട്ടങ്ങൾ തുല്യമായി ലഭ്യമാകുമെന്നും.

തുല്യ ധാർമ്മിക മൂല്യം

വംശം, ലിംഗഭേദം, ജാതി, മതം, സമ്പത്ത് അല്ലെങ്കിൽ കഴിവ് എന്നിവ കണക്കിലെടുക്കാതെ എല്ലാ മനുഷ്യർക്കും തുല്യ ധാർമ്മിക സ്ഥാനങ്ങളുണ്ട്, തുല്യ ബഹുമാനം അർഹിക്കുന്നു.

•നീതിയും വിവേചനമില്ലായ്മയും: വംശം, ലിംഗഭേദം, പ്രായം, മതം തുടങ്ങിയ സെൻസിറ്റീവ് ഗുണങ്ങളെ അടിസ്ഥാനമാക്കിയുള്ള വിവേചനപരമായ ഫലങ്ങളിലേക്ക് നയിക്കുന്ന അന്യായമായ പക്ഷപാതങ്ങൾ ഒഴിവാക്കിക്കൊണ്ട്, എല്ലാ വ്യക്തികൾക്കും തുല്യമായ പരിഗണന AI സംവിധാനങ്ങൾ ഉറപ്പാക്കണം. ബാഹ്യ ഗുണങ്ങൾ പരിഗണിക്കാതെ, തീരുമാനങ്ങൾ എല്ലാ ആളുകളെയും ഗ്രൂപ്പുകളെയും ന്യായമായി പരിഗണിക്കണം.

തുല്യ പരിഗണന എന്നത് താരതമ്യപ്പെടുത്താവുന്ന സാഹചര്യങ്ങളിൽ വ്യക്തികൾക്ക് ഒരേ തരത്തിലുള്ള പരിഗണന ലഭിക്കണമെന്ന് നിഷ്കർഷിക്കുന്ന ഒരു ധാർമ്മിക ആദർശമാണ്, അന്യായമായ പക്ഷപാതമോ പക്ഷപാതമോ ഇല്ലാതെ. AI ധാർമ്മികതയിൽ, വംശം, ലിംഗഭേദം അല്ലെങ്കിൽ സാമൂഹിക സാമ്പത്തിക നില പോലുള്ള ബാഹ്യമായ വ്യക്തിഗത ഗുണങ്ങളെ അടിസ്ഥാനമാക്കി ഓട്ടോമേറ്റഡ് തിരഞ്ഞെടുപ്പുകൾ വ്യക്തികളോട് വിവേചനം കാണിക്കുന്നില്ലെന്ന് ഇത് ഉറപ്പുനൽകുന്നു, ഇത് തീരുമാനമെടുക്കൽ പ്രക്രിയകളിൽ നീതി ഉറപ്പാക്കുന്നു.

നിഷ്പക്ഷമായ തീരുമാനമെടുക്കൽ എന്നത് ഒരു ധാർമ്മിക ആദർശമാണ്, അത് തിരഞ്ഞെടുപ്പുകൾ വസ്തുനിഷ്ഠമായും, തുല്യമായും, പക്ഷപാതരഹിതമായും, പ്രസക്തമായ വസ്തുതകളെയും മാനദണ്ഡങ്ങളെയും മാത്രം ആശ്രയിച്ച് നടത്തണമെന്ന് ആവശ്യപ്പെടുന്നു. AI ധാർമ്മികതയിൽ, സിസ്റ്റങ്ങൾ വ്യക്തികളെയോ ഗ്രൂപ്പുകളെയോ അന്യായമായി അനുകൂലിക്കുകയോ പ്രതികൂലിക്കുകയോ ചെയ്യുന്നില്ലെന്ന് ഇത് ഉറപ്പുനൽകുന്നു.

•സുതാര്യതയും വിശദീകരണക്ഷമതയും(XAI): സുതാര്യതയും വിശദീകരണക്ഷമതയും AI ധാർമ്മികതയുടെയും വിന്യാസത്തിന്റെയും അടിസ്ഥാന തത്വങ്ങളാണ്, ഇത് AI സിസ്റ്റങ്ങളെ മനസ്സിലാക്കാവുന്നതും വ്യാഖ്യാനിക്കാവുന്നതും മനുഷ്യർക്ക് ഉത്തരവാദിത്തമുള്ളതാക്കുന്നു. AI വിധിന്യായങ്ങൾ സുതാര്യവും യുക്തിസഹവും പരിശോധനയ്ക്ക് വിധേയവുമാണെന്ന് അവ ഉറപ്പു നൽകുന്നു.

