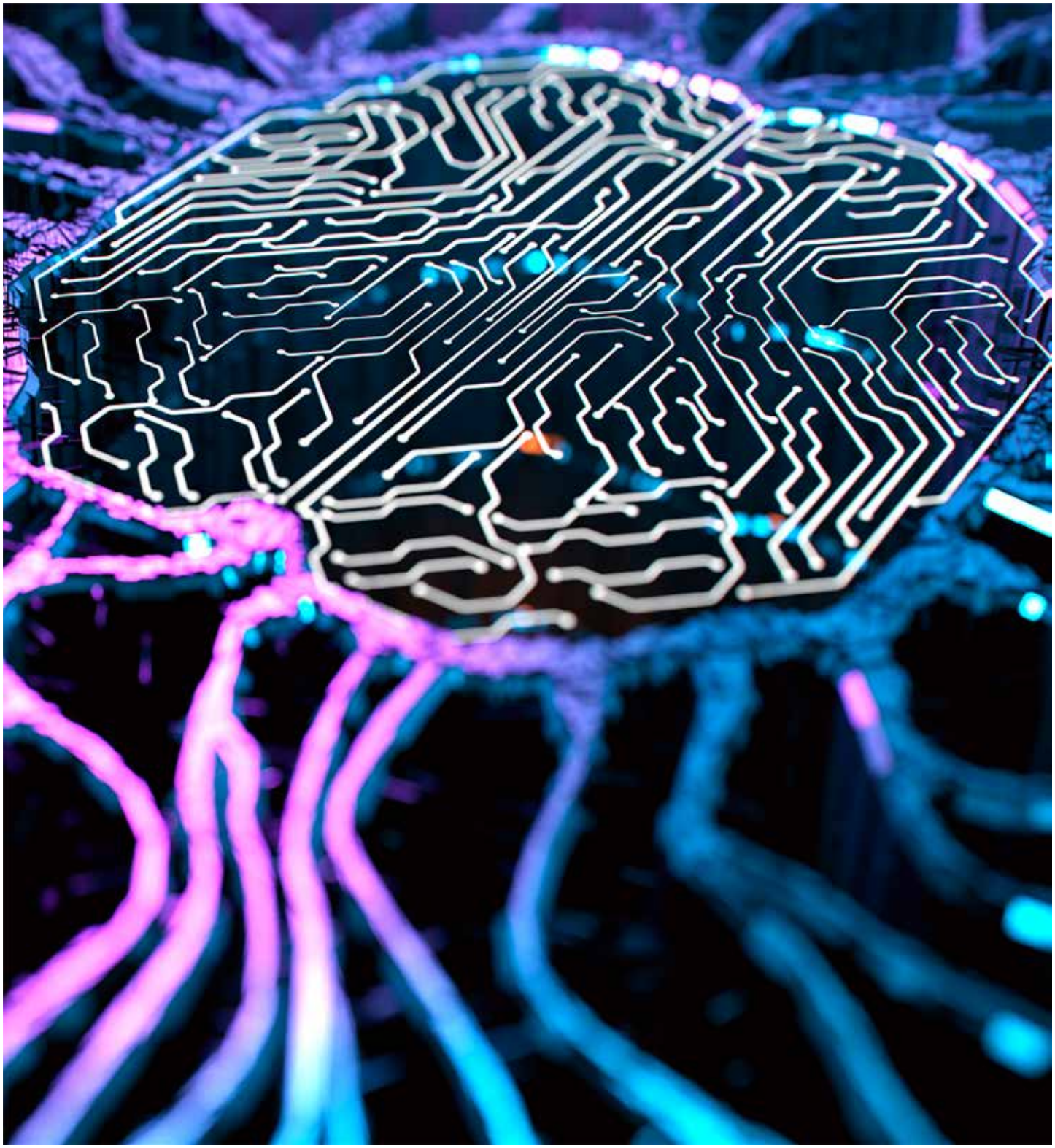
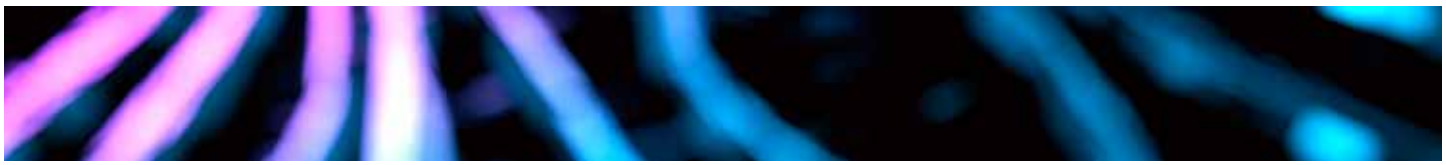
ਸੁਰੱਖਿਅਤ AI ਪ੍ਰਣਾਲੀ ਬਣਾਉਣ ਲਈ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼





National Cyber Security Centre
a part of GCHQ

CISA
CYBERSECURITY & INFRASTRUCTURE SECURITY AGENCY

Australian Government
Australian Signals Directorate

ASD AUSTRALIAN SIGNALS DIRECTORATE
ACSC Australian Cyber Security Centre

Communications Security Establishment
Canadian Centre for Cyber Security
Centre de la sécurité des télécommunications
Centre canadien pour la cybersécurité

CSIRT
Equipo de Respuesta ante Incidentes de Seguridad Informática

National Cyber and Information Security Agency

Republic of Estonia
Information System Authority

Department of Justice
Federal Bureau of Investigation

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Federal Office for Information Security

INCD
Israel National Cyber Directorate

AGENZIA PER LA CYBERSICUREZZA NAZIONALE
ACN
SAPIENTIA ADIUVAT

NISC
内閣サイバーセキュリティセンター
National center of Incident readiness and Strategy for Cybersecurity

National Cyber Security Centre

NITDA

NSM
NORWEGIAN NATIONAL CYBER SECURITY CENTRE

National Security Agency
United States of America

NASK

국가정보원
National Intelligence Service

Ministerstwo Cyfryzacji

CSA SINGAPORE
Cyber Security Agency of Singapore

ਇਸ ਦਸਤਾਵੇਜ਼ ਬਾਰੇ

ਇਹ ਦਸਤਾਵੇਜ਼ UK ਰਾਸ਼ਟਰੀ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਕੇਂਦਰ (NCSC), US ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਅਤੇ ਬੁਨਿਆਦੀ ਢਾਂਚਾ ਸੁਰੱਖਿਆ ਏਜੰਸੀ (CISA), ਅਤੇ ਹੇਠਾਂ ਦਿੱਤੇ ਅੰਤਰਰਾਸ਼ਟਰੀ ਭਾਈਵਾਲਾਂ ਦੁਆਰਾ ਪ੍ਰਕਾਸ਼ਿਤ ਕੀਤਾ ਗਿਆ ਹੈ:

