විශ්වාසදායක කෘතිම බුද්ධිය(TAI):

මිනිස් විභවය අගුළු හැරීම විශ්වාසය සමඟ ආරම්භ වේ...

විශ්වාසදායක AI යනු කුමක්ද?

විශ්වාසදායක AI යනු පැහැදිලි කළ හැකි, සාධාරණ, අර්ථ නිරූපණය කළ හැකි, ශක්තිමත්, විනිවිද පෙනෙන, ආරක්ෂිත සහ ආරක්ෂිත කෘතිම බුද්ධි පද්ධති වේ. මෙම ගුණාංග පාර්ශවකරුවන් සහ අවසාන පරිශීලකයින් අතර AI පද්ධති කෙරෙහි විශ්වාසය සහ විශ්වාසය ඇති කරයි.

විශ්වාසදායක කෘත්‍රිම බුද්ධිය හෝ TAI ආකෘති යෙදවීම හා සම්බන්ධ විභව අවදානම් අවම කළ හැකිය. මෙම AI අවදානම් වලට පුද්ගලයින්ට, සංවිධානවලට සහ පරිසර පද්ධතිවලට හානි ඇතුළත් වේ. එවැනි හානි සිදු වූ විට, ඒවා නිශ්චිත AI ආකෘති කෙරෙහි පමණක් නොව සමස්තයක් ලෙස කෘත්‍රිම බුද්ධිය කෙරෙහි ඇති විශ්වාසය අඩපණ කළ හැකිය.

විශ්වාසදායක AI, frameworks මඟින් සංවිධානවලට AI තාක්ෂණයන් සංවර්ධනය කිරීම, සම්මත කර ගැනීම සහ ඇගයීම සඳහා මඟ පෙන්විය හැකිය. එක්සත් ජනපදයේ ජාතික ප්‍රමිති සහ තාක්ෂණ ආයතනය (NIST), යුරෝපීය කොමිසමේ AI පිළිබඳ ඉහළ මට්ටමේ විශේෂඥ කණ්ඩායම සහ ආර්ථික සහයෝගීතාව සහ සංවර්ධනය සඳහා වූ සංවිධානය (OECD) ඇතුළු රාජ්‍ය සහ අන්තර් රාජ්‍ය සංවිධාන කිහිපයක් එවැනි රාමු ස්ථාපිත කර ඇත.

ඊට අමතරව, ව්‍යාපාරවලට ඔවුන්ගේ AI පද්ධතිවල විශ්වාසනීයත්වය වැඩි දියුණු කිරීම සඳහා විවිධ උපාය මාර්ග සහ මෙවලම් ක්‍රියාත්මක කළ හැකිය. උදාහරණයක් ලෙස, අඛණ්ඩ අධීක්ෂණය, ලේඛනගත කිරීම සහ AI පාලන රාමු සියල්ලම අවදානම අවම කිරීමට උපකාරී වේ.

විශ්වාසදායක AI වැදගත් වන්නේ ඇයි?

තාක්ෂණයක් ක්‍රියා කරන ආකාරය තේරුම් ගැනීම බොහෝ විට එහි කාර්යක්ෂමතාව විශ්වාස කිරීමට යතුරයි. නමුත් ගැඹුරු ඉගෙනුම් ආකෘති වැනි බොහෝ AI සහ යන්ත්‍ර ඉගෙනුම් (ML) පද්ධති සත්‍ය කළු පෙට්ටි ලෙස ක්‍රියා කරයි; ඔවුන් දත්ත අවශෝෂණය කර ප්‍රතිදාන නිර්මාණය කරයි, ඒවා එම ප්‍රතිදාන වෙත පැමිණෙන ආකාරය පිළිබඳ විනිවිදභාවයක් හෝ සුළු වශයෙන් අඩුය.

එහි ප්‍රතිඵලයක් ලෙස විශ්වාස හිඟයන් බහුලව දක්නට ලැබේ. 2023 සමීක්ෂණයකින් හෙළි වූයේ ව්‍යාපාරික නායකයින්ගෙන් 40% කට වඩා වැඩි පිරිසක් AI විශ්වසනීයත්වය පිළිබඳ කනස්සල්ල පළ කළ බවයි. මේ අතර, පාරිභෝගිකයින් AI අවිශ්වාසය ද පෙන්නුම් කර ඇත: 2024 අධ්‍යයනයකින් හෙළි වූයේ නිෂ්පාදනයක ලේබල් කිරීමේදී “කෘතිම බුද්ධිය” යන යෙදුම ඇතුළත් කිරීමෙන් සාප්පු හිමියන් එම නිෂ්පාදනය මිලදී ගැනීමට ඇති ඉඩකඩ අඩු කළ හැකි බවයි.

AI තාක්ෂණයන් සංවර්ධනය කිරීම මගින් AI තාක්ෂණයන් කෙරෙහි පාරිභෝගික සහ පාරිභෝගික විශ්වාසය අඩුවීමට හේතු විය හැකි අවදානම් කිහිපයක් ඉස්මතු කරයි, එනම්:

· පක්ෂග්‍රාහීත්වයන් - AI විසින් ගනු ලබන තීරණ වලට බලපාන පක්ෂග්‍රාහීත්වයන්

· වගවීමක් නොමැතිකම - AI තීරණ සඳහා මානව වගවීමක් හෝ වගකීමක් නොමැතිකම

· තක්කඩි හැසිරීම - අර්ධ වශයෙන් හෝ සම්පූර්ණයෙන්ම ස්වාධීන නියෝජිතයින්ගේ අනපේක්ෂිත හැසිරීම

· ඉතා සංකීර්ණ - AI ඇල්ගොරිතමවල පැහැදිලි කිරීමේ හැකියාව නොමැතිකම

ඉහළ අවදානමක් ඇති අවස්ථාවන්හිදී වැරදි හෝ හානිකර ප්‍රතිඵල නිපදවන AI පද්ධති පිළිබඳ සැබෑ ලෝක උදාහරණ AI විශ්වාසනීය ගැටළු තවදුරටත් අවුස්සයි. එක් ප්‍රසිද්ධ සෞඛ්‍ය සේවා උදාහරණයක, AI ආකෘතියක් විශ්වාසනීය ලෙස සෙප්සිස් රෝග විනිශ්චය කිරීමට අසමත් විය. පුහුණු පසුබිමක ආකෘතිය හොඳින් ක්‍රියාත්මක වුවද, රෝහල් රෝගීන්ගෙන් තුනෙන් දෙකකට වඩා සෙප්සිස් හඳුනා ගත්තේ නැත.

වෙනත් අවස්ථාවල දී, AI ආකෘති පක්ෂග්‍රාහී ඇල්ගොරිතම තීරණ ගැනීම පෙන්නුම් කර ඇති අතර, සුළුතර ප්‍රජාවන් අසමාන ලෙස ඉලක්ක කරන පුරෝකථන පොලිස් පද්ධති සහ කාන්තාවන්ට වඩා පිරිමි අපේක්ෂකයින්ට අනුග්‍රහය දක්වන අයදුම්කරු ලුහුබැඳීමේ පද්ධති ඇතුළුව. තවද, AI චැට්බොට් විසින් සංවේදී, පුද්ගලික දත්ත නොදැනුවත්වම හෙළි කිරීම සහ හිමිකාර ආයතනික තොරතුරු සොරකම් කිරීම සඳහා AI ආකෘතිවල ඇති අවදානම් උපයෝගී කර ගනිමින් හැකර්වරුන් භාවිතා කිරීම වැනි ආරක්ෂක සැලකිලිමත්වීම් තිබේ.

AI ආකෘති දුර්වල ලෙස ක්‍රියා කරන විට හෝ හානිකර ප්‍රතිඵල ලබා දෙන විට, එය එම ආකෘති කෙරෙහි පමණක් නොව, සාමාන්‍යයෙන් කෘත්‍රිම බුද්ධිය කෙරෙහි ඇති විශ්වාසය අඩපණ කළ හැකි අතර, අනාගත සංවර්ධනයට සහ AI භාවිතයට බාධා කළ හැකිය. විශ්වාසදායක AI පද්ධති සාක්ෂාත් කර ගැනීම සහ අනාගත AI සංවර්ධනයට සහාය වීම යනු රූපක AI කළු පෙට්ටිය තුළ ආලෝකය විහිදුවීමයි. මෙය කොටස්කරුවන්ට ඔවුන්ගේ AI යෙදුම් කෙරෙහි විශ්වාසදායක, නිවැරදි ප්‍රතිඵල ලබා දීමට සහ මුල් අභිප්‍රාය සමඟ පක්ෂග්‍රාහී හෝ නොගැලපෙන ප්‍රතිඵලවල අවදානම් අවම කිරීමට හැකියාව ලබා දෙයි.

විශ්වාසදායක AI හි මූලධර්ම මොනවාද?