വ്യാഖ്യാനിക്കാവുന്നതും വിശദീകരിക്കാവുന്നതുമായ AI സൃഷ്ടിക്കുന്നത് AI വിധിന്യായങ്ങളെക്കുറിച്ചുള്ള മനുഷ്യ ഗ്രാഹ്യത്തെ സുഗമമാക്കുകയും അലൈൻമെന്റ് പ്രശ്നങ്ങൾ നേരത്തേ കണ്ടെത്തുകയും ചെയ്യുന്നു. AI സിസ്റ്റങ്ങളുടെ പ്രവർത്തന സംവിധാനങ്ങളും (സുതാര്യത) ചില തീരുമാനങ്ങൾക്കോ ഔട്ട്പുട്ടുകൾക്കോ പിന്നിലെ യുക്തിയും (വിശദീകരിക്കൽ) ഡെവലപ്പർമാരും ഉപഭോക്താക്കളും മനസ്സിലാക്കേണ്ടത് അത്യാവശ്യമാണ്. സങ്കീർണ്ണമായ മോഡലുകൾ മനുഷ്യന്റെ ധാരണയിൽ നിന്ന് ഒഴിഞ്ഞുമാറുന്ന "ബ്ലാക്ക് ബോക്സ്" പ്രശ്നത്തെ ഇത് അടിസംബോധന ചെയ്യുന്നു.

തീരുമാനമെടുക്കലിൽ ഒരു AI സിസ്റ്റം ഉപയോഗിക്കുന്ന യുക്തി, രീതിശാസ്ത്രങ്ങൾ, മാനദണ്ഡങ്ങൾ എന്നിവ വ്യക്തവും മനസ്സിലാക്കാവുന്നതും വ്യക്തികൾക്ക് - പ്രത്യേകിച്ച് ഉപയോക്താക്കൾ, ഓഡിറ്റർമാർ, റെഗുലേറ്റർമാർ എന്നിവർക്ക് - എത്രത്തോളം ലഭ്യമാണെന്ന് അൽഗോരിതമിക് സുതാര്യത സൂചിപ്പിക്കുന്നു.

•ഉത്തരവാദിത്തം: വ്യക്തമായ ഉത്തരവാദിത്ത രേഖകൾ സ്ഥാപിക്കപ്പെടണം. AI സിസ്റ്റങ്ങൾക്ക് നിയമപരമോ ധാർമ്മികമോ ആയ

ഉത്തരവാദിത്തം വഹിക്കാൻ കഴിയാത്തതിനാൽ, സിസ്റ്റത്തിന്റെ ഫലങ്ങളുടെയും അതിന്റെ ഫലമായുണ്ടാകുന്ന ഏതൊരു ദോഷത്തിന്റെയും ഉത്തരവാദിത്തം മനുഷ്യർ (ഡിസൈനർമാർ, ഡെവലപ്പർമാർ, ഡിപ്ലോയർമാർ) ഏറ്റെടുക്കണം.

•സ്വകാര്യതയും ഡാറ്റാ ഭരണവും: ഒരു വ്യക്തിയുടെ സ്വകാര്യതയ്ക്കുള്ള അവകാശത്തെയും അവരുടെ വ്യക്തിപരമായി തിരിച്ചറിയാൻ കഴിയുന്ന വിവരങ്ങളുടെ (PII) നിയന്ത്രണത്തെയും മാനിച്ചുകൊണ്ട്, ഡാറ്റ ശേഖരിക്കുകയും സംഭരിക്കുകയും പ്രോസസ്സ് ചെയ്യുകയും സുരക്ഷിതമായും ഉത്തരവാദിത്തത്തോടെയും ഉപയോഗിക്കുകയും ചെയ്യുന്നുവെന്ന് ഉറപ്പാക്കാൻ കർശനമായ നിയമങ്ങൾ നടപ്പിലാക്കണം. ഡാറ്റ മിനിമൈസേഷൻ എന്നത് സ്വകാര്യതയുടെയും ഡാറ്റ ഭരണത്തിന്റെയും ഒരു അടിസ്ഥാന തത്വമാണ്, ഒരു പ്രത്യേക ആവശ്യത്തിന് ആവശ്യമായ ഡാറ്റ മാത്രമേ ശേഖരിക്കുകയും പ്രോസസ്സ് ചെയ്യുകയും നിലനിർത്തുകയും ചെയ്യാവൂ എന്ന് ഇത് ഉറപ്പിക്കുന്നു.