- ਰਾਸ਼ਟਰੀ ਸੁਰੱਖਿਆ ਏਜੰਸੀ (NSA)
- ਫ਼ੈਡਰਲ ਬਿਊਰੋ ਆਫ਼ ਇਨਵੈਸਟੀਗੇਸ਼ਨ (FBI)
- ਆਸਟ੍ਰੇਲੀਅਨ ਸਿਗਨਲ ਡਾਇਰੈਕਟੋਰੇਟ ਦਾ ਆਸਟ੍ਰੇਲੀਅਨ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਕੇਂਦਰ (ACSC)
- ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਲਈ ਕੈਨੇਡੀਅਨ ਸੈਂਟਰ (CCCS)
- ਨਿਊਜ਼ੀਲੈਂਡ ਰਾਸ਼ਟਰੀ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਕੇਂਦਰ (NCSC-NZ)
- ਚਿਲੀ ਦੀ ਸਰਕਾਰ ਦਾ CSIRT
- ਚੈਕੀਆ ਦੀ ਰਾਸ਼ਟਰੀ ਸਾਈਬਰ ਅਤੇ ਸੂਚਨਾ ਸੁਰੱਖਿਆ ਏਜੰਸੀ (NUKIB)
- ਐਸਟੋਨੀਆ ਦੀ ਸੂਚਨਾ ਪ੍ਰਣਾਲੀ ਅਥਾਰਟੀ (RIA) ਅਤੇ ਐਸਟੋਨੀਆ ਦਾ ਰਾਸ਼ਟਰੀ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਕੇਂਦਰ (NCSC-EE)
- ਫਰਾਂਸੀਸੀ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਏਜੰਸੀ (ANSSI)
- ਸੂਚਨਾ ਸੁਰੱਖਿਆ ਲਈ ਜਰਮਨੀ ਦਾ ਸੰਘੀ ਦਫ਼ਤਰ (BSI)
- ਇਜ਼ਰਾਈਲੀ ਰਾਸ਼ਟਰੀ ਸਾਈਬਰ ਡਾਇਰੈਕਟੋਰੇਟ (INCD)
- ਇਟਲੀ ਦੀ ਰਾਸ਼ਟਰੀ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਏਜੰਸੀ (ACN)
- ਜਾਪਾਨ ਦਾ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਘਟਨਾ ਤਿਆਰੀ ਅਤੇ ਰਣਨੀਤੀ ਲਈ ਰਾਸ਼ਟਰੀ ਕੇਂਦਰ (NISC)
- ਜਾਪਾਨ ਦਾ ਵਿਗਿਆਨ, ਤਕਨਾਲੋਜੀ ਅਤੇ ਨਵੀਨਤਾ ਨੀਤੀ ਦਾ ਸਕੱਤਰੇਤ, ਕੈਬਨਿਟ ਦਫ਼ਤਰ
- ਨਾਈਜੀਰੀਆ ਦੀ ਰਾਸ਼ਟਰੀ ਸੂਚਨਾ ਤਕਨਾਲੋਜੀ ਵਿਕਾਸ ਏਜੰਸੀ (NITDA)
- ਨਾਰਵੇ ਦਾ ਰਾਸ਼ਟਰੀ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਕੇਂਦਰ (NCSC-NO)
- ਪੋਲੈਂਡ ਦਾ ਡਿਜ਼ੀਟਲ ਮਾਮਲਿਆਂ ਦਾ ਮੰਤਰਾਲਾ
- ਪੋਲੈਂਡ ਦਾ NASK ਨੈਸ਼ਨਲ ਰਿਸਰਚ ਇੰਸਟੀਚਿਊਟ (NASK)
- ਕੋਰੀਆ ਦੀ ਗਣਰਾਜ ਨੈਸ਼ਨਲ ਇੰਟੈਲੀਜੈਂਸ ਸਰਵਿਸ (NIS)
- ਸਿੰਗਾਪੁਰ ਦੀ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਏਜੰਸੀ (CSA)

ਮਾਨਤਾਵਾਂ

ਹੇਠ ਲਿਖੀਆਂ ਸੰਸਥਾਵਾਂ ਨੇ ਇਹਨਾਂ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ਾਂ ਦੇ ਵਿਕਾਸ ਵਿੱਚ ਯੋਗਦਾਨ ਪਾਇਆ:

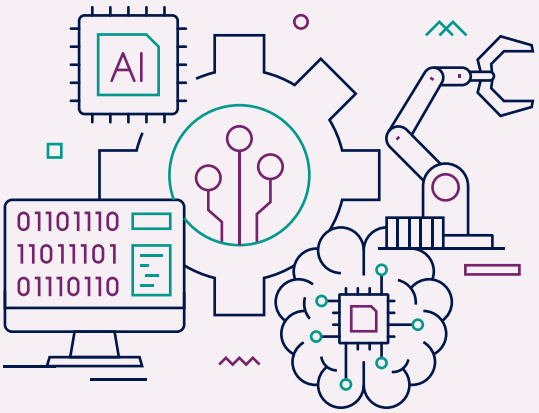
- ਐਲਨ ਟਿਊਰਿੰਗ ਇੰਸਟੀਚਿਊਟ
- Anthropic
- Databricks
- ਜਾਰਜਟਾਊਨ ਯੂਨੀਵਰਸਿਟੀ ਦਾ ਸੈਂਟਰ ਫਾਰ ਸਕਿਓਰਿਟੀ ਐਂਡ ਐਮਰਜਿੰਗ ਟੈਕਨਾਲੋਜੀ
- Google
- Google DeepMind
- IBM
- ImBue
- Microsoft
- OpenAI
- Palantir
- RAND
- Scale AI
- ਕਾਰਨੇਗੀ ਮੈਲਨ ਯੂਨੀਵਰਸਿਟੀ ਦਾ ਸਾਫ਼ਟਵੇਅਰ ਇੰਜੀਨੀਅਰਿੰਗ ਇੰਸਟੀਚਿਊਟ
- ਸਟੈਨਫੋਰਡ ਸੈਂਟਰ ਫਾਰ AI ਸੇਫ਼ਟੀ
- ਭੂ-ਰਾਜਨੀਤੀ, ਤਕਨਾਲੋਜੀ ਅਤੇ ਸ਼ਾਸਨ 'ਤੇ ਸਟੈਨਫੋਰਡ ਪ੍ਰੋਗਰਾਮ

ਬੇਦਾਅਵਾ

ਇਸ ਦਸਤਾਵੇਜ਼ ਵਿਚਲੀ ਜਾਣਕਾਰੀ NCSC ਅਤੇ ਲੇਖਕ ਸੰਗਠਨਾਂ ਦੁਆਰਾ "ਜਿਵੇਂ ਹੈ ਉਵੇਂ ਹੀ" ਪ੍ਰਦਾਨ ਕੀਤੀ ਗਈ ਹੈ ਜੋ ਕਾਨੂੰਨ ਦੁਆਰਾ ਲੋੜੀਂਦੇ ਹੋਣ ਤੋਂ ਇਲਾਵਾ ਇਸਦੀ ਵਰਤੋਂ ਕਾਰਨ ਹੋਏ ਕਿਸੇ ਵੀ ਘਾਟੇ, ਸੱਟ ਜਾਂ ਨੁਕਸਾਨ ਲਈ ਜ਼ਿੰਮੇਵਾਰ ਨਹੀਂ ਹੋਣਗੇ। ਇਸ ਦਸਤਾਵੇਜ਼ ਵਿਚਲੀ ਜਾਣਕਾਰੀ NCSC ਅਤੇ ਅਧਿਕਾਰਤ ਏਜੰਸੀਆਂ ਦੁਆਰਾ ਕਿਸੇ ਵੀ ਤੀਜੀ ਧਿਰ ਦੀ ਸੰਸਥਾ, ਉਤਪਾਦ, ਜਾਂ ਸੇਵਾ ਦੀ ਪੁਸ਼ਟੀ ਜਾਂ ਸਿਫ਼ਾਰਸ਼ ਨਹੀਂ ਕਰਦੀ ਜਾਂ ਉਨ੍ਹਾਂ ਵੱਲ ਸੰਕੇਤ ਨਹੀਂ ਕਰਦੀ ਹੈ। ਵੈੱਬਸਾਈਟਾਂ ਅਤੇ ਤੀਜੀ ਧਿਰ ਦੀਆਂ ਸਮੱਗਰੀਆਂ ਦੇ ਲਿੰਕ ਅਤੇ ਹਵਾਲੇ ਸਿਰਫ਼ ਜਾਣਕਾਰੀ ਲਈ ਦਿੱਤੇ ਗਏ ਹਨ ਅਤੇ ਬਾਕੀਆਂ ਉੱਪਰ ਅਜਿਹੇ ਸਰੋਤਾਂ ਦੀ ਤਸਦੀਕ ਜਾਂ ਸਿਫ਼ਾਰਸ਼ ਨਹੀਂ ਦਰਸਾਉਂਦੇ ਹਨ।

ਇਹ ਦਸਤਾਵੇਜ਼ TLP:CLEAR ਆਧਾਰ 'ਤੇ ਉਪਲਬਧ ਕਰਵਾਇਆ ਗਿਆ ਹੈ (https://www.first.org/tlp/)।

ਤਤਕਰਾ



ਕਾਰਜਕਾਰੀ ਸੰਖੇਪ-ਸਾਰ

ਇਹ ਦਸਤਾਵੇਜ਼ ਕਿਸੇ ਵੀ ਅਜਿਹੀ ਪ੍ਰਣਾਲੀ ਦੇ ਪ੍ਰਦਾਤਾਵਾਂ ਲਈ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ਾਂ ਦੀ ਸਿਫ਼ਾਰਸ਼ ਕਰਦਾ ਹੈ ਜੋ ਆਰਟੀਫੀਸ਼ੀਅਲ ਇੰਟੈਲੀਜੈਂਸ (AI) ਦੀ ਵਰਤੋਂ ਕਰਦੇ ਹਨ, ਭਾਵੇਂ ਉਹ ਪ੍ਰਣਾਲੀ ਬਿਲਕੁਲ ਨਵੇਂ ਸਿਰੇ ਤੋਂ ਬਣਾਈ ਗਈ ਹੈ ਜਾਂ ਦੂਜਿਆਂ ਦੁਆਰਾ ਪ੍ਰਦਾਨ ਕੀਤੇ ਗਏ ਟੂਲਾਂ ਅਤੇ ਸੇਵਾਵਾਂ ਉੱਪਰ ਨਿਰਭਰ ਕਰਦਿਆਂ ਹੋਇਆ ਬਣਾਈ ਗਈ ਹੈ। ਇਹਨਾਂ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ਾਂ ਨੂੰ ਲਾਗੂ ਕਰਨ ਨਾਲ ਪ੍ਰਦਾਤਾਵਾਂ ਨੂੰ ਅਜਿਹੀਆਂ AI ਪ੍ਰਣਾਲੀਆਂ ਬਣਾਉਣ ਵਿੱਚ ਮੱਦਦ ਮਿਲੇਗੀ ਜੋ ਤੈਅ ਕੀਤੇ ਉਦੇਸ਼ ਅਨੁਸਾਰ ਕੰਮ ਕਰਦੀਆਂ ਹਨ, ਲੋੜ ਪੈਣ 'ਤੇ ਉਪਲਬਧ ਹੁੰਦੀਆਂ ਹਨ, ਅਤੇ ਅਣਅਧਿਕਾਰਤ ਧਿਰਾਂ ਨੂੰ ਸੰਵੇਦਨਸ਼ੀਲ ਡੇਟਾ ਦਾ ਖੁਲਾਸਾ ਕੀਤੇ ਬਿਨਾਂ ਕੰਮ ਕਰਦੀਆਂ ਹਨ।