විවිධ සංවිධාන සහ රාමු විශ්වාසදායක AI සඳහා විවිධ මාර්ගෝපදේශක මූලධර්ම සහ ඉලක්ක අවධාරණය කරයි. විශ්වාසදායක AI හි නිතර උපුටා දක්වන මූලධර්මවලට ඇතුළත් වන්නේ:

· වගවීම

· පැහැදිලි කිරීමේ හැකියාව

· සාධාරණත්වය

· අර්ථ නිරූපණය සහ විනිවිදභාවය

· පෞද්ගලිකත්වය

· විශ්වසනීයත්වය

· ශක්තිමත් බව සහ ආරක්ෂාව

· ආරක්ෂාව

වගවීම: AI හි මෙය AI ක්‍රියාකාරීන් ඔවුන්ගේ ජීවන චක්‍ර පුරාවට AI පද්ධති නිසි ලෙස ක්‍රියාත්මක කිරීම සඳහා වගකිව යුතු බවට පත් කරයි. AI තාක්ෂණය සංවර්ධනය කිරීම, යෙදවීම හෝ ක්‍රියාත්මක කිරීම සඳහා සම්බන්ධ වන පුද්ගලයින් සහ සංවිධාන මෙයට ඇතුළත් වේ.

පැහැදිලි කිරීමේ හැකියාව: AI පැහැදිලි කිරීමේ හැකියාව යනු ආකෘතියක ප්‍රතිදානයන් සත්‍යාපනය කිරීම හෝ සාධාරණීකරණය කිරීම ගැන ය. යන්ත්‍ර ඉගෙනුම් ඇල්ගොරිතම මගින් නිර්මාණය කරන ලද ප්‍රතිඵල සහ ප්‍රතිදානය තේරුම් ගැනීමට සහ විශ්වාස කිරීමට මානව පරිශීලකයින්ට හැකියාව ලබා දෙන පැහැදිලි කළ හැකි AI ලෙස සාමූහිකව හැඳින්වෙන විවිධ පැහැදිලි කිරීමේ ක්‍රම තිබේ.

සාධාරණත්වය: AI හි සාධාරණත්වය යනු පුද්ගලයන්ට සහ කණ්ඩායම්වලට සාධාරණ ලෙස සැලකීමයි. එයට ඇල්ගොරිතම සහ දත්ත පක්ෂග්‍රාහීත්වය අවම කිරීම ඇතුළත් වේ. යන්ත්‍ර ඉගෙනුම් ඇල්ගොරිතමවල පද්ධතිමය දෝෂ අසාධාරණ හෝ වෙනස් කොට සැලකීමේ ප්‍රතිඵල ඇති කරන විට ඇල්ගොරිතම පක්ෂග්‍රාහීත්වය ඇති වන අතර,

දත්ත පක්ෂග්‍රාහීත්වය AI ආකෘතියක භාවිතා කරන පුහුණු දත්තවල විකෘති හෝ නියෝජනය නොකරන ස්වභාවය අදහස් කරයි. අර්ථ නිරූපණය සහ විනිවිදභාවය: AI අර්ථ නිරූපණය කිරීමේ හැකියාව මිනිසුන්ට AI ආකෘතිවල තීරණ ගැනීමේ ක්‍රියාවලීන් වඩා හොඳින් තේරුම් ගැනීමට සහ පැහැදිලි කිරීමට උපකාරී වේ. අර්ථකථනය යනු විනිවිදභාවය ගැන වන අතර, පරිශීලකයින්ට ආකෘතියක ගෘහ නිර්මාණ ශිල්පය, එය භාවිතා කරන විශේෂාංග සහ එය ඒවා ඒකාබද්ධ කරන ආකාරය පුරෝකථන ලබා දීමට තේරුම් ගැනීමට ඉඩ සලසයි. සමහර ආකෘති සහජයෙන්ම අර්ථකථනය කළ හැකි වුවද, අනෙක් ඒවාට අර්ථකථන ක්‍රම භාවිතා කිරීම අවශ්‍ය වේ.

පෞද්ගලිකත්වය: AI පෞද්ගලිකත්වය යනු AI විසින් රැස් කරන ලද, භාවිතා කරන, බෙදා ගන්නා හෝ ගබඩා කරන පුද්ගලික හෝ සංවේදී තොරතුරු ආරක්ෂා කිරීමයි. AI පෞද්ගලිකත්වය දත්ත පෞද්ගලිකත්වයට සමීපව සම්බන්ධ වේ. තොරතුරු පෞද්ගලිකත්වය ලෙසද හැඳින්වෙන දත්ත රහස්‍යතාවය යනු පුද්ගලයෙකුට තම පෞද්ගලික දත්ත පාලනය කළ යුතු බවට වන මූලධර්මයයි. ගුප්ත විද්‍යාවේ සිට ෆෙඩරේටඩ් ඉගෙනීම දක්වා ක්‍රම ගණනාවක් හරහා AI සහ දත්ත රහස්‍යතාව පවත්වා ගැනීම වැඩිදියුණු කළ හැකිය.

විශ්වාසය: විශ්වසනීයත්වය යනු යම් යම් කොන්දේසි යටතේ දී ඇති කාල පරිච්ඡේදයක් සඳහා අසාර්ථක නොවී, අපේක්ෂිත හෝ අවශ්‍ය පරිදි ක්‍රියා කිරීමේ හැකියාව ලෙස අර්ථ දැක්විය හැකිය. අපේක්ෂිත කොන්දේසි යටතේ භාවිතා කරන විට, විශ්වාසදායක AI පද්ධති, එම පද්ධතිවල සම්පූර්ණ ආයු කාලය ඇතුළත් විය හැකි, දී ඇති කාල පරිච්ඡේදයක් තුළ නිවැරදි ප්‍රතිඵල ලබා දිය යුතුය.

ශක්තිමත් බව සහ ආරක්ෂාව: ආරක්ෂිත, ශක්තිමත් AI පද්ධතිවලට විරුද්ධවාදී ප්‍රහාර සහ අනවසර ප්‍රවේශයන්ට එරෙහිව ආරක්ෂණ යාන්ත්‍රණ ඇත, සයිබර් ආරක්ෂණ අවදානම් සහ අවදානම් අවම කිරීම. අනපේක්ෂිත හානියක් සිදු නොකර අසාමාන්‍ය තත්වයන් යටතේ ක්‍රියා කිරීමට සහ අනපේක්ෂිත සිදුවීමකින් පසු සාමාන්‍ය ක්‍රියාකාරිත්වයට නැවත පැමිණීමට ඔවුන්ට හැකිය.

ආරක්ෂාව: ආරක්ෂිත AI පද්ධති මිනිස් ජීවිතයට, සෞඛ්‍යයට, දේපළට හෝ පරිසරයට අනතුරක් නොකරයි. ඒවා කල්තියා නිර්මාණය කර ඇත්තේ හානිවලින් මිනිසුන් ආරක්ෂා කිරීම සඳහා වන අතර අනාරක්ෂිත ප්‍රතිඵල අවම කරන පියවර ඇතුළත් වේ, පද්ධතියක් භාවිතයෙන් ඉවත් කිරීමේ හැකියාව ඇතුළුව.

සමස්තයක් වශයෙන්, විශ්වාසදායක AI ඉතා පොහොසත් අර්ථයක් ඇති අතර එය දෘෂ්ටිකෝණ කිහිපයකින් අර්ථ දැක්විය හැකිය. සාධාරණ AI, පැහැදිලි කළ හැකි AI යනාදිය ඇතුළුව පවතින බොහෝ පාරිභාෂික පදවල සංකල්ප එහි අඩංගු වේ. විශ්වාසදායක AI පිළිබඳ යෝජිත සංකල්පය සහ සදාචාරාත්මක AI, ප්‍රයෝජනවත් AI සහ වගකිවයුතු AI යන සංකල්ප අතර විශාල අතිච්ඡාදනයන් ද පවතී. ඒ සියල්ල මිනිස් සමාජයට නිරසාර ලෙස ප්‍රතිලාභ ලබා දෙන විශ්වාසදායක AI ගොඩනැගීම අරමුණු කරයි. කෙසේ වෙතත්, මෙම සංකල්ප අතර යම් යම් වෙනස්කම් පවතී, මන්ද ඒවා විවිධ දෘෂ්ටිකෝණවලින් යෝජනා කර ඇත. උදාහරණයක් ලෙස, ප්‍රයෝජනවත් AI සහ වගකිවයුතු AI හි, AI භාවිතා කරන්නන් සහ පාලකයින් සඳහා සමහර මූලධර්ම සහ අවශ්‍යතා ද ඇති අතර යෝජිත විශ්වාසදායක AI හි, අවධානය AI තාක්ෂණයේ මූලධර්ම කෙරෙහිම යොමු වේ. මෙම සංකල්ප අතර සම්බන්ධතා නිරූපණය කරයි.

විශ්වාසදායක AI මගින් අවම කළ හැකි අවදානම් මොනවාද?