സ്വകാര്യതയിലും ഡാറ്റാ ഭരണത്തിലും വിവരമറിയിച്ചുള്ള സമ്മതം നിർണ്ണായകമായ ഒരു നൈതികവും നിയമപരവുമായ തത്വമാണ്, അതിനാൽ വ്യക്തികൾ അവരുടെ സ്വകാര്യ ഡാറ്റയുടെ ശേഖരണം, ഉപയോഗം, സംഭരണം, പങ്കിടൽ എന്നിവയെക്കുറിച്ച് സമഗ്രമായി മനസ്സിലാക്കുകയും അത്തരം രീതികൾക്ക് സജീവമായി സമ്മതം നൽകുകയും വേണം. സ്വകാര്യതയുടെയും ഡാറ്റാ ഭരണത്തിന്റെയും അടിസ്ഥാന തത്വമാണ് രഹസ്യാത്മകത, അത് അനധികൃത ആക്സസ്, വെളിപ്പെടുത്തൽ, ദുരുപയോഗം അല്ലെങ്കിൽ ലംഘനങ്ങൾ എന്നിവയിൽ നിന്ന് വ്യക്തിപരവും സെൻസിറ്റീവുമായ വിവരങ്ങൾ സംരക്ഷിക്കണമെന്ന് നിർബന്ധമാക്കുന്നു.

•സുരക്ഷയും പ്രതിരോധശേഷിയും: സാധാരണവും അപ്രതീക്ഷിതവുമായ സാഹചര്യങ്ങളിൽ വിശ്വസനീയമായും സുരക്ഷിതമായും പ്രവചനാതീതമായും പ്രവർത്തിക്കാൻ AI സംവിധാനങ്ങൾ രൂപകൽപ്പന ചെയ്തിരിക്കണം, അവ ശാരീരികമോ മാനസികമോ സാമൂഹികമോ ആയ ദോഷങ്ങൾ ഉണ്ടാക്കുന്നില്ലെന്ന് ഉറപ്പാക്കണം.

• മനുഷ്യ ഏജൻസിയും മേൽനോട്ടവും: പ്രത്യേകിച്ച് നിർണ്ണായക സാഹചര്യങ്ങളിൽ, മനുഷ്യന്റെ അധികാരം, സ്വയംഭരണം അല്ലെങ്കിൽ തിരുമാനമെടുക്കൽ എന്നിവയിൽ വിടുവീഴ്ച ചെയ്യുന്നതിനോ മാറ്റിസ്ഥാപിക്കുന്നതിനോ പകരം മനുഷ്യന്റെ കഴിവുകൾ വർദ്ധിപ്പിക്കുന്ന ഒരു ഉപകരണമായി AI പ്രവർത്തിക്കണം. AI ധർമ്മീകതയിലും വിന്യാസത്തിലും മനുഷ്യന്റെ ഉത്തരവാദിത്തം, മേൽനോട്ടം, AI സിസ്റ്റങ്ങളുടെ പ്രവർത്തനങ്ങളിൽ ഇടപെടാനുള്ള കഴിവ് എന്നിവ ഉറപ്പുനൽകുന്ന ഒരു അനിവാര്യ തത്വമാണിത്. ദോഷകരമോ മനുഷ്യർവമല്ലാത്തതോ ആയ പ്രവർത്തനങ്ങൾ നടത്തുന്നതിൽ നിന്ന് സ്വയംഭരണ AI-യെ ഇത് തടയുന്നു.