ਇਸ ਦਸਤਾਵੇਜ਼ ਦਾ ਉਦੇਸ਼ ਮੁੱਖ ਤੌਰ 'ਤੇ AI ਪ੍ਰਣਾਲੀਆਂ ਦੇ ਪ੍ਰਦਾਤਾ ਹਨ ਜੋ ਕਿਸੇ ਸੰਸਥਾ ਦੁਆਰਾ ਆਯੋਜਿਤ ਕੀਤੇ ਮਾਡਲਾਂ ਦੀ ਵਰਤੋਂ ਕਰ ਰਹੇ ਹਨ, ਜਾਂ ਬਾਹਰੀ ਐਪਲੀਕੇਸ਼ਨ ਪ੍ਰੋਗਰਾਮਿੰਗ ਇੰਟਰਫੇਸ (APIs) ਦੀ ਵਰਤੋਂ ਕਰ ਰਹੇ ਹਨ। ਅਸੀਂ **ਸਾਰੇ** ਹਿੱਤਧਾਰਕਾਂ (ਡੇਟਾ ਵਿਗਿਆਨੀਆਂ, ਡਿਵੈਲਪਰਾਂ, ਮੈਨੇਜਰਾਂ, ਫ਼ੈਸਲਾ ਲੈਣ ਵਾਲਿਆਂ ਅਤੇ ਜ਼ੋਖਮ ਦੇ ਮਾਲਕਾਂ ਸਮੇਤ) ਨੂੰ ਆਪਣੀਆਂ AI ਪ੍ਰਣਾਲੀਆਂ ਦੇ **ਡਿਜ਼ਾਈਨ**, **ਵਿਕਾਸ**, **ਤੈਨਾਤੀ** ਅਤੇ **ਸੰਚਾਲਨ** ਬਾਰੇ ਜਾਣਕਾਰੀ ਭਰਪੂਰ ਫ਼ੈਸਲੇ ਲੈਣ ਵਿੱਚ ਮੱਦਦ ਕਰਨ ਲਈ ਇਹਨਾਂ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ਾਂ ਨੂੰ ਪੜ੍ਹਨ ਲਈ ਬੇਨਤੀ ਕਰਦੇ ਹਾਂ।

ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ਾਂ ਬਾਰੇ

AI ਪ੍ਰਣਾਲੀਆਂ ਵਿੱਚ ਸਮਾਜ ਲਈ ਬਹੁਤ ਸਾਰੇ ਲਾਭ ਲਿਆਉਣ ਦੀ ਸਮਰੱਥਾ ਹੈ। ਹਾਲਾਂਕਿ, AI ਦੇ ਮੌਕਿਆਂ ਨੂੰ ਪੂਰੀ ਤਰ੍ਹਾਂ ਸਾਕਾਰ ਕਰਨ ਲਈ, ਇਸਨੂੰ ਸੁਰੱਖਿਅਤ ਅਤੇ ਜ਼ਿੰਮੇਵਾਰ ਤਰੀਕੇ ਨਾਲ ਵਿਕਸਤ, ਤੈਨਾਤ ਅਤੇ ਸੰਚਾਲਿਤ ਕੀਤਾ ਜਾਣਾ ਲਾਜ਼ਮੀ ਹੈ।

AI ਪ੍ਰਣਾਲੀਆਂ ਨਵੀਆਂ ਕਿਸਮਾਂ ਦੀਆਂ ਸੁਰੱਖਿਆ ਕਮਜ਼ੋਰੀਆਂ ਦੇ ਅਧੀਨ ਹਨ ਜਿਨ੍ਹਾਂ ਨੂੰ ਮਿਆਰੀ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਖ਼ਤਰਿਆਂ ਦੇ ਨਾਲ ਵਿਚਾਰਿਆ ਜਾਣਾ ਚਾਹੀਦਾ ਹੈ। ਜਦੋਂ ਵਿਕਾਸ ਦੀ ਰਫ਼ਤਾਰ ਤੇਜ਼ ਹੁੰਦੀ ਹੈ - ਜਿਵੇਂ ਕਿ AI ਨਾਲ ਹੈ - ਸੁਰੱਖਿਆ ਅਕਸਰ ਇੱਕ ਦੂਜੇ ਦਰਜੇ ਦਾ (ਘੱਟ ਮਹੱਤਵਪੂਰਨ) ਵਿਚਾਰ ਹੋ ਸਕਦੀ ਹੈ। ਸੁਰੱਖਿਆ ਨਾ ਸਿਰਫ਼ ਵਿਕਾਸ ਦੇ ਪੜਾਅ ਵਿੱਚ, ਸਗੋਂ ਇਸ ਪ੍ਰਣਾਲੀ ਦੇ ਪੂਰੇ ਜੀਵਨ ਚੱਕਰ ਦੌਰਾਨ ਇੱਕ ਮੁੱਖ ਲੋੜ ਹੋਣੀ ਲਾਜ਼ਮੀ ਹੈ।

ਇਸ ਕਾਰਨ ਕਰਕੇ, ਇਹਨਾਂ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ਾਂ ਨੂੰ AI ਪ੍ਰਣਾਲੀ ਦੇ ਵਿਕਾਸ ਦੇ ਜੀਵਨ ਚੱਕਰ ਵਿੱਚ ਚਾਰ ਮੁੱਖ ਖੇਤਰਾਂ ਵਿੱਚ ਵੰਡਿਆ ਗਿਆ ਹੈ: **ਸੁਰੱਖਿਅਤ ਡਿਜ਼ਾਈਨ**, **ਸੁਰੱਖਿਅਤ ਵਿਕਾਸ**, **ਸੁਰੱਖਿਅਤ ਤੈਨਾਤੀ**, ਅਤੇ **ਸੁਰੱਖਿਅਤ ਸੰਚਾਲਨ ਅਤੇ ਰੱਖ-ਰਖਾਅ**। ਹਰੇਕ ਭਾਗ ਲਈ ਅਸੀਂ ਵਿਚਾਰਨ ਯੋਗ ਨੁਕਤਿਆਂ ਅਤੇ ਕਮੀਆਂ ਦਾ ਸੁਝਾਅ ਦਿੰਦੇ ਹਾਂ ਜੋ ਸੰਸਥਾਤਮਕ AI ਪ੍ਰਣਾਲੀ ਦੀ ਵਿਕਾਸ ਪ੍ਰਕਿਰਿਆ ਦੇ ਸਮੁੱਚੇ ਜ਼ੋਖਮ ਨੂੰ ਘਟਾਉਣ ਵਿੱਚ ਮੱਦਦ ਕਰਨਗੀਆਂ।

1. **ਸੁਰੱਖਿਅਤ ਡਿਜ਼ਾਈਨ**
   ਇਸ ਭਾਗ ਵਿੱਚ ਉਹ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ ਸ਼ਾਮਲ ਹਨ ਜੋ AI ਪ੍ਰਣਾਲੀ ਦੇ ਵਿਕਾਸ ਜੀਵਨ ਚੱਕਰ ਦੇ ਡਿਜ਼ਾਈਨ ਪੜਾਅ 'ਤੇ ਲਾਗੂ ਹੁੰਦੇ ਹਨ। ਇਸ ਵਿੱਚ ਪ੍ਰਣਾਲੀ ਅਤੇ ਮਾਡਲ ਡਿਜ਼ਾਈਨ 'ਤੇ ਵਿਚਾਰ ਕਰਨ ਲਈ ਜ਼ੋਖਮਾਂ ਅਤੇ ਖ਼ਤਰੇ ਦੀ ਮਾਡਲਿੰਗ ਦੇ ਨਾਲ-ਨਾਲ ਖ਼ਾਸ ਵਿਸ਼ਿਆਂ ਅਤੇ ਟ੍ਰੇਡ-ਆਫ (ਕੁੱਝ ਚੰਗਾ ਕਰਨ ਲਈ ਕੁੱਝ ਬੁਰਾ ਸਵੀਕਾਰ ਕਰਨ) ਨੂੰ ਸਮਝਣਾ ਵੀ ਸ਼ਾਮਲ ਹੈ।