විශ්වාසදායක ගුණාංග නොමැති AI පද්ධති පුළුල් පරාසයක අවදානම් ඇති කරයි. එක්සත් ජනපද වාණිජ දෙපාර්තමේන්තුවේ කොටසක් වන ජාතික ප්‍රමිති සහ තාක්ෂණ ආයතනය (NIST), AI අවදානම් කළමනාකරණය සඳහා මිණුම් ලකුණක් බවට පත්ව ඇති රාමුවක් සකස් කළේය. එය

AI පද්ධති වලින් සිදුවිය හැකි හානි පිළිබඳ අවදානම් පහත කාණ්ඩවලට සංවිධානය කරයි:

· මිනිසුන්ට හානියක්

· සංවිධානයකට හානියක්

· පරිසර පද්ධතියකට හානියක්

ජනතාවට සිදුවන හානිය: මෙම කාණ්ඩයට පුද්ගලයන්ගේ සිවිල් නිදහස, අයිතිවාසිකම්, ශාරීරික හෝ මානසික ආරක්ෂාව හෝ ආර්ථික අවස්ථාවන්ට සිදුවන හානිය ඇතුළත් වේ. ප්‍රජාතන්ත්‍රවාදී සහභාගීත්වයට හෝ අධ්‍යාපනික ප්‍රවේශයට සිදුවන හානිය ලෙස සමාජවලට සිදුවන වෙනස්කම් සහ බලපෑම් හරහා කණ්ඩායම් කෙරෙහි වන බලපෑම් ද එයට ඇතුළත් වේ.

සංවිධානයකට සිදුවන හානිය: මෙම කාණ්ඩයට යොමු වන්නේ සංවිධානයක ව්‍යාපාරික මෙහෙයුම් වලට සිදුවන හානිය, ආරක්ෂක කඩකිරීම් හෝ මුදල් අලාභයෙන් සිදුවන හානිය සහ එහි කීර්ති නාමයට සිදුවන හානියයි.

පරිසර පද්ධතියකට සිදුවන හානිය: මෙම කාණ්ඩයට සිදුවන හානිය “අන්තර් සම්බන්ධිත සහ අන්තර් රඳා පවතින මූලද්‍රව්‍ය සහ සම්පත්” වලට සිදුවන හානියයි. NIST විශේෂයෙන් ගෝලීය මූල්‍ය පද්ධතියට, සැපයුම් දාමයට හෝ “අන්තර් සම්බන්ධිත පද්ධති” මෙන්ම ස්වාභාවික සම්පත්, පරිසරය සහ ග්‍රහලෝකයට සිදුවන හානිය උපුටා දක්වයි. AI පද්ධති වලින් පක්ෂග්‍රාහී හෝ සාවද්‍ය ප්‍රතිදානයන් බහුවිධ හානිවලට හේතු විය හැක. පෙර උදාහරණයකට ආපසු යන විට, පක්ෂග්‍රාහී අයදුම්කරු-ලුහුබැඳීමේ පද්ධති මගින් පුද්ගලයන්ගේ ආර්ථික අවස්ථාවන්ට හානි කළ හැකි අතරම සංවිධානයක කීර්ති නාමයට ද හානි කළ හැකිය. විශාල භාෂා ආකෘතියක් (LLM)

සමාගමක මෙහෙයුම් අඩාල කරන අනිෂ්ට මෘදුකාංග ක්‍රියාත්මක කිරීමට රැවටෙන්නේ නම්, එය සමාගමට සහ එය අයත් සැපයුම් දාමයට හානි කළ හැකිය. විශ්වාසදායක AI පද්ධති එවැනි භයානක අවස්ථා සහ ප්‍රතිවිපාක වළක්වා ගැනීමට උපකාරී විය හැකිය.

NIST ට අනුව, “විශ්වාසදායක AI පද්ධති සහ ඒවායේ වගකිවයුතු භාවිතය ඍණාත්මක අවදානම් අවම කර මිනිසුන්, සංවිධාන සහ පරිසර පද්ධති සඳහා ප්‍රතිලාභ සඳහා දායක විය හැකිය.”

විශ්වාසදායක AI vs. සදාචාරාත්මක AI vs. වගකිවයුතු AI

විශ්වාසදායක AI, සදාචාරාත්මක AI සහ වගකිවයුතු AI යන පද බොහෝ විට එකිනෙකට වෙනස් ලෙස භාවිතා වේ. තවද එක් එක් සංකල්පයේ අර්ථ දැක්වීම් මූලාශ්‍රය අනුව වෙනස් විය හැකි අතර බොහෝ විට සැලකිය යුතු අතිච්ඡාදන ඇතුළත් වන බැවින්, 3 අතර නීරණාත්මක වෙනස්කම් ඇඳීම අභියෝගාත්මක විය හැකිය.

නිදසුනක් ලෙස, විශ්වාසදායක AI සහ සදාචාරාත්මක AI හි පොදු අර්ථ දැක්වීම් සාධාරණත්වය සහ පෞද්ගලිකත්වය වැනි මූලධර්ම එක් එක් සංකල්පයට පදනම ලෙස ලැයිස්තුගත කරයි. ඒ හා සමානව, වගවීම සහ විනිවිදභාවය යනු බොහෝ විට විශ්වාසදායක AI සහ වගකිවයුතු AI යන දෙකටම සම්බන්ධ වන ගුණාංග වේ.

AI මත පදනම් වූ සංකල්ප 3 අතර වෙනස හඳුනා ගැනීමට එක් ක්‍රමයක් නම්, ඒවායේ මූලික මූලධර්මවලින් ඔබ්බට ගොස් ඒවා භාවිතා කරන ආකාරය කෙරෙහි අවධානය යොමු කිරීමයි:

✓ විශ්වාසදායක AI බොහෝ විට සාක්ෂාත් කර ගන්නා දෙයක් ලෙස රාමු කර ඇත; එහි පරිශීලකයින් සමඟ විශ්වාසය ස්ථාපිත කරන්නේ AI ය.

✓ ඊට වෙනස්ව, සදාචාරාත්මක AI යනු, ඒවායේ සැලසුම් සහ සංවර්ධනය අතරතුර අන්තර්ගත මානව වටිනාකම් සහ සදාචාරාත්මක ප්‍රමිතීන් පිළිබිඹු කරන, සදාචාරාත්මක සලකා බැලීම් ඇති AI පද්ධති ලෙස විස්තර කර ඇත.

✓ වගකිවයුතු AI යනු AI යෙදුම් සහ වැඩ ප්‍රවාහයන් තුළ එම ආචාර ධර්ම ඇතුළත් කිරීමේ ප්‍රායෝගික මාධ්‍යයන් ඇතුළත් කිරීමක් ලෙස අර්ථ දැක්විය හැකිය.

විශ්වාසදායක AI සාක්ෂාත් කර ගැනීම සඳහා උපාය මාර්ග

AI ඇල්ගොරිතම සහ දත්ත කට්ටල ඇතුළුව ඔවුන්ගේ කෘතිම බුද්ධි පද්ධති විශ්වාසදායක AI හි මූලධර්මවලට අනුකූලව ක්‍රියාත්මක වන බව සහතික කිරීමට සංවිධානවලට වැදගත් පියවර ගත හැකිය.

· තක්සේරුව: AI-සක්‍රීය ව්‍යාපාර ක්‍රියාවලීන් තක්සේරු කිරීම සමාගම්වලට විවිධ විශ්වාසනීය මිනුම්වල වැඩිදියුණු කිරීමට ඉඩ ඇති ස්ථාන තීරණය කිරීමට උපකාරී වේ.

· අඛණ්ඩ අධීක්ෂණය: AI පක්ෂග්‍රාහීත්වය සහ ආකෘති ප්ලාවිතය වැනි ගැටළු සඳහා අඛණ්ඩ අධීක්ෂණය හරහා, සංවිධාන අසාධාරණ හෝ සාවද්‍ය ක්‍රියාවලීන් හෝ ප්‍රතිදානයන් කල්තියා ආමන්ත්‍රණය කරන අතර එමඟින් සාධාරණත්වය සහ විශ්වසනීයත්වයට සහාය වේ.

· අවදානම් කළමනාකරණය: අවදානම් කළමනාකරණ රාමුවක් සහ මෙවලම් ක්‍රියාත්මක කිරීම

AI ශක්තිමත් බව බල ගැන්වීම සඳහා ආරක්ෂක උල්ලංඝනයන් සහ රහස්‍යතා උල්ලංඝනයන් හඳුනා ගැනීම සහ අවම කිරීම සඳහා ඉඩ සලසයි.

· ලේඛනගත කිරීම: දත්ත විද්‍යාව සහ AI ජීවන චක්‍රය හරහා ස්වයංක්‍රීය ලියකියවිලි කර්මාන්ත සහ නියාමන විගණන සඳහා භාවිතා කළ හැකි අතර, වගවීම සහ විනිවිදභාවය සක්‍රීය කරයි.

· AI පාලන රාමු: AI පාලන රාමුවලට දත්ත සහ ආකෘති කළමනාකරණය පිළිබඳ ක්‍රියා පටිපාටි ඇතුළත් වන අතර,සංවිධානයක් තුළ සංවර්ධකයින් සහ දත්ත විද්‍යාඥයින් අභ්‍යන්තර ප්‍රමිතීන් සහ රජයේ රෙගුලාසි දෙකම අනුගමනය කරන බව සහතික කිරීමට උපකාරී වේ.

AI පාලන මෘදුකාංග සහ විවෘත මූලාශ්‍ර මෙවලම් කට්ටල මඟින් සංවිධානවලට ඔවුන්ගේ AI පද්ධතිවල විශ්වසනීයත්වය වැඩි දියුණු කිරීම සඳහා මෙම සහ වෙනත් පියවර ගැනීමට උපකාරී වේ. නිවැරදි පියවර සහ ආරක්ෂණ ක්‍රම ක්‍රියාත්මක වීමත් සමඟ, ව්‍යාපාරවලට AI හි බලය උපයෝගී කර ගන්නා විට අවදානම් අවම කර ගත හැකිය.