ഹ്യൂമൻ-ഇൻ-ദി-ലൂപ്പ് അല്ലെങ്കിൽ ഹ്യൂമൻ-ഓൺ-ദി-ലൂപ്പ് സിസ്റ്റങ്ങൾ സംരക്ഷിക്കുന്നത്, ആവശ്യമുള്ളപ്പോൾ വ്യക്തികൾക്ക് AI (ആർട്ടിഫിഷ്യൽ ഇന്റലിജൻസ്) മേൽനോട്ടം വഹിക്കാനോ ഇടപെടാനോ നിർജ്ജീവമാക്കാനോ കഴിയുമെന്ന് ഉറപ്പാക്കുന്നു. മാനുഷിക മൂല്യങ്ങളും സാമൂഹിക സാഹചര്യങ്ങളും വികസിക്കുമ്പോൾ, കാലക്രമേണ വിന്യാസം നിലനിർത്തുന്നതിന് AI സിസ്റ്റങ്ങൾ പതിവായി ഓഡിറ്റുകൾക്കും അപ്ഡേറ്റുകൾക്കും വിധേയമാകണം. ഉത്തരവാദിത്തവും മനുഷ്യാവകാശ സംരക്ഷണവും ഉറപ്പാക്കിക്കൊണ്ട് AI വികസനം ധർമ്മീക തത്വങ്ങൾ, നിയമ ചട്ടങ്ങളുകൾ, ജനാധിപത്യ മാനദണ്ഡങ്ങൾ എന്നിവ പാലിക്കണം.

മനുഷ്യ അഭിലാഷങ്ങളുമായി സ്വയംഭരണ AI സിസ്റ്റങ്ങളുടെ ലക്ഷ്യങ്ങൾ വിന്യസിക്കുക എന്നത് കൃത്രിമബുദ്ധിയുടെ പുരോഗതിയിൽ ഒരു പ്രധാന ബുദ്ധിമുട്ടാണ്. AI സിസ്റ്റങ്ങൾ കൂടുതൽ ശക്തവും സ്വയംഭരണാധികാരവുമാകുമ്പോൾ, അവയുടെ പ്രവർത്തനങ്ങൾ മാനുഷിക മൂല്യങ്ങൾ, സുരക്ഷ, ക്ഷേമം എന്നിവയുമായി യോജിക്കുന്നുവെന്ന് ഉറപ്പാക്കേണ്ടത് അത്യാവശ്യമാണ്. വ്യക്തമായ ധർമ്മീക തത്വങ്ങൾ, സുരക്ഷാ പ്രോട്ടോക്കോളുകൾ, മാനുഷിക ആവശ്യങ്ങൾ എന്നിവ ഉൾക്കൊള്ളുന്ന AI സിസ്റ്റങ്ങളുടെ സൂക്ഷ്മമായ രൂപകൽപ്പനയിലൂടെ ഈ വിന്യാസം കൈവരിക്കാനാകും.

കേന്ദ്രീകൃത ലക്ഷ്യങ്ങൾ. കൃത്രിമബുദ്ധിയെ മനുഷ്യ ലക്ഷ്യങ്ങളുമായി യോജിപ്പിക്കുന്നതിന് ധർമ്മീക രൂപകൽപ്പന, മനുഷ്യ മേൽനോട്ടം, സുതാര്യത, അന്താരാഷ്ട്ര സഹകരണം എന്നിവ ആവശ്യമാണ്, ഇത് സ്വയംഭരണ AI സാങ്കേതികവിദ്യകൾ മനുഷ്യരാശിയുടെ ഭാവിക്ക് പ്രയോജനകരമായ ഫലങ്ങൾ നൽകുന്നുവെന്ന് ഉറപ്പാക്കുന്നു.

നിങ്ങളുടെ ബുദ്ധിയെ ഉണർത്തൂ

1 AI വിന്യാസത്തിന്റെ പ്രധാന ലക്ഷ്യം എന്താണ്?

എ) AI സിസ്റ്റങ്ങളെ കൂടുതൽ വേഗതയേറിയതും കാര്യക്ഷമവുമാക്കുന്നതിന്

ബി) AI സംവിധാനങ്ങൾ മാനുഷിക മൂല്യങ്ങൾക്കും ഉദ്ദേശ്യങ്ങൾക്കും അനുസൃതമായി പ്രവർത്തിക്കുന്നുവെന്ന് ഉറപ്പാക്കാൻ