2. **ਸੁਰੱਖਿਅਤ ਵਿਕਾਸ**
   ਇਸ ਭਾਗ ਵਿੱਚ ਅਜਿਹੇ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ ਸ਼ਾਮਲ ਹਨ ਜੋ AI ਪ੍ਰਣਾਲੀ ਵਿਕਸਤ ਕਰਨ ਦੇ ਜੀਵਨ ਚੱਕਰ ਦੇ ਵਿਕਾਸ ਪੜਾਅ 'ਤੇ ਲਾਗੂ ਹੁੰਦੇ ਹਨ, ਜਿਸ ਵਿੱਚ ਸਪਲਾਈ ਚੇਨ ਸੁਰੱਖਿਆ, ਦਸਤਾਵੇਜ਼, ਅਤੇ ਸੰਪਤੀ ਅਤੇ ਤਕਨੀਕੀ ਕਰਜ਼ਾ ਪ੍ਰਬੰਧਨ ਸ਼ਾਮਲ ਹਨ।

3. **ਸੁਰੱਖਿਅਤ ਤੈਨਾਤੀ**
   ਇਸ ਭਾਗ ਵਿੱਚ ਅਜਿਹੇ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ ਸ਼ਾਮਲ ਹਨ ਜੋ AI ਪ੍ਰਣਾਲੀ ਵਿਕਸਤ ਕਰਨ ਦੇ ਜੀਵਨ ਚੱਕਰ ਦੇ ਤੈਨਾਤੀ ਪੜਾਅ 'ਤੇ ਲਾਗੂ ਹੁੰਦੇ ਹਨ, ਜਿਸ ਵਿੱਚ ਬੁਨਿਆਦੀ ਢਾਂਚੇ ਅਤੇ ਮਾਡਲਾਂ ਨੂੰ ਅਣਅਧਿਕਾਰਤ ਪਹੁੰਚ, ਖ਼ਤਰੇ ਜਾਂ ਨੁਕਸਾਨ ਤੋਂ ਬਚਾਉਣਾ, ਘਟਨਾ ਪ੍ਰਬੰਧਨ ਦੀਆਂ ਪ੍ਰਕਿਰਿਆਵਾਂ ਦਾ ਵਿਕਾਸ ਕਰਨਾ, ਅਤੇ ਜ਼ਿੰਮੇਵਾਰੀ ਨਾਲ ਜਾਣਕਾਰੀ ਦੇ ਖੁਲਾਸੇ ਕਰਨੇ ਸ਼ਾਮਿਲ ਹਨ।

4. **ਸੁਰੱਖਿਅਤ ਸੰਚਾਲਨ ਅਤੇ ਰੱਖ-ਰਖਾਅ**
   ਇਸ ਭਾਗ ਵਿੱਚ ਅਜਿਹੇ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ ਹਨ ਜੋ AI ਪ੍ਰਣਾਲੀ ਵਿਕਸਤ ਕਰਨ ਦੇ ਜੀਵਨ ਚੱਕਰ ਦੇ ਸੁਰੱਖਿਅਤ ਸੰਚਾਲਨ ਅਤੇ ਰੱਖ-ਰਖਾਅ ਦੇ ਪੜਾਅ 'ਤੇ ਲਾਗੂ ਹੁੰਦੇ ਹਨ। ਇਹ ਲੌਗਿੰਗ (ਰੋਜ਼ਨਾਮਚੇ ਵਿਚ ਦਰਜ ਕਰਨ) ਅਤੇ ਨਿਗਰਾਨੀ, ਅੱਪਡੇਟ ਪ੍ਰਬੰਧਨ ਅਤੇ ਜਾਣਕਾਰੀ ਸਾਂਝਾ ਕਰਨ ਸਮੇਤ, ਇਸ ਪ੍ਰਣਾਲੀ ਦੇ ਲਾਗੂ ਹੋਣ ਤੋਂ ਬਾਅਦ ਖ਼ਾਸ ਤੌਰ 'ਤੇ ਢੁੱਕਵੀਆਂ ਕਾਰਵਾਈਆਂ ਬਾਰੇ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ ਪ੍ਰਦਾਨ ਕਰਦਾ ਹੈ।

ਇਹ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ ਇੱਕ 'ਪਹਿਲਾਂ-ਤੋਂ-ਨਿਰਧਾਰਤ ਸੁਰੱਖਿਅਤ' ਪਹੁੰਚ ਦੀ ਪਾਲਣਾ ਕਰਦੇ ਹਨ, ਅਤੇ NCSC ਦੇ ਸੁਰੱਖਿਅਤ ਵਿਕਾਸ ਅਤੇ ਤੈਨਾਤੀ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼, NIST ਦੇ ਸੁਰੱਖਿਅਤ ਸਾਫ਼ਟਵੇਅਰ ਡਿਵੈਲਪਮੈਂਟ ਫਰੇਮਵਰਕ, ਅਤੇ CISA, NCSC ਅਤੇ ਅੰਤਰਰਾਸ਼ਟਰੀ ਸਾਈਬਰ ਏਜੰਸੀਆਂ ਦੁਆਰਾ ਪ੍ਰਕਾਸ਼ਿਤ 'ਡਿਜ਼ਾਇਨ ਦੁਆਰਾ ਸੁਰੱਖਿਅਤ ਸਿਧਾਂਤਾਂ' ਵਿੱਚ ਪਰਿਭਾਸ਼ਿਤ ਕੰਮ ਕਰਨ ਦੇ ਤਰੀਕਿਆਂ ਨਾਲ ਨੇੜਿਓਂ ਜੁੜੇ ਹੋਏ ਹਨ। ਉਹ ਪਹਿਲ ਦਿੰਦੇ ਹਨ:

- ਗਾਹਕਾਂ ਲਈ ਸੁਰੱਖਿਆ ਨਤੀਜਿਆਂ ਦੀ ਜ਼ਿੰਮੇਵਾਰੀ ਲੈਣ ਨੂੰ
- ਬੁਨਿਆਦੀ ਤਬਦੀਲੀ ਲਿਆਉਣ ਵਾਲੀ ਪਾਰਦਰਸ਼ਤਾ ਅਤੇ ਜਵਾਬਦੇਹੀ ਨੂੰ ਗਲੇ ਲਗਾਉਣ ਨੂੰ
- ਸੰਸਥਾਤਮਕ ਢਾਂਚੇ ਅਤੇ ਲੀਡਰਸ਼ਿਪ ਨਿਰਮਾਣ ਨੂੰ, ਇਸ ਲਈ ਡਿਜ਼ਾਈਨ ਦੁਆਰਾ ਸੁਰੱਖਿਅਤ ਬਣਾਉਣਾ ਇੱਕ ਪ੍ਰਮੁੱਖ ਵਪਾਰਕ ਪਹਿਲ ਹੁੰਦਾ ਹੈ

ਜਾਣ-ਪਛਾਣ

ਆਰਟੀਫੀਸ਼ੀਅਲ ਇੰਟੈਲੀਜੈਂਸ (AI) ਪ੍ਰਣਾਲੀਆਂ ਵਿੱਚ ਸਮਾਜ ਲਈ ਬਹੁਤ ਸਾਰੇ ਲਾਭ ਲਿਆਉਣ ਦੀ ਸਮਰੱਥਾ ਹੈ। ਹਾਲਾਂਕਿ, AI ਦੇ ਮੌਕਿਆਂ ਨੂੰ ਪੂਰੀ ਤਰ੍ਹਾਂ ਸਾਕਾਰ ਕਰਨ ਲਈ, ਇਸਨੂੰ ਸੁਰੱਖਿਅਤ ਅਤੇ ਜ਼ਿੰਮੇਵਾਰ ਤਰੀਕੇ ਨਾਲ ਵਿਕਸਤ, ਤੈਨਾਤ ਅਤੇ ਸੰਚਾਲਿਤ ਕੀਤਾ ਜਾਣਾ ਲਾਜ਼ਮੀ ਹੈ। AI ਪ੍ਰਣਾਲੀਆਂ ਦੀ ਸੁਰੱਖਿਆ, ਲਚਕਤਾ, ਗੁਪਤਤਾ, ਨਿਰਪੱਖਤਾ, ਪ੍ਰਭਾਵਸ਼ੀਲਤਾ ਅਤੇ ਭਰੋਸੇਯੋਗਤਾ ਲਈ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਇੱਕ ਜ਼ਰੂਰੀ ਪੂਰਵ-ਸ਼ਰਤ ਹੈ।