ආර්ථික වර්ධනයට විශ්වාසදායක AI හි බලපෑම

AI හි කළමනාකරණය නොකළ සංවර්ධනය මගින් ඇති වන අවදානම් ආමන්ත්‍රණය නොකළහොත්, පාරිභෝගික සහ සේවක විශ්වාසය අඩුවීම AI තාක්ෂණයන් දිගුකාලීන වර්ධනයට සහ භාවිතයට බාධාවක් විය හැකි අතර AI-සක්‍රීය විසඳුම් සඳහා ආයෝජනය කිරීමෙන් පුද්ගලික අංශය අධෛර්යමත් කළ හැකි අතර AI හි ප්‍රතිලාභ සහ සමස්ත ආර්ථික වර්ධනයට සීමා කළ හැකිය.

AI වෙළඳපොළ අවහිර කිරීමේ ප්‍රතිවිපාක සැලකිය යුතු අතර, AI හි නායකත්වය පවත්වා ගැනීමට අපොහොසත් වුවහොත් එක්සත් ජනපදයට ගෝලීය වශයෙන් තරඟ කිරීමට ඇති හැකියාවට දිගුකාලීන අවදානම් ඇත. උදාහරණයක් ලෙස ජාත්‍යන්තර දත්ත සංස්ථාව (IDC) ඇස්තමේන්තු කර ඇත්තේ, AI වෙළඳපොළ සඳහා ආදායම 2021 දී ඩොලර් බිලියන 327.5 දක්වා ළඟා වනු ඇතැයි පුරෝකථනය කර ඇති අතර, එය AI පද්ධති ක්‍රියාත්මක කිරීමෙන් සහ භාවිතා කිරීමෙන් ලැබෙන ද්විතියික ප්‍රතිලාභ පවා නොසලකයි. AI සඳහා ආරෝපණය කළ හැකි ආර්ථික

බලපෑම වසර පහක් තුළ ඩොලර් බිලියන 447 ත් 1.43 ත් අතර ඇස්තමේන්තු කර ඇති අතර, එය AI R&D සඳහා රජයේ ආයෝජන වැඩි කිරීමෙන් වේගවත් කළ හැකිය. විශ්වාසදායක AI ප්‍රවේශයක් මඟින් AI පද්ධති කෙරෙහි විශ්වාසය අඩු කළ හැකි අවදානම් අවම කර ගත හැකි අතර මෙම නීරණාත්මක අංශයේ නවෝත්පාදනයන් මැඩපවත්වන අතර එමඟින් ආර්ථික වර්ධනයට සහ ජනතාවගේ සෞඛ්‍යය, ආරක්ෂාව සහ යහපැවැත්ම වැඩිදියුණු කළ හැකිය.

වැඩි වූ AI පර්යේෂණ සහ සංවර්ධන බලපෑම ආකෘතිකරණය කිරීම

පෙර තාක්ෂණ විප්ලවයන්ට සාපේක්ෂව, AI හට පෙර බොහෝ තාක්ෂණයන්ට වඩා ඉහළ සමස්ත ආර්ථික වටිනාකමක් ඇති බවට සාක්ෂි නිබේ. වසර 20ක කාලයක් තුළ උපුටා දක්වමින්, AI මත ප්‍රතිලාභය පෙර ඵලදායිතාව වැඩි කරන පරිවර්තනවල දක්නට ලැබෙන රටාවට සමාන යැයි උපකල්පනය කළහොත්, ඩොලර් බිලියන 25 ක නිර්දේශිත ෆෙඩරල් AI පර්යේෂණ සහ සංවර්ධන වියදම්වල ප්‍රතිඵලයක් ලෙස 2045 වන විට ඩොලර් බිලියන 94 ත් 156 ත් අතර වර්ධක ආර්ථික බලපෑමක් ඇති වනු ඇත.

එපමණක් නොව, 2019 යුරෝපීය විශ්ව විද්‍යාල ආයතනයේ පත්‍රිකාවක් යෝජනා කරන්නේ ස්වයං වැඩිදියුණු කිරීමේ හැකියාව නිසා, AI හුදෙක් නව ආකාරයේ ස්වයංක්‍රීයකරණයක් ලෙස නොව, "සම්පූර්ණයෙන්ම නව නිෂ්පාදන ආදානයක්" ලෙස සංකල්පනය කළ යුතු බවයි, එය ශ්‍රම ඵලදායිතාව හෝ ප්‍රාග්ධන ආයෝජන මත ප්‍රතිලාභ පමණක් නොව සම්පූර්ණයෙන්ම නව කාර්යයන් කළ හැකි කරයි. මෙයින් ගම්‍ය වන්නේ ඵලදායිතාවයට AI හි දායකත්වය පෙර පරිවර්තනීය තාක්ෂණයන්ට වඩා සැලකිය යුතු ලෙස ඉක්මවා යා හැකි බවයි. AI රොබෝ විද්‍යාව හෝ ජංගම දුරකථන වලට සමාන දායකත්වයක් පමණක් ලබා දීමට ගත වුවද, මෙය 2025 වන විට දළ දේශීය නිෂ්පාදිතයේ වර්ධනයෙන් ඩොලර් බිලියන 477 කට වඩා නියෝජනය කළ හැකිය. කෙසේ වෙතත්, AI සැබවින්ම පෙර පරම්පරාවල තාක්ෂණයන් මෙන් නොව පෙර සියලුම තොරතුරු තාක්ෂණ ආයෝජනවලට වඩා ඵලදායිතා වැඩිදියුණු කිරීම් ලබා දෙන්නේ

නම්, ඵලදායිතා වර්ධනයට 1.2-ලක්ෂ්‍ය දායකත්වයක් ලබා දෙන්නේ නම්, බලපෑම 2025 වන විට අතිරේක දළ දේශීය නිෂ්පාදිතයෙන් ඩොලර් ට්‍රිලියන 1.4 ක් තරම් විශාල විය හැකිය.

අවසාන වශයෙන්, පෙර පරිවර්තනීය තාක්ෂණයන්ට සාපේක්ෂව AI පර්යේෂණ සහ සංවර්ධන ආයෝජනවලින් ඉහළ ප්‍රතිඵලයක් ලබා ගැනීමේ සම්භාවිතාව, පුද්ගලික අංශය AI පර්යේෂණ සහ සංවර්ධන කටයුතු සඳහා සිදු කරන සැලකිය යුතු ආයෝජන මගින් ද වේගවත් කළ හැකිය. 2025 වන විට එක්සත් ජනපදයේ පෞද්ගලික සමාගම් AI පර්යේෂණ සහ සංවර්ධන කටයුතු සඳහා වාර්ෂිකව ඩොලර් බිලියන 100 කට ආසන්න මුදලක් වැය කරනු ඇතැයි අපේක්ෂා කෙරේ. AI වෙතින් ඵලදායිතාවයේ සමස්ත වැඩිවීමෙන් කොටසක් පමණක් මහජන පර්යේෂණ සහ සංවර්ධන ආයෝජනවලට සෘජුවම ආරෝපණය කළ හැකි වුවද, AI පර්යේෂණ සහ සංවර්ධන කටයුතු සඳහා ෆෙඩරල් ආයෝජන AI ආර්ථිකයක අඛණ්ඩ වර්ධනයට පහසුකම් සැලසීම සඳහා සැලකිය යුතු ප්‍රමුඛතාවයක් බවට පත් කරයි.

විශ්වාසදායක AI සක්‍රීය කිරීම සඳහා රාජ්‍ය ප්‍රතිපත්ති

විශ්වාසදායක AI හි ගෝලීය ප්‍රමුඛයෙකු වන අතරම AI හි අවදානම් පිළිබඳ උත්සුකයන් අවම කරන අතරම AI නවෝත්පාදනවල ප්‍රතිලාභවලට සහාය වීමට එක්සත් ජනපදයට සැලකිය යුතු අවස්ථාවක් තිබේ. විශ්වාසදායක AI සංවර්ධනයට පහසුකම් සැලසීම කෙරෙහි අවධානය යොමු කරන ලද ක්‍රියාකාරී ප්‍රතිපත්ති උපාය මාර්ගයකට:

· මානව කේන්ද්‍රීය AI නවෝත්පාදන සංවර්ධනය සඳහා ප්‍රමුඛ කාර්යභාරයක් ඉටු කිරීම සඳහා පෞද්ගලික අංශයේ විශේෂඥතාව උපයෝගී කර ගැනීම

· එම නවෝත්පාදනයන්ගෙන් පැන නගින නව ව්‍යාපාරික අවස්ථා මතුවීම ප්‍රවර්ධනය කිරීම

· AI ආර්ථිකයේ ප්‍රතිලාභ සඳහා ඇමරිකානු සේවකයින්ගේ අයිතිය වැඩි කිරීම

· AI නවෝත්පාදන අවදානම් වලට වඩා වැඩි ආර්ථික හා සමාජීය ප්‍රතිලාභ ලබා දෙන බවට පාරිභෝගිකයින්ගේ සහ මහජනතාවගේ විශ්වාසය ගොඩනැගීම