സി) മേൽനോട്ടമില്ലാതെ AI സ്വയംഭരണം വർദ്ധിപ്പിക്കുന്നതിന്

ഡി) AI വികസനത്തിന്റെ ചെലവ് കുറയ്ക്കുന്നതിന്

2. AI സിസ്റ്റങ്ങളിൽ മാനുഷിക മൂല്യങ്ങൾ ഉൾപ്പെടുത്തുന്നത് പ്രധാനമായിരിക്കുന്നത് എന്തുകൊണ്ട്?

എ) AI സിസ്റ്റങ്ങളെ കൂടുതൽ വേഗത്തിലും ലാഭകരമായും മാറ്റുന്നതിന്

ബി) ധാർമിക നിയന്ത്രണങ്ങളില്ലാതെ AI സ്വയംഭരണം വർദ്ധിപ്പിക്കുന്നതിന്

സി) മനുഷ്യ മേൽനോട്ടത്തിന്റെ ആവശ്യകത കുറയ്ക്കുന്നതിന്

ഡി) മനുഷ്യന്റെ അന്തസ്സ്, സുരക്ഷ, നീതി, ക്ഷേമം എന്നിവയെ AI ബഹുമാനിക്കുന്നുണ്ടെന്ന് ഉറപ്പാക്കുക.

3. AI സിസ്റ്റങ്ങളുടെ തുടർച്ചയായ നിരീക്ഷണം പ്രധാനമാണ് കാരണം:

എ) കാലക്രമേണ മനുഷ്യ-അനുയോജ്യമായ പെരുമാറ്റത്തിൽ നിന്ന് AI സംവിധാനങ്ങൾ വ്യതിചലിച്ചേക്കാം.

ബി) AI ലക്ഷ്യങ്ങൾ ഒരിക്കലും മാറില്ല.

സി) മനുഷ്യർക്ക് എല്ലാ ഉത്തരവാദിത്തവും AI-യെ ഏൽപ്പിക്കാൻ കഴിയും

ഡി) നിരീക്ഷണം AI കാര്യക്ഷമതയെ മെച്ചപ്പെടുത്തുന്നതിനായിരിക്കണം

4. AI-യിലെ സുതാര്യതയുടെയും വിശദീകരണത്തിന്റെയും ഉദ്ദേശ്യം എന്താണ്?

A) ഉപയോക്താക്കളിൽ നിന്ന് AI തീരുമാനമെടുക്കൽ മറയ്ക്കാൻ

ബി) AI തീരുമാനങ്ങൾ മനസ്സിലാക്കാവുന്നതും, ഉത്തരവാദിത്തമുള്ളതും, വിശ്വസനീയവുമാക്കുന്നതിന്

സി) AI വേഗത്തിൽ പ്രവർത്തിക്കുന്നതിന്

ഡി) മാനുഷിക മൂല്യങ്ങൾ പഠിക്കുന്നതിൽ നിന്ന് AI തടയുന്നതിന്

5. AI സിസ്റ്റങ്ങളിലെ സുതാര്യതയുടെ പ്രാഥമിക ലക്ഷ്യം എന്താണ്?

എ) ഉപയോക്താക്കളിൽ നിന്ന് AI തീരുമാനമെടുക്കൽ പ്രക്രിയകൾ മറയ്ക്കാൻ

ബി) AI പ്രോസസ്സിംഗ് വേഗത വർദ്ധിപ്പിക്കുന്നതിന്

സി) AI തീരുമാനമെടുക്കൽ ദൃശ്യവും മനസ്സിലാക്കാവുന്നതുമായി മാറ്റുന്നതിന്

ഡി) മനുഷ്യന്റെ മേൽനോട്ടമില്ലാതെ സ്വതന്ത്രമായി പ്രവർത്തിക്കാൻ AI-യെ അനുവദിക്കുക.

നിങ്ങളുടെ കഴിവുകൾ പുറത്തെടുക്കുക

1. AI ലക്ഷ്യങ്ങളെ മാനുഷിക മൂല്യങ്ങളുമായി യോജിപ്പിക്കുന്നതിന്റെ പ്രാധാന്യം എന്താണ്? 2. AI സംവിധാനങ്ങൾ മനുഷ്യ ലക്ഷ്യങ്ങളുമായി യോജിച്ച് പ്രവർത്തിക്കുന്നുവെന്ന് ഉറപ്പാക്കുന്നതുമായി ബന്ധപ്പെട്ട ബുദ്ധിമുട്ടുകൾ പരിശോധിക്കുക.