ਹਾਲਾਂਕਿ, AI ਪ੍ਰਣਾਲੀਆਂ ਨਵੀਆਂ ਕਿਸਮਾਂ ਦੀਆਂ ਸੁਰੱਖਿਆ ਕਮਜ਼ੋਰੀਆਂ ਦੇ ਅਧੀਨ ਹਨ ਜਿਨ੍ਹਾਂ ਨੂੰ ਮਿਆਰੀ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਖ਼ਤਰਿਆਂ ਦੇ ਨਾਲ-ਨਾਲ ਵਿਚਾਰਿਆ ਜਾਣਾ ਚਾਹੀਦਾ ਹੈ। ਜਦੋਂ ਵਿਕਾਸ ਦੀ ਰਫ਼ਤਾਰ ਤੇਜ਼ ਹੁੰਦੀ ਹੈ - ਜਿਵੇਂ ਕਿ AI ਨਾਲ ਹੈ - ਸੁਰੱਖਿਆ ਅਕਸਰ ਇੱਕ ਦੂਜੇ ਦਰਜੇ ਦਾ (ਘੱਟ ਮਹੱਤਵਪੂਰਨ) ਵਿਚਾਰ ਹੋ ਸਕਦੀ ਹੈ। ਸੁਰੱਖਿਆ ਨਾ ਸਿਰਫ਼ ਵਿਕਾਸ ਦੇ ਪੜਾਅ ਵਿੱਚ, ਸਗੋਂ ਇਸ ਪ੍ਰਣਾਲੀ ਦੇ ਪੂਰੇ ਜੀਵਨ ਚੱਕਰ ਦੌਰਾਨ ਇੱਕ ਮੁੱਖ ਲੋੜ ਹੋਣੀ ਲਾਜ਼ਮੀ ਹੈ।

ਇਹ ਦਸਤਾਵੇਜ਼ ਕਿਸੇ ਵੀ ਅਜਿਹੀ ਪ੍ਰਣਾਲੀ ਦੇ ਪ੍ਰਦਾਤਾਵਾਂ ਲਈ 'ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ਾਂ ਦੀ ਸਿਫ਼ਾਰਸ਼ ਕਰਦਾ ਹੈ ਜੋ AI ਦੀ ਵਰਤੋਂ ਕਰਦੇ ਹਨ, ਭਾਵੇਂ ਉਹ ਪ੍ਰਣਾਲੀ ਬਿਲਕੁਲ ਨਵੇਂ ਸਿਰੇ ਤੋਂ ਬਣਾਈ ਗਈ ਹੈ ਜਾਂ ਦੂਜਿਆਂ ਦੁਆਰਾ ਪ੍ਰਦਾਨ ਕੀਤੇ ਗਏ ਟੂਲਾਂ ਅਤੇ ਸੇਵਾਵਾਂ ਉੱਪਰ ਨਿਰਭਰ ਕਰਦਿਆਂ ਹੋਇਆ ਬਣਾਈ ਗਈ ਹੈ। ਇਹਨਾਂ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ਾਂ ਨੂੰ ਲਾਗੂ ਕਰਨ ਨਾਲ ਪ੍ਰਦਾਤਾਵਾਂ ਨੂੰ ਅਜਿਹੀਆਂ AI ਪ੍ਰਣਾਲੀਆਂ ਬਣਾਉਣ ਵਿੱਚ ਮੱਦਦ ਮਿਲੇਗੀ ਜੋ ਤੈਅ ਕੀਤੇ ਅਨੁਸਾਰ ਕੰਮ ਕਰਦੀਆਂ ਹਨ, ਲੋੜ ਪੈਣ 'ਤੇ ਉਪਲਬਧ ਹੁੰਦੀਆਂ ਹਨ, ਅਤੇ ਅਣਅਧਿਕਾਰਤ ਧਿਰਾਂ ਨੂੰ ਸੰਵੇਦਨਸ਼ੀਲ ਡੇਟਾ ਦਾ ਖੁਲਾਸਾ ਕੀਤੇ ਬਿਨਾਂ ਕੰਮ ਕਰਦੀਆਂ ਹਨ।

ਇਹਨਾਂ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ਾਂ ਨੂੰ ਪਹਿਲਾਂ ਤੋਂ ਬਣੀ ਹੋਈ ਸਾਈਬਰ ਸੁਰੱਖਿਆ, ਜ਼ੋਖਮ ਪ੍ਰਬੰਧਨ, ਅਤੇ ਘਟਨਾ ਪ੍ਰਤੀ ਜਵਾਬ ਦੇਣ ਦੇ ਸਭ ਤੋਂ ਵਧੀਆ ਤਰੀਕੇ ਨਾਲ ਜੋੜਕੇ ਵਿਚਾਰਿਆ ਜਾਣਾ ਚਾਹੀਦਾ ਹੈ। ਖ਼ਾਸ ਤੌਰ 'ਤੇ, ਅਸੀਂ ਪ੍ਰਦਾਤਾਵਾਂ ਨੂੰ US ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਅਤੇ ਬੁਨਿਆਦੀ ਢਾਂਚਾ ਸੁਰੱਖਿਆ ਏਜੰਸੀ (CISA), UK ਨੈਸ਼ਨਲ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਕੇਂਦਰ (NCSC), ਅਤੇ ਸਾਡੇ ਸਾਰੇ ਅੰਤਰਰਾਸ਼ਟਰੀ ਭਾਈਵਾਲਾਂ ਦੁਆਰਾ ਵਿਕਸਿਤ ਕੀਤੇ 'ਡਿਜ਼ਾਈਨ ਦੁਆਰਾ ਸੁਰੱਖਿਅਤ'[2] ਸਿਧਾਂਤਾਂ ਦੀ ਪਾਲਣਾ ਕਰਨ ਦੀ ਅਪੀਲ ਕਰਦੇ ਹਾਂ। ਇਹ ਸਿਧਾਂਤ ਪਹਿਲ ਦਿੰਦੇ ਹਨ:

- ਗਾਹਕਾਂ ਲਈ ਸੁਰੱਖਿਆ ਨਤੀਜਿਆਂ ਦੀ ਜ਼ਿੰਮੇਵਾਰੀ ਲੈਣ ਨੂੰ
- ਬੁਨਿਆਦੀ ਤਬਦੀਲੀ ਲਿਆਉਣ ਵਾਲੀ ਪਾਰਦਰਸ਼ਤਾ ਅਤੇ ਜਵਾਬਦੇਹੀ ਨੂੰ ਗਲੇ ਲਗਾਉਣ ਨੂੰ
- ਸੰਸਥਾਤਮਕ ਢਾਂਚੇ ਅਤੇ ਲੀਡਰਸ਼ਿਪ ਨਿਰਮਾਣ ਨੂੰ, ਇਸ ਲਈ ਡਿਜ਼ਾਈਨ ਦੁਆਰਾ ਸੁਰੱਖਿਅਤ ਬਣਾਉਣਾ ਇੱਕ ਪ੍ਰਮੁੱਖ ਵਪਾਰਕ ਪਹਿਲ ਹੁੰਦਾ ਹੈ।

'ਡਿਜ਼ਾਇਨ ਦੁਆਰਾ ਸੁਰੱਖਿਅਤ' ਸਿਧਾਂਤਾਂ ਦੀ ਪਾਲਣਾ ਕਰਨ ਲਈ ਪ੍ਰਣਾਲੀ ਦੇ ਜੀਵਨ ਚੱਕਰ ਦੌਰਾਨ ਕਾਫ਼ੀ ਜ਼ਿਆਦਾ ਸਰੋਤਾਂ ਦੀ ਲੋੜ ਹੁੰਦੀ ਹੈ। ਇਸਦਾ ਮਤਲਬ ਹੈ ਕਿ ਡਿਵੈਲਪਰਾਂ ਨੂੰ ਪ੍ਰਣਾਲੀ ਡਿਜ਼ਾਈਨ ਦੀ ਹਰੇਕ ਪਰਤ 'ਤੇ, ਅਤੇ ਵਿਕਾਸ ਜੀਵਨ ਚੱਕਰ ਦੇ ਸਾਰੇ ਪੜਾਵਾਂ 'ਤੇ ਗਾਹਕਾਂ ਦੀ ਸੁਰੱਖਿਆ ਨੂੰ ਪਹਿਲ ਦੇਣ ਵਾਲੀਆਂ **ਵਿਸ਼ੇਸ਼ਤਾਵਾਂ**, **ਵਿਧੀਆਂ**, ਅਤੇ **ਲਾਗੂ ਕਰਨ** ਵਾਲੇ ਟੂਲਾਂ ਵਿੱਚ ਲਾਜ਼ਮੀ ਨਿਵੇਸ਼ ਕਰਨਾ ਚਾਹੀਦਾ ਹੈ। ਅਜਿਹਾ ਕਰਨਾ ਬਾਅਦ ਵਿੱਚ ਮਹਿੰਗੇ ਪੈਣ ਵਾਲੇ ਪੁਨਰ-ਡਿਜ਼ਾਈਨ ਨੂੰ ਰੋਕੇਗਾ, ਨਾਲ ਹੀ ਨਾਲ ਨੇੜਲੇ ਸਮੇਂ ਵਿੱਚ ਗਾਹਕਾਂ ਅਤੇ ਉਹਨਾਂ ਦੇ ਡੇਟਾ ਨੂੰ ਸੁਰੱਖਿਅਤ ਕਰੇਗਾ।

AI ਸੁਰੱਖਿਆ ਵੱਖਰੀ ਕਿਉਂ ਹੈ?