විශ්වාසදායක AI සංවර්ධනයට හැකියාව ලබා දිය හැකි විවිධ ප්‍රතිපත්ති විසඳුම් කාණ්ඩ තිබේ, ඒවාට ඇතුළත් වන්නේ:

· AI හි මූලික පර්යේෂණ

· AI ඇල්ගොරිතමවල කාර්ය සාධනය සහ විශ්වසනීයත්වය සඳහා ප්‍රමිතීන්

· AI ආකෘති වැඩිදියුණු කළ පුහුණුව සඳහා රජයේ දත්ත කට්ටල වෙත ප්‍රවේශය

· දාර දෘඩාංග සහ උපාංග සඳහා අන්තර් ක්‍රියාකාරීත්ව ප්‍රමිතීන්

· නව AI ආකෘති සංවර්ධනය කිරීම සහ පුහුණු කිරීම සඳහා බෙදාගත් පරිගණක සහ වලාකුළු සම්පත් වෙත ප්‍රවේශය

· සංවර්ධනය සරල කිරීමට හෝ වේගවත් කිරීමට විවෘත මූලාශ්‍ර මෙවලම් සහ රාමු

· AI කුසලතා සහ වෘත්තීය තේරීම ප්‍රවර්ධනය කිරීම සඳහා තරුණයින් සඳහා විෂයමාලා

• AI කුසලතා අවශ්‍ය වන රැකියා භූමිකාවන්ට මාරුවීමට සහාය වීම සඳහා වැඩිහිටියන් ඉලක්ක කරගත් නැවත පුහුණු කිරීම හෝ අඛණ්ඩ අධ්‍යාපන වැඩසටහන්

· AI භාවිතය සහ යෙදවීම සඳහා පොදු රාමු ප්‍රවර්ධනය කිරීම සඳහා පවතින ජාත්‍යන්තර හවුල්කාරිත්වයන් ස්ථාපිත කිරීම හෝ සහාය දීම.

සමෘද්ධිමත් සහ තිරසාර AI-සක්‍රීය ආර්ථිකයක් ගොඩනැගීම සඳහා, AI පද්ධති සංවර්ධනය හා යෙදවීමේදී විශ්වාසදායක AI සංකල්ප ඇතුළත් කිරීමට නවෝත්පාදකයින් දිරිමත් කිරීම සඳහා බුද්ධිමත් ප්‍රතිපත්තිමය විසඳුම් අවශ්‍ය වනු ඇත.

Microsoft Trustworthy AI

Microsoft හි, ඔවුන් විශ්වාසදායක AI සහතික කිරීමට කැපවී සිටින අතර කර්මාන්තයේ ප්‍රමුඛ සහායක තාක්ෂණය ගොඩනඟමින් සිටිති. ඔවුන්ගේ කැපවීම් සහ හැකියාවන් අත්වැල් බැඳගෙන

තම ගනුදෙනුකරුවන් සහ සංවර්ධකයින් සෑම ස්ථරයකම ආරක්ෂා කර ඇති බව සහතික කරයි.

Trustworthy AI හි ප්‍රධාන ලක්ෂණයක් වන්නේ Microsoft Azure AI අන්තර්ගත ආරක්ෂාවේභූමිභාග හඳුනාගැනීමේ විශේෂාංගය වන අතර එය දැන් සැබෑ කාලය තුළ මායාවන් ගැටළු නිවැරදි කිරීමට නිවැරදි කිරීමේ විකල්පයක් ඇති අතර වඩාත් නිවැරදි සහ විශ්වාසදායක AI ප්‍රතිචාර සහතික කරයි. රහසිගත අනුමාන කිරීමේ විශේෂාංගය අනුමාන කිරීමේ ක්‍රියාවලිය අතරතුර සංවේදී පාරිභෝගික දත්ත ආරක්ෂිතව සහ පෞද්ගලිකව පවතින බව සහතික කරන අතර, පුහුණු AI ආකෘතිය නව දත්ත මත පදනම්ව අනාවැකි හෝ තීරණ ගනී.

ඔවුන්ගේ කැපවීම් මත ගොඩනඟමින්, Microsoft AI පද්ධතිවල ආරක්ෂාව, ආරක්ෂාව සහ පෞද්ගලිකත්වය ශක්තිමත් කිරීම සඳහා නව නිෂ්පාදන හැකියාවන් නිවේදනය කරයි.

ආරක්ෂාව: ආරක්ෂාව මයික්‍රොසොෆ්ට් හි ඔවුන්ගේ ප්‍රමුඛතාවය වන අතර, ඔවුන්ගේ පුළුල් කරන ලද ආරක්ෂිත අනාගත මුලපිරීම

(SFI) සමාගම පුරා කැපවීම් සහ ගනුදෙනුකරුවන් වඩාත් ආරක්ෂිත කිරීමට ඔවුන්ට හැඟෙන වගකීම අවධාරණය කරයි. මෙය අන් සියල්ලටම වඩා ආරක්ෂාවට ප්‍රමුඛත්වය දීමට ඔවුන්ගේ ප්‍රතිඥාව ඉටු කරන අතර මූලධර්ම තුනකින් මඟ පෙන්වනු ලැබේ: සැලසුම අනුව සුරක්ෂිත කිරීම, පෙරනිමියෙන් සුරක්ෂිත කිරීම සහ ආරක්ෂිත මෙහෙයුම්.

ඔවුන්ගේ පළමු පාර්ශවීය පිරිනැමීම් වන මයික්‍රොසොෆ්ට් ඩිෆෙන්ඩර් සහ පර්වීව් වලට අමතරව, කඩිනම් එන්නත් සහ ප්‍රකාශන හිමිකම් උල්ලංඝනය කිරීම් වැළැක්වීමට උපකාරී වන බිල්ට්-ඉන් කාර්යයන් වැනි මූලික ආරක්ෂක පාලනයන් සමඟ AI සේවාවන් පැමිණේ. ඒවා මත ගොඩනඟමින්, ඔවුන් නව හැකියාවන් දෙකක් නිවේදනය කරයි:

· ක්‍රියාකාරී අවදානම් තක්සේරු කිරීම් සඳහා සහාය වීම සඳහා Azure AI ස්ටුඩියෝවේ ඇඟයීම්.

· වෙබ් සෙවුම මඟින් Copilot ප්‍රතිචාරය වැඩි දියුණු කරන ආකාරය පරිපාලකයින්ට සහ පරිශීලකයින්ට වඩා හොඳින් තේරුම් ගැනීමට උපකාර කිරීම සඳහා Microsoft 365 Copilot වෙබ් විමසුම් වලට විනිවිදභාවය ලබා දෙනු ඇත. ඉක්මනින් පැමිණේ.

මයික්‍රොසොෆ්ට් හි ආරක්ෂක හැකියාවන් දැනටමත් පාරිභෝගිකයින් විසින් භාවිතා කරනු ලැබේ. එන්ජින් නිෂ්පාදනය සහ පිරිසිදු බලශක්ති තාක්ෂණයන් සංවර්ධනය කිරීම සඳහා ප්‍රසිද්ධ වසර 105 ක් පැරණි සමාගමක් වන කමින්ස්, දත්ත වර්ගීකරණය, ටැග් කිරීම සහ ලේබල් කිරීම ස්වයංක්‍රීය කිරීම මගින් ඔවුන්ගේ දත්ත ආරක්ෂාව සහ පාලනය ශක්තිමත් කිරීම සඳහා මයික්‍රොසොෆ්ට් පර්වීව් වෙත යොමු විය. මෘදුකාංග ඉංජිනේරු සහ ව්‍යාපාර උපදේශන සමාගමක් වන ඊපිඒඑම් සිස්ටම්ස්, මයික්‍රොසොෆ්ට් වෙතින් ලැබෙන දත්ත ආරක්ෂාව නිසා පරිශීලකයින් 300 දෙනෙකු සඳහා මයික්‍රොසොෆ්ට් 365 කොපිලට් යෙදවීය. තොරතුරු තාක්ෂණ ජ්‍යෙෂ්ඨ අධ්‍යක්ෂ ජේ.ටී. සොඩානෝ, "අනෙකුත් විශාල භාෂා ආකෘති (LLM) හා සසඳන විට මයික්‍රොසොෆ්ට් 365 සඳහා කොපිලට් සමඟ අපි බොහෝ විශ්වාසයෙන් සිටියෙමු, මන්ද අපි මයික්‍රොසොෆ්ට් පර්වීව් හි විනිවිදභාවය කර ඇති එකම තොරතුරු සහ දත්ත ආරක්ෂණ ප්‍රතිපත්ති කොපිලට් සඳහා අදාළ වන බව අපි දනිමු."

ආරක්ෂාව: ආරක්ෂාව සහ පෞද්ගලිකත්වය යන දෙකම ඇතුළුව, 2018 දී පිහිටුවන ලද මයික්‍රොසොෆ්ට් හි පුළුල් වගකිවයුතු AI මූලධර්ම, ඔවුන් සමාගම පුරා ආරක්ෂිතව AI ගොඩනඟා යෙදවිය යුතු ආකාරය මඟ පෙන්වයි.

ප්‍රායෝගිකව මෙයින් අදහස් කරන්නේ හානිකර අන්තර්ගතය, පක්ෂග්‍රාහීත්වය, අනිසි භාවිතය සහ වෙනත් අනපේක්ෂිත අවදානම් වැනි අනවශ්‍ය හැසිරීම් වළක්වා ගැනීම සඳහා පද්ධති නිසි ලෙස ගොඩනැගීම, පරීක්ෂා කිරීම සහ අධීක්ෂණය කිරීමයි. වසර ගණනාවක් පුරා, මෙම මූලධර්ම ආරක්ෂා කිරීමට සහ AI ආරක්ෂිතව ගොඩනැගීමට සහ යෙදවීමට අවශ්‍ය පාලන ව්‍යුහය, ප්‍රතිපත්ති, මෙවලම් සහ ක්‍රියාවලීන් ගොඩනැගීම සඳහා ඔවුන් සැලකිය යුතු ආයෝජන සිදු කර ඇත.

මයික්‍රොසොෆ්ට් හිදී, ඔවුන්ගේ වගකිවයුතු AI මූලධර්ම ආරක්ෂා කිරීමේ මෙම ගමනේදී ඔවුන්ගේ ඉගෙනුම් ඔවුන්ගේ ගනුදෙනුකරුවන් සමඟ බෙදා ගැනීමට ඔවුන් කැපවී සිටිති. මයික්‍රොසොෆ්ට් ඔවුන්ගේම හොඳම භාවිතයන් සහ ඉගෙනුම් භාවිතා කරමින් පුද්ගලයින්ට සහ සංවිධානවලට ඔවුන් උත්සාහ කරන ඉහළ ප්‍රමිතීන් බෙදා ගන්නා AI යෙදුම් ගොඩනැගීමට හැකියාවන් සහ මෙවලම් ලබා දෙයි.

අද, මයික්‍රොසොෆ්ට් අවදානම් අවම කරන අතරම AI හි ප්‍රතිලාභ හඹා යාමට පාරිභෝගිකයින්ට උපකාර කිරීම සඳහා නව හැකියාවන් බෙදා ගනී:

· පරිශීලකයින් දැකීමට පෙර සැබෑ කාලය තුළ මායාවන් ගැටළු නිරාකරණය කිරීමට උපකාරී වන Microsoft Azure AI අන්තර්ගත ආරක්ෂාවේ භූගතභාවය හඳුනාගැනීමේ විශේෂාංගයේ නිවැරදි කිරීමේ හැකියාව.

· පාරිභෝගිකයින්ට උපාංගවල Azure AI අන්තර්ගත ආරක්ෂාව ඇතුළත් කිරීමට ඉඩ සලසන Embedded Content Safety. වලාකුළු සම්බන්ධතාවය අනරමැදි හෝ ලබා ගත නොහැකි විය හැකි උපාංග අවස්ථා සඳහා මෙය වැදගත් වේ.

· පාරිභෝගිකයින්ට ප්‍රතිදානවල ගුණාත්මකභාවය සහ අදාළත්වය අගය කිරීම සහ ඔවුන්ගේ AI යෙදුම ආරක්ෂිත ද්‍රව්‍ය කොපමණ වාරයක් ප්‍රතිදානය කරයිද යන්න තක්සේරු කිරීමට උපකාර කිරීම සඳහා Azure AI Studio හි නව ඇගයීම්.

· පෙර පැවති අන්තර්ගතය සහ කේතය හඳුනා ගැනීමට උපකාර කිරීම සඳහා කේතය සඳහා ආරක්ෂිත ද්‍රව්‍ය හඳුනාගැනීම දැන් Azure AI අන්තර්ගත ආරක්ෂාව තුළ පෙරදසුනෙහි ඇත. මෙම විශේෂාංගය සංවර්ධකයින්ට GitHub ගබඩාවල පොදු මූලාශ්‍ර කේතය ගවේෂණය කිරීමට, සහයෝගීතාවය සහ විනිවිදභාවය පෝෂණය කිරීමට සහ වඩාත් දැනුවත් කේතීකරණ තීරණ සක්‍රීය කිරීමට උපකාරී වේ.

කර්මාන්ත පුරා සිටින ගනුදෙනුකරුවන් දැනටමත් වඩාත් ආරක්ෂිත සහ විශ්වාසදායක AI යෙදුම් ගොඩනැගීම සඳහා Microsoft විසඳුම් භාවිතා කරන ආකාරය දැකීම පුදුම සහගතය. උදාහරණයක් ලෙස, 3Dක්‍රීඩා සඳහා වේදිකාවක් වන Unity, ක්‍රීඩා සංවර්ධනය පහසු කරන AI සහායකයෙකු වන Muse Chat ගොඩනැගීමට Microsoft Azure OpenAI සේවාව භාවිතා කළේය. Muse Chat Azure AI අන්තර්ගත ආරක්ෂාව තුළ අන්තර්ගත-පෙරහන් ආකෘති භාවිතා කරයි. Muse Chat, මෘදුකාංගයේ වගකීමෙන් යුත් භාවිතය සහතික කිරීම සඳහා Azure AI අන්තර්ගත ආරක්ෂාව තුළ අන්තර්ගත-පෙරහන් ආකෘති භාවිතා කරයි. මීට අමතරව, සන්නාම හවුල්කරුවන් 900 කට ආසන්න සංඛ්‍යාවක් සිටින UK-පාදක විලාසිතා සිල්ලර වෙළෙන්දෙකු වන ASOS, පාරිභෝගිකයින්ට නව පෙනුමක් සොයා ගැනීමට උපකාර වන AI යෙදුමක් හරහා උසස් තත්ත්වයේ අන්තර්ක්‍රියා සඳහා සහාය වීම සඳහා Azure AI අන්තර්ගත ආරක්ෂාව තුළ ඇති එකම බිල්ට්-ඉන් අන්තර්ගත පෙරහන් භාවිතා කළේය.

අධ්‍යාපන අවකාශයේ ද ඔවුන් බලපෑම දකිමින් සිටී. නිව් යෝර්ක් නගරයේ පොදු පාසල් මයික්‍රොසොෆ්ට් සමඟ හවුල් වී අධ්‍යාපන සන්දර්භය සඳහා ආරක්ෂිත සහ සුදුසු කතාබස් පද්ධතියක් සංවර්ධනය කළ අතර, ඔවුන් දැන් පාසල්වල එය නියමු කර ඇත. දකුණු ඕස්ට්‍රේලියානු අධ්‍යාපන දෙපාර්තමේන්තුව ද ඒ හා සමානව EdChat සමඟ පන්ති කාමරයට උත්පාදක AI ගෙන ආ අතර, සිසුන්ට සහ ගුරුවරුන්ට ආරක්ෂිත භාවිතය සහතික කිරීම සඳහා එකම යටිතල පහසුකම් මත විශ්වාසය තැබීය.

රහස්‍යතාවය: දත්ත AI හි පදනම වන අතර, Microsoft හි ප්‍රමුඛතාවය වන්නේ පරිශීලකපාලනය, විනිවිදභාවය සහ නීතිමය සහ නියාමන ආරක්ෂණ ඇතුළත් ඔවුන්ගේ දිගුකාලීන රහස්‍යතා මූලධර්ම හරහා පාරිභෝගික දත්ත ආරක්ෂා කර අනුකූල වන බව සහතික කිරීමට උපකාර කිරීමයි. මෙය මත ගොඩනැගීම සඳහා, අද ඔවුන් නිවේදනය කරන්නේ:

· ඔවුන්ගේ Azure OpenAI සේවා විස්පර් ආකෘතියේ පෙරදසුනෙහි රහස්‍ය අනුමාන කිරීම, එබැවින් පාරිභෝගිකයින්ට සත්‍යාපනය කළ හැකි අන්තයේ සිට අවසානය දක්වා පෞද්ගලිකත්වයට සහාය වන උත්පාදක AI යෙදුම් සංවර්ධනය කළ හැකිය. රහසිගත අනුමාන කිරීම මඟින් සංවේදී පාරිභෝගික දත්ත ආරක්ෂිතව සහ අනුමාන කිරීමේ ක්‍රියාවලිය අතරතුර පුද්ගලිකව පවතින බව සහතික කරයි, එනම් පුහුණු AI ආකෘතියක් නව දත්ත මත පදනම්ව පුරෝකථන හෝ තීරණ ගන්නා විටය. සෞඛ්‍ය සේවා, මූල්‍ය සේවා, සිල්ලර වෙළඳාම, නිෂ්පාදන සහ බලශක්තිය වැනි ඉහළ නියාමනය කරන ලද කර්මාන්ත සඳහා මෙය විශේෂයෙන් වැදගත් වේ.

· NVIDIA H100 Tensor Core GPU සහිත Azure රහස්‍ය VM වල සාමාන්‍ය ලබා ගැනීමේ හැකියාව, එමඟින් පාරිභෝගිකයින්ට GPU මත සෘජුවම දත්ත සුරක්ෂිත කිරීමට ඉඩ සලසයි. මෙය ඔවුන්ගේ රහස්‍ය පරිගණක විසඳුම් මත ගොඩනගා ඇති අතර එමඟින් පාරිභෝගික දත්ත සංකේතනය කර ආරක්ෂිත පරිසරයක ආරක්ෂිතව පවතින බව සහතික කරයි, එවිට කිසිවෙකුට අවසරයකින් තොරව තොරතුරු හෝ පද්ධතියට ප්‍රවේශය නොලැබේ.