ਇਸ ਦਸਤਾਵੇਜ਼ ਵਿੱਚ ਅਸੀਂ ਮਸ਼ੀਨ ਲਰਨਿੰਗ (ML) ਐਪਲੀਕੇਸ਼ਨਾਂ ਦਾ ਹਵਾਲਾ ਦੇਣ ਲਈ ਵਿਸ਼ੇਸ਼ ਤੌਰ 'ਤੇ 'AI' ਸ਼ਬਦ ਦੀ ਵਰਤੋਂ ਕਰਦੇ ਹਾਂ[3]। ML ਦੀਆਂ ਸਾਰੀਆਂ ਕਿਸਮਾਂ ਰੁੰਜਾਇਸ਼ ਦੇ ਦਾਇਰੇ ਵਿੱਚ ਹਨ। ਅਸੀਂ ML ਐਪਲੀਕੇਸ਼ਨਾਂ ਨੂੰ ਅਜਿਹੀਆਂ ਐਪਲੀਕੇਸ਼ਨਾਂ ਵਜੋਂ ਪਰਿਭਾਸ਼ਿਤ ਕਰਦੇ ਹਾਂ ਜੋ:

- ਅਜਿਹੇ ਸਾਫ਼ਟਵੇਅਰ ਭਾਗਾਂ (ਮਾਡਲ) ਨੂੰ ਸ਼ਾਮਲ ਕਰਦੇ ਹਨ ਜੋ ਕੰਪਿਊਟਰਾਂ ਨੂੰ ਕਿਸੇ ਮਨੁੱਖ ਦੁਆਰਾ ਸਪੱਸ਼ਟ ਤੌਰ 'ਤੇ ਪ੍ਰੋਗਰਾਮ ਕੀਤੇ ਜਾਣ ਵਾਲੇ ਨਿਯਮਾਂ ਤੋਂ ਬਗ਼ੈਰ ਡੇਟਾ ਵਿੱਚ ਪੈਟਰਨਾਂ ਨੂੰ ਪਛਾਣਨ ਅਤੇ ਸੰਦਰਭ ਲਿਆਉਣ ਦੀ ਆਗਿਆ ਦਿੰਦੇ ਹਨ।
- ਅੰਕੜਿਆਂ ਦੇ ਤਰਕ ਦੇ ਆਧਾਰ 'ਤੇ ਪੂਰਵ-ਅਨੁਮਾਨ ਲਗਾਉਂਦੇ, ਸਿਫ਼ਾਰਸ਼ਾਂ ਜਾਂ ਫ਼ੈਸਲੇ ਕਰਦੇ ਹਨ

ਮੌਜੂਦਾ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਖ਼ਤਰਿਆਂ ਦੇ ਨਾਲ-ਨਾਲ, AI ਪ੍ਰਣਾਲੀਆਂ ਨਵੀਆਂ ਕਿਸਮਾਂ ਦੀਆਂ ਕਮਜ਼ੋਰੀਆਂ ਦੇ ਅਧੀਨ ਹਨ। 'ਵਿਰੋਧਮਈ ਮਸ਼ੀਨ ਲਰਨਿੰਗ' (AML) ਸ਼ਬਦ ਨੂੰ, ਹਾਰਡਵੇਅਰ, ਸਾਫ਼ਟਵੇਅਰ, ਵਰਕਫਲੋਅ ਅਤੇ ਸਪਲਾਈ ਚੇਨਾਂ ਸਮੇਤ ML ਭਾਗਾਂ ਵਿੱਚ ਬੁਨਿਆਦੀ ਕਮਜ਼ੋਰੀਆਂ ਦੇ ਸ਼ੋਸ਼ਣ ਦਾ ਵਰਣਨ ਕਰਨ ਲਈ ਵਰਤਿਆ ਜਾਂਦਾ ਹੈ। AML ਹਮਲਾਵਰਾਂ ਨੂੰ ML ਪ੍ਰਣਾਲੀਆਂ ਵਿੱਚ ਅਣਇੱਛਤ ਵਿਵਹਾਰ ਪੈਦਾ ਕਰਨ ਦੇ ਯੋਗ ਬਣਾਉਂਦਾ ਹੈ ਜਿਸ ਵਿੱਚ ਸ਼ਾਮਲ ਹੋ ਸਕਦੇ ਹਨ:

- ਮਾਡਲ ਦੇ ਵਰਗੀਕਰਨ ਜਾਂ ਪਿਛੇ ਹੱਟਣ ਦੀ ਕਾਰਗੁਜ਼ਾਰੀ ਨੂੰ ਪ੍ਰਭਾਵਿਤ ਕਰਨਾ
- ਉਪਭੋਗਤਾਵਾਂ ਨੂੰ ਅਣਅਧਿਕਾਰਤ ਕਾਰਵਾਈਆਂ ਕਰਨ ਦੀ ਆਗਿਆ ਦੇਣਾ
- ਸੰਵੇਦਨਸ਼ੀਲ ਮਾਡਲ ਦੀ ਜਾਣਕਾਰੀ ਦਾ ਨਿਚੋੜ ਕੱਢਣਾ

ਇਹਨਾਂ ਪ੍ਰਭਾਵਾਂ ਨੂੰ ਪ੍ਰਾਪਤ ਕਰਨ ਦੇ ਬਹੁਤ ਸਾਰੇ ਤਰੀਕੇ ਹਨ, ਜਿਵੇਂ ਕਿ ਵੱਡੇ ਭਾਸ਼ਾ ਮਾਡਲ (LLM) ਡੋਮੇਨ ਵਿੱਚ ਝਟਪਟ ਇੰਜੈਕਸ਼ਨ ਹਮਲੇ, ਜਾਂ ਜਾਣਬੁੱਝ ਕੇ ਸਿਖਲਾਈ ਡੇਟਾ ਜਾਂ ਉਪਭੋਗਤਾ ਫੀਡਬੈਕ (ਜਿਸ ਨੂੰ 'ਡੇਟਾ ਪੋਇਜ਼ਨਿੰਗ (ਡੇਟਾ ਜ਼ਹਿਰੀਲਾ ਕਰਨ)' ਵਜੋਂ ਜਾਣਿਆ ਜਾਂਦਾ ਹੈ) ਨੂੰ ਖ਼ਰਾਬ ਕਰਨਾ।

ਇਹ ਦਸਤਾਵੇਜ਼ ਨੂੰ ਕਿਸਨੂੰ ਪੜ੍ਹਨਾ ਚਾਹੀਦਾ ਹੈ?