· EU සහ US සඳහා Azure OpenAI දත්ත කලාප ඉක්මනින් පැමිණෙන අතර උත්පාදක AI යෙදුම්වල දත්ත සැකසීම සහ ගබඩා කිරීම කළමනාකරණය කිරීම පහසු කරමින් Azure OpenAI සේවාව මඟින් සපයනු ලබන පවතින දත්ත පදිංචිය මත ගොඩනැගේ. මෙම නව ක්‍රියාකාරිත්වය පාරිභෝගිකයින්ට භූගෝලීය ප්‍රදේශයක් තුළ සියලුම Azure කලාප හරහා උත්පාදක AI යෙදුම් පරිමාණය කිරීමේ නම්‍යශීලීභාවය ලබා දෙන අතරම, EU හෝ US තුළ දත්ත සැකසීම සහ ගබඩා කිරීම පාලනය කිරීමේ හැකියාව ඔවුන්ට ලබා දෙයි.

රහසිගත පරිගණකකරණය කෙරෙහි පාරිභෝගික උනන්දුව සහ රහසිගත GPU සඳහා උද්යෝගය වැඩි වීම Microsoft විසින් දැක ඇති අතර, එහි ආකෘති විශ්ලේෂණය කරන දත්තවල රහස්‍යභාවය සහතික කරන අතරම, උසස් AI-බලගැන්වූ ආරක්ෂක විසඳුම් ගොඩනැගීම සඳහා Azure NVIDIA H100 Tensor Core GPU සමඟ රහසිගත VM භාවිතා කරන යෙදුම් ආරක්ෂක සැපයුම්කරු F5 වෙතින් ද ඇතුළුව. සහ බහුජාතික බැංකු සංස්ථාවක් වන Royal Bank of Canada (RBC) Azure රහසිගත පරිගණකකරණය ඔවුන්ගේම වේදිකාවට ඒකාබද්ධ කර ඇති අතර, පාරිභෝගික පෞද්ගලිකත්වය ආරක්ෂා කරමින් සංකේතාත්මක දත්ත විශ්ලේෂණය කරයි. NVIDIA H100 Tensor Core GPU සමඟ Azure රහසිගත VM වල සාමාන්‍ය ලබා ගැනීමේ හැකියාව සමඟ, RBC හට දැන් මෙම උසස් AI මෙවලම් භාවිතා කර වඩාත් කාර්යක්ෂමව ක්‍රියා කිරීමට සහ වඩාත් බලවත් AI මාදිලි සංවර්ධනය කිරීමට හැකිය.

මයික්‍රොසොෆ්ට් ආයතනයට විශ්වාස කළ හැකි සහ අපේක්ෂා කරන සියලුම AI අවශ්‍ය වේ. සේවකයින්ගේ අත්දැකීම් පොහොසත් කිරීම සහ ව්‍යාපාර ක්‍රියාවලීන් නැවත සකස් කිරීමේ සිට පාරිභෝගික සහභාගීත්වය නැවත සොයා ගැනීම සහ ඔවුන්ගේ එදිනෙදා ජීවිතය නැවත සිතා බැලීම දක්වා විශ්වාසදායක ආකාරයකින් AI භාවිතා කිරීමට මිනිසුන්ට බලය ලබා දුන් විට කළ හැකි දේ ඔවුන් දැක තිබේ. ආරක්ෂාව, ආරක්ෂාව සහ පෞද්ගලිකත්වය වැඩිදියුණු කරන නව හැකියාවන් සමඟින්, මයික්‍රොසොෆ්ට් පාරිභෝගිකයින්ට ලෝකයේ සෑම පුද්ගලයෙකුටම සහ සංවිධානයකටම වැඩි යමක් අත්කර ගැනීමට උපකාරී වන විශ්වාසදායක AI විසඳුම් භාවිතා කිරීමට සහ ගොඩනැගීමට හැකියාව ලබා දෙයි. අවසාන වශයෙන්, විශ්වාසදායක AI ඔවුන් Microsoft හි කරන සියල්ල ආවරණය කරයි, ඔවුන් අවස්ථා පුළුල් කිරීමට, විශ්වාසය උපයා ගැනීමට, මූලික අයිතිවාසිකම් ආරක්ෂා කිරීමට සහ ඔවුන් කරන සෑම දෙයකම තිරසාරභාවය ඉදිරියට ගෙන යාමට කටයුතු කරන විට එය ඔවුන්ගේ මෙහෙවරට අත්‍යවශ්‍ය වේ.

විශ්වාසදායක උත්පාදක AI යෙදුම් ගොඩනැගීමට උපකාර කිරීම සඳහා Azure AI හි නව මෙවලම්

ගනුදෙනුකරුවන්ට AI ගුණාත්මකභාවය සහ ආරක්ෂක අභියෝගවලට මුහුණ දීමට උපකාර කිරීම සඳහා, ඔවුන් ජනක AI යෙදුම් සංවර්ධකයින් සඳහා Azure AI Studio වෙත දැන් ලබා ගත හැකි හෝ ඉක්මනින් පැමිණෙන නව මෙවලම් නිවේදනය කරයි:

· ඔබේ ආකෘතියට බලපෑම් කිරීමට පෙර වක්‍ර කඩිනම් ප්‍රහාර හඳුනා ගැනීම සඳහා නව ආකෘතියක් ඇතුළුව, කඩිනම් එන්නත් ප්‍රහාර හඳුනාගෙන අවහිර කිරීමට කඩිනම් පලිහ, ඉක්මනින් පැමිණෙන අතර දැන් Azure AI අන්තර්ගත ආරක්ෂාවේ පෙරදසුනෙහි ඇත.

· ආකෘති ප්‍රතිදානයන්හි “මායාවන්” හඳුනා ගැනීම සඳහා භූගත බව හඳුනා ගැනීම.

· ආරක්ෂිත, වගකිවයුතු ප්‍රතිදාන දෙසට ඔබේ ආකෘතියේ හැසිරීම මෙහෙයවීම සඳහා ආරක්ෂිත පද්ධති පණිවිඩ.

· Jailbreak ප්‍රහාරවලට සහ අන්තර්ගත අවදානම් ජනනය කිරීමට යෙදුමක අවදානම තක්සේරු කිරීමට ආරක්ෂිත ඇගයීම්, දැන් පෙරදසුනෙහි ඇත.

· කුමන ආකෘති ආදාන, ප්‍රතිදානයන් සහ අවසාන පරිශීලකයින් අන්තර්ගත පෙරහන් අවුලුවාලනවාද යන්න තේරුම් ගැනීමට අවදානම් සහ ආරක්ෂක අධීක්ෂණය Azure OpenAI සේවාවේ පෙරදසුනෙහි ඇත.

Nvidia Trustworthy AI

විශ්වාසනීය AI මූලධර්ම ඔවුන්ගේ අවසානය දක්වා සංවර්ධනයට පදනම් වන අතර හවුල්කරුවන්ට, ගනුදෙනුකරුවන්ට සහ සංවර්ධකයින්ට ඔවුන්ගේ හොඳම කාර්යය කිරීමට හැකි වන තාක්ෂණික විශිෂ්ටත්වය සඳහා අත්‍යවශ්‍ය වේ. ඔවුන් උත්පාදක AI සේවා සඳහා දත්ත කර්මාන්තශාලා ගොඩනඟමින්, පරිගණක දැක්ම සඳහා අපක්ෂපාතී දත්ත කට්ටල සකස් කිරීමට සහ වලංගු කිරීමට මෙවලම්, විශාල භාෂා ආකෘති සඳහා ස්වාභාවික භාෂා සැකසුම් දත්ත පරිමාණය කිරීම සඳහා පුස්තකාල, රහස්‍ය පරිගණකකරණය වැනි බිම් මට්ටමේ ආරක්ෂක විශේෂාංග සහ මානව ප්‍රතිපෝෂණ සමඟ ආකෘති පෙළගැස්වීම සඳහා විවෘත මූලාශ්‍ර ශිල්පීය ක්‍රම වැනි නවෝත්පාදනයන් ගොඩනඟමින් සිටිති. Nvidia හි විශ්වාසනීය AI විසඳුම්වලට ඇතුළත් වන්නේ:

Model card++

AI ආදර්ශ කාඩ්පතක් යනු යන්ත්‍ර ඉගෙනුම් (ML) ආකෘති ක්‍රියා කරන ආකාරය විස්තර කරන ලේඛනයකි. ආදර්ශ කාඩ්පත් මඟින් ML ආකෘතියේ පාර-දත්ත පිළිබඳ සවිස්තරාත්මක තොරතුරු සපයන අතර එය පදනම් වූ දත්ත කට්ටල, එය පුහුණු කරන ලද කාර්ය සාධන මිනුම් සහ ගැඹුරු ඉගෙනුම් පුහුණු ක්‍රමවේදය ද ඇතුළත් වේ. ආදර්ශ කාඩ්පත් ආකෘති විනිවිදභාවය සහ විශ්වසනීයත්වය දිරිමත් කිරීම සඳහා නිර්මාණය කර ඇති අතර, ඒවා සංවර්ධක අවබෝධය වැඩිදියුණු කිරීමට සහ තීරණ ගැනීමේ ක්‍රියාවලීන් ප්‍රමිතිකරණය කිරීමට කොටස්කරුවන් විසින් ද භාවිතා කරනු ලැබේ.