ਇਸ ਦਸਤਾਵੇਜ਼ ਮੁੱਖ ਤੌਰ 'ਤੇ AI ਪ੍ਰਣਾਲੀਆਂ ਦੇ ਪ੍ਰਦਾਤਾਵਾਂ ਲਈ ਉਦੇਸ਼ਿਤ ਹੈ, ਭਾਵੇਂ ਉਹ ਕਿਸੇ ਸੰਸਥਾ ਦੁਆਰਾ ਆਯੋਜਿਤ ਕੀਤੇ ਮਾਡਲਾਂ 'ਤੇ ਆਧਾਰਿਤ ਹੋਵੇ ਜਾਂ ਬਾਹਰੀ ਐਪਲੀਕੇਸ਼ਨ ਪ੍ਰੋਗਰਾਮਿੰਗ ਇੰਟਰਫੇਸ (APIs) ਦੀ ਵਰਤੋਂ ਕਰ ਰਿਹਾ ਹੋਵੇ। ਹਾਲਾਂਕਿ, ਅਸੀਂ **ਸਾਰੇ** ਹਿੱਤਧਾਰਕਾਂ (ਡੇਟਾ ਵਿਗਿਆਨੀਆਂ, ਡਿਵੈਲਪਰਾਂ, ਮੈਨੇਜਰਾਂ, ਫ਼ੈਸਲਾ ਲੈਣ ਵਾਲਿਆਂ ਅਤੇ ਜ਼ੋਖਮ ਦੇ ਮਾਲਕਾਂ ਸਮੇਤ) ਨੂੰ ਆਪਣੀਆਂ ਮਸ਼ੀਨ ਲਰਨਿੰਗ AI ਪ੍ਰਣਾਲੀਆਂ ਦੇ **ਡਿਜ਼ਾਈਨ**, **ਤੈਨਾਤੀ** ਅਤੇ **ਸੰਚਾਲਨ** ਬਾਰੇ ਜਾਣਕਾਰੀ ਭਰਪੂਰ ਫ਼ੈਸਲੇ ਲੈਣ ਵਿੱਚ ਮੱਦਦ ਕਰਨ ਲਈ ਇਹਨਾਂ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ਾਂ ਨੂੰ ਪੜ੍ਹਨ ਲਈ ਬੇਨਤੀ ਕਰਦੇ ਹਾਂ।

ਭਾਵੇਂ ਕਿ, ਸਾਰੇ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ ਸਾਰੀਆਂ ਸੰਸਥਾਵਾਂ 'ਤੇ ਸਿੱਧੇ ਤੌਰ 'ਤੇ ਲਾਗੂ ਨਹੀਂ ਹੋਣਗੇ। AI ਪ੍ਰਣਾਲੀ ਨੂੰ ਨਿਸ਼ਾਨਾ ਬਣਾਉਣ ਵਾਲੇ ਵਿਰੋਧਤਾ ਦੇ ਆਧਾਰ 'ਤੇ ਸੂਝਤਾ ਦਾ ਪੱਧਰ ਅਤੇ ਹਮਲੇ ਦੇ ਢੰਗ ਵੱਖੋ-ਵੱਖਰੇ ਹੋਣਗੇ, ਇਸ ਲਈ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ਾਂ ਨੂੰ ਤੁਹਾਡੀ ਸੰਸਥਾ ਦੀ ਵਰਤੋਂ ਦੇ ਮਾਮਲਿਆਂ ਅਤੇ ਖ਼ਤਰੇ ਦੇ ਪ੍ਰੋਫ਼ਾਈਲ ਦੇ ਨਾਲ-ਨਾਲ ਵਿਚਾਰਿਆ ਜਾਣਾ ਚਾਹੀਦਾ ਹੈ।

ਸੁਰੱਖਿਅਤ AI ਨੂੰ ਵਿਕਸਿਤ ਕਰਨ ਲਈ ਕੌਣ ਜ਼ਿੰਮੇਵਾਰ ਹੈ?

ਆਧੁਨਿਕ AI ਸਪਲਾਈ ਚੇਨਾਂ ਵਿੱਚ ਅਕਸਰ ਬਹੁਤ ਸਾਰੇ ਅਦਾਕਾਰ ਕਾਰਕ ਹੁੰਦੇ ਹਨ। ਇੱਕ ਆਸਾਨ ਪਹੁੰਚ ਦੋ ਇਕਾਈਆਂ ਨੂੰ ਮੰਨਦੀ ਹੈ:

- 'ਪ੍ਰਦਾਤਾ' ਜੋ ਡੇਟਾ ਦੀ ਵਿਵਸਥਾ ਕਰਨ, ਐਲਗੋਰਿਦਮਿਕ ਵਿਕਾਸ, ਡਿਜ਼ਾਈਨ, ਤੈਨਾਤੀ ਅਤੇ ਰੱਖ-ਰਖਾਅ ਲਈ ਜ਼ਿੰਮੇਵਾਰ ਹੁੰਦਾ ਹੈ
- 'ਉਪਭੋਗਤਾ', ਜੋ ਜਾਣਕਾਰੀ ਪ੍ਰਦਾਨ ਕਰਦਾ ਹੈ ਅਤੇ ਨਤੀਜਾ ਪ੍ਰਾਪਤ ਕਰਦਾ ਹੈ

ਹਾਲਾਂਕਿ ਇਹ ਪ੍ਰਦਾਤਾ-ਉਪਭੋਗਤਾ ਪਹੁੰਚ ਬਹੁਤ ਸਾਰੀਆਂ ਐਪਲੀਕੇਸ਼ਨਾਂ ਵਿੱਚ ਵਰਤੀ ਜਾਂਦੀ ਹੈ, ਇਹ ਤੇਜ਼ੀ ਨਾਲ ਘੱਟਦੀ ਜਾ ਰਹੀ ਹੈ[4], ਕਿਉਂਕਿ ਪ੍ਰਦਾਤਾ ਆਪਣੀ ਪ੍ਰਣਾਲੀ ਵਿੱਚ ਤੀਜੀਆਂ ਧਿਰਾਂ ਦੁਆਰਾ ਪ੍ਰਦਾਨ ਕੀਤੇ ਗਏ ਸਾਫ਼ਟਵੇਅਰ, ਡੇਟਾ, ਮਾਡਲਾਂ ਅਤੇ/ਜਾਂ ਰਿਮੋਟ ਸੇਵਾਵਾਂ ਨੂੰ ਸ਼ਾਮਲ ਕਰਨ ਦੀ ਕੋਸ਼ਿਸ਼ ਕਰ ਸਕਦੇ ਹਨ। ਇਹ ਗੁੰਝਲਦਾਰ ਸਪਲਾਈ ਚੇਨਾਂ ਅੰਤਮ ਉਪਭੋਗਤਾਵਾਂ ਲਈ ਇਹ ਸਮਝਣਾ ਮੁਸ਼ਕਲ ਬਣਾਉਂਦੀਆਂ ਹਨ ਕਿ ਸੁਰੱਖਿਅਤ AI ਦੀ ਜ਼ਿੰਮੇਵਾਰੀ ਕਿਸ ਦੇ ਸਿਰ ਹੈ।

ਉਪਭੋਗਤਾ (ਚਾਹੇ 'ਆਖਰੀ ਉਪਭੋਗਤਾ', ਜਾਂ ਇੱਕ ਬਾਹਰੀ AI ਹਿੱਸੇ ਨੂੰ ਸ਼ਾਮਲ ਕਰਨ ਵਾਲੇ ਪ੍ਰਦਾਤਾਵਾਂ [5]) ਕੋਲ ਆਮ ਤੌਰ 'ਤੇ ਉਹਨਾਂ ਦੁਆਰਾ ਵਰਤੀਆਂ ਜਾ ਰਹੀਆਂ ਪ੍ਰਣਾਲੀਆਂ ਨਾਲ ਜੁੜੇ ਜ਼ੋਖਮਾਂ ਨੂੰ ਪੂਰੀ ਤਰ੍ਹਾਂ ਸਮਝਣ, ਮੁਲਾਂਕਣ ਕਰਨ ਜਾਂ ਹੱਲ ਕਰਨ ਲਈ ਲੋੜੀਂਦੀ ਦਿੱਖ ਅਤੇ/ਜਾਂ ਮੁਹਾਰਤ ਨਹੀਂ ਹੁੰਦੀ ਹੈ। ਜਿਵੇਂ ਕਿ, 'ਡਿਜ਼ਾਇਨ ਦੁਆਰਾ ਸੁਰੱਖਿਅਤ' ਸਿਧਾਂਤਾਂ ਦੇ ਅਨੁਸਾਰ, **AI ਤੱਤਾਂ ਦੇ ਪ੍ਰਦਾਤਾਵਾਂ ਨੂੰ ਸਪਲਾਈ ਲੜੀ ਤੋਂ ਅੱਗੇ ਉਪਭੋਗਤਾਵਾਂ ਦੇ ਸੁਰੱਖਿਆ ਨਤੀਜਿਆਂ ਦੀ ਜ਼ਿੰਮੇਵਾਰੀ ਲੈਣੀ ਚਾਹੀਦੀ ਹੈ।**

ਪ੍ਰਦਾਤਾਵਾਂ ਨੂੰ ਉਹਨਾਂ ਦੇ ਮਾਡਲਾਂ, ਪਾਈਪਲਾਈਨਾਂ ਅਤੇ/ਜਾਂ ਪ੍ਰਣਾਲੀਆਂ ਦੇ ਅੰਦਰ ਜਿੱਥੇ ਸੰਭਵ ਹੋਵੇ ਸੁਰੱਖਿਆ ਨਿਯੰਤਰਣ ਅਤੇ ਕਮੀਆਂ ਤੋਂ ਬਚਾਅ ਨੂੰ ਲਾਗੂ ਕਰਨਾ ਚਾਹੀਦਾ ਹੈ, ਅਤੇ ਜਿੱਥੇ ਸੈਟਿੰਗਾਂ ਦੀ ਵਰਤੋਂ ਕੀਤੀ ਜਾਂਦੀ ਹੈ, ਸਭ ਤੋਂ ਸੁਰੱਖਿਅਤ ਵਿਕਲਪ ਨੂੰ ਡਿਫੌਲਟ ਵਜੋਂ ਲਾਗੂ ਕਰਨਾ ਚਾਹੀਦਾ ਹੈ। ਜਿੱਥੇ ਜ਼ੋਖਮਾਂ ਨੂੰ ਘੱਟ ਨਹੀਂ ਕੀਤਾ ਜਾ ਸਕਦਾ, ਪ੍ਰਦਾਤਾ ਨੂੰ ਇਹਨਾਂ ਗੱਲਾਂ ਲਈ ਜ਼ਿੰਮੇਵਾਰ ਹੋਣਾ ਚਾਹੀਦਾ ਹੈ:

- ਉਪਭੋਗਤਾਵਾਂ ਨੂੰ ਉਨ੍ਹਾਂ ਸਪਲਾਈ ਲੜੀ ਵਿਚਲੇ ਜ਼ੋਖਮਾਂ ਬਾਰੇ ਹੋਰ ਜਾਣਕਾਰੀ ਦੇਣਾ ਜੋ ਉਹ ਆਪ ਅਤੇ (ਜੇ ਲਾਗੂ ਹੋਵੇ ਤਾਂ) ਉਹਨਾਂ ਦੇ ਆਪਣੇ ਉਪਭੋਗਤਾ ਸਵੀਕਾਰ ਕਰ ਰਹੇ ਹਨ
- ਉਹਨਾਂ ਨੂੰ ਇਹ ਸਲਾਹ ਦੇਣਾ ਕਿ ਉਸ ਭਾਗ ਨੂੰ ਸੁਰੱਖਿਅਤ ਢੰਗ ਨਾਲ ਕਿਵੇਂ ਵਰਤਣਾ ਹੈ

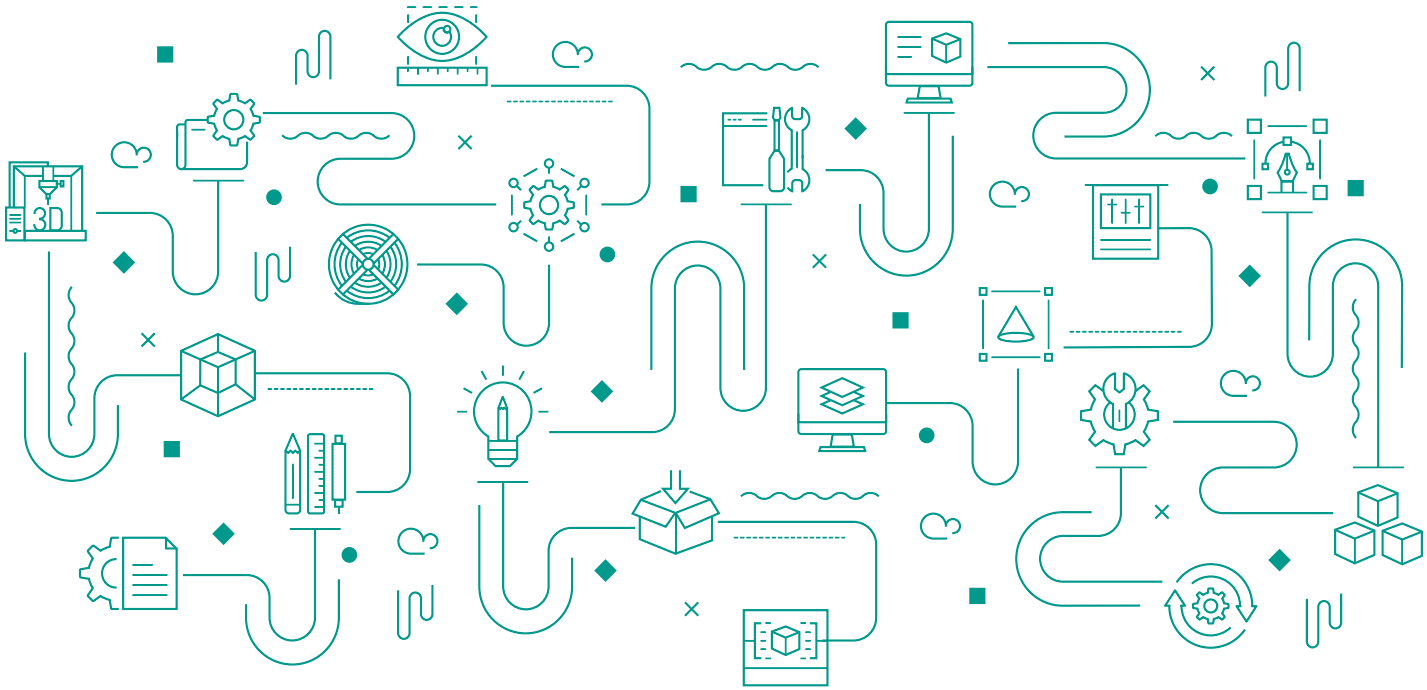
ਜਿੱਥੇ ਪ੍ਰਣਾਲੀ ਦੀ ਅਣਅਧਿਕਾਰਤ ਪਹੁੰਚ ਸਪੱਸ਼ਟ ਜਾਂ ਵਿਆਪਕ ਭੌਤਿਕ ਨੁਕਸਾਨ ਜਾਂ ਸਾਖ਼ ਨੂੰ ਨੁਕਸਾਨ ਪਹੁੰਚਾ ਸਕਦੀ ਹੈ, ਕਾਰੋਬਾਰੀ ਗਤੀਵਿਧੀਆਂ ਕਾਫ਼ੀ ਜ਼ਿਆਦਾ ਮਹੱਤਵਪੂਰਨ ਨੁਕਸਾਨ, ਸੰਵੇਦਨਸ਼ੀਲ ਜਾਂ ਗੁਪਤ ਜਾਣਕਾਰੀ ਦੇ ਲੀਕ ਹੋਣ ਅਤੇ/ਜਾਂ ਕਾਨੂੰਨੀ ਉਲਝਣਾਂ ਦਾ ਕਾਰਨ ਬਣ ਸਕਦੀਆਂ ਹਨ, ਉੱਥੇ AI ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਜ਼ੋਖਮਾਂ ਨੂੰ **ਅਤਿ ਖ਼ਤਰਨਾਕ** ਮੰਨਿਆ ਜਾਣਾ ਚਾਹੀਦਾ ਹੈ।

ਸੁਰੱਖਿਅਤ AI ਪ੍ਰਣਾਲੀ ਬਣਾਉਣ ਲਈ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼

ਇਹਨਾਂ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ਾਂ ਨੂੰ AI ਪ੍ਰਣਾਲੀ ਦੇ ਵਿਕਾਸ ਦੇ ਜੀਵਨ ਚੱਕਰ ਵਿੱਚ ਚਾਰ ਮੁੱਖ ਖੇਤਰਾਂ ਵਿੱਚ ਵੰਡਿਆ ਗਿਆ ਹੈ: **ਸੁਰੱਖਿਅਤ ਡਿਜ਼ਾਈਨ**, **ਸੁਰੱਖਿਅਤ ਵਿਕਾਸ**, **ਸੁਰੱਖਿਅਤ ਤੈਨਾਤੀ**, ਅਤੇ **ਸੁਰੱਖਿਅਤ ਸੰਚਾਲਨ ਅਤੇ ਰੱਖ-ਰਖਾਅ**। ਹਰੇਕ ਖੇਤਰ ਲਈ, ਅਸੀਂ ਵਿਚਾਰਨ ਯੋਗ ਨੁਕਤਿਆਂ ਅਤੇ ਕਮੀਆਂ ਤੋਂ ਬਚਾਅ ਦਾ ਸੁਝਾਅ ਦਿੰਦੇ ਹਾਂ ਜੋ ਸੰਸਥਾਤਮਕ AI ਪ੍ਰਣਾਲੀ ਦੀ ਵਿਕਾਸ ਪ੍ਰਕਿਰਿਆ ਦੇ ਸਮੁੱਚੇ ਜ਼ੋਖਮ ਨੂੰ ਘਟਾਉਣ ਵਿੱਚ ਮੱਦਦ ਕਰਨਗੀਆਂ।

ਇਸ ਦਸਤਾਵੇਜ਼