AI ආදර්ශ කාඩ්පත් සංවර්ධකයින් සඳහා පමණක් ගොඩනගා නොගත යුතුය; සමාගම් තාක්ෂණික නොවන පුද්ගලයින්ට සහ තාක්ෂණික විශේෂඥයින්ටඑකම ආකාරයෙන් ප්‍රවේශ විය හැකි සහ කියවිය හැකි ආදර්ශ කාඩ්පත් ද ගොඩනගා ගත යුතුය.

ආදර්ශ කාඩ්පත් හරහා AI බලමුලු ගැන්වීම සමාගම්වලට විශ්වාසදායක AI හි දියුණුව සඳහා ගත හැකි තීරණාත්මක සහ විනිවිද පෙනෙන පියවරකි(රූපය 03).

ආදර්ශ කාඩ්පත් දී ඇති කර්මාන්තයකට හෝ වසමකට සීමා නොවේ. ඒවා පරිගණක දැක්ම, කථනය, නිර්දේශ පද්ධති සහ අනෙකුත් AI වැඩ ප්‍රවාහ සඳහා භාවිතා කළ හැකිය. උසස් අධ්‍යාපනය සහ පර්යේෂණ සහ ඉහළ කාර්යසාධනයක් සහිත පරිගණක අවකාශයන්හි ක්‍රියාකාරී භාවිතයට අමතරව, ආකෘති කාඩ්පත් මෝටර් රථ, සෞඛ්‍ය සේවා සහ රොබෝ විද්‍යා යෙදුම් ඇතුළු බහු කර්මාන්ත හරහා උපයෝගීතාවයක් ඇත. ආකෘති කාඩ්පත්වලට:

· සිසුන්ට ඉගැන්වීම සහ සැබෑ ලෝක භාවිත අවස්ථා තේරුම් ගැනීමට ඔවුන්ට උපකාර කිරීම

· ප්‍රතිපත්ති සම්පාදකයින් දැනුවත් කිරීම සහ ආකෘති නොවන සංවර්ධකයින් සඳහා අපේක්ෂිත භාවිතය පැහැදිලි කිරීම

AI හි ප්‍රතිලාභ සෙවීමට උනන්දුවක් දක්වන අය දැනුවත් කිරීම

Omniverse Replicator

Omniverse Replicator යනු අභිරුචි කෘතිම දත්ත උත්පාදනය නල මාර්ග සහ සේවාවන් සංවර්ධනය කිරීම සඳහා රාමුවකි. ස්වයංක්‍රීය වාහන, රොබෝ විද්‍යාව සහ බුද්ධිමත් වීඩියෝ විශ්ලේෂණ යෙදුම්වල භාවිතා වන AI සංජානන ජාලවල පුහුණුව සහ කාර්ය සාධනය වැඩි දියුණු කිරීම සඳහා වටිනා ක්‍රමයක් ලෙස සේවය කරන භෞතිකව නිවැරදි 3D කෘතිම දත්ත ජනනය කළ හැකිය (රූපය 04).

Universal Scene Description (USD), PhysX, සහ Material Definition Language (MDL) වැනි විවෘත මූලාශ්‍ර ප්‍රමිතීන් භාවිතා කරමින් පවතින නල මාර්ග සමඟ පහසුවෙන් ඒකාබද්ධ වීමට Replicator නිර්මාණය කර ඇත.AI ආකෘති පුහුණු කිරීම සඳහා නිශ්චිත අවශ්‍යතා සපුරාලීම සඳහා විස්තර කරන්නන් සහ ලේඛකයින්ගේ විස්තීරණ ලේඛනයක් ඇත.

ඔවුන් දැන් GitHub හි පහසු Replicator උදාහරණ ලබා දෙයි. අන්තර්ගත උදාහරණ අවශ්‍ය වන කොටස්, සම්පූර්ණ ස්ක්‍රිප්ට් සහ USD දර්ශන ඇත.

NeMo Guardrails

NeMo Guardrails මඟින් සංවර්ධකයින්ට වැඩසටහන්ගත කළ හැකි නීති එකතු කිරීමෙන් චැට්බොට්වරුන්ට පහසුවෙන් මඟ පෙන්වීමට හැකියාව ලබා දෙයි. යෙදුමක් තුළ අපේක්ෂිත පරිශීලක අන්තර්ක්‍රියා නිර්වචනය කිරීම සඳහා. එය LLM-පාදක සංවාද යෙදුම් සඳහා ආරක්ෂාව, ආරක්ෂාව සහ විශ්වසනීයත්වයේ ස්ථරයක් එක් කරමින් LangChain සඳහා ස්වදේශීයව සහාය දක්වයි. පයිතන් පුස්තකාලයක් භාවිතයෙන් සංවර්ධකයින්ට පරිශීලක අන්තර්ක්‍රියා නිර්වචනය කළ හැකි අතර මෙම ආරක්ෂක වැටවල් ඕනෑම යෙදුමකට පහසුවෙන් ඒකාබද්ධ කළ හැකිය (රූපය 05).

NeMo සේවාවක් ලෙස ද ලබා ගත හැකිය. එය Nvidia AI පදනම්වල කොටසකි, ඔවුන්ගේම දත්ත කට්ටල සහ වසම් දැනුම මත පදනම්ව අභිරුචි උත්පාදක AI ආකෘති නිර්මාණය කිරීමට සහ ක්‍රියාත්මක කිරීමට කැමති ව්‍යාපාර සඳහා වලාකුළු සේවා පවුලකි.

විශ්වාසදායක, ආරක්ෂිත සහ ආරක්ෂිත LLM වල බලය සැමට ප්‍රවේශ විය හැකි බවට පත් කිරීම සඳහා AI ප්‍රජාව සමඟ වැඩ කිරීමට ඔවුන් බලාපොරොත්තු වේ.

Apple Trustworthy AI

ඇපල් සමාගම ආරක්ෂිත, ආරක්ෂිත සහ විශ්වාසදායක AI සංවර්ධනය කිරීම සඳහා ධවල මන්දිරයේ ස්වේච්ඡා කැපවීමට අත්සන් තබා ඇත. සමාගම ඉක්මනින්ම එහි උත්පාදක AI පිරිනැමීම වන Apple Intelligence දියත් කරනු ඇත, එය එහි මූලික නිෂ්පාදනවලට ඒකාබද්ධ කරමින්, තාක්ෂණය Apple හි බිලියන 2 ක පරිශීලකයින් ඉදිරියේ තබයි.

2023 ජූලි මාසයේදී උත්පාදක AI සංවර්ධනය කිරීම සඳහා ධවල මන්දිරයේ මාර්ගෝපදේශවලට කැප වූ Amazon, Anthropic, Google, Inflection, Meta, Microsoft, සහ OpenAI ඇතුළු තවත් තාක්ෂණ සමාගම් 15 ක් සමඟ Apple එක් වේ. ඒ වන විට, Apple තවමත් iOS තුළට AI ඒකාබද්ධ කිරීම සඳහා එහි සැලසුම් විස්තර කර නොතිබුණි, නමුත් ජූනි WWDC හිදී, එය එහි අභිප්‍රායන් පැහැදිලි කළේය: සමාගම උත්පාදක AI මත සම්පූර්ණයෙන්ම ඉදිරියට යන අතර, එය ChatGPT iPhone තුළට ඇතුළත් කරන හවුල්කාරිත්වයකින් ආරම්භ වේ. ෆෙඩරල් නියාමකයින්ගේ නිතර ඉලක්කයක් ලෙස, Apple සමාගමට අවශ්‍ය වන්නේ ධවල මන්දිරයේ AI නීති රීති පිළිපැදීමට ඇති කැමැත්ත කලින් ඇඟවීමට, AI පිළිබඳ අනාගත නියාමන සටන් ඇතිවීමට පෙර අනුග්‍රහය දැක්වීමට විය හැකිය.

Largest AI Companies by Market Cap (January 2025)

2025 ජනවාරි 14 වන විට, අපි ප්‍රසිද්ධියේ වෙළඳාම් කරන ලද AI සමාගම් 70 ක මුළු වෙළඳපල ප්‍රාග්ධනීකරණය ගණනය කර ඇති අතර එය ඩොලර් ට්‍රිලියන 15.556 කි.

අපගේ වෙළඳපල ප්‍රාග්ධනීකරණ දත්ත වලට අනුව, ඇපල් (AAPL) ඩොලර් ට්‍රිලියන 3.512 ක වෙළඳපල ප්‍රාග්ධනීකරණයකට ළඟා වූ අතර ගෝලීය AI සමාගම් අතර ඉහළම ස්ථානය ලබා ගන්නේය. NVIDIA (NVDA) සහ Microsoft (MSFT) ඉතා පසුපසින් සිටිති. දත්ත අවසන් වරට යාවත්කාලීන කරන ලද්දේ 2025 ජනවාරි 14 වන දින (රූපය 06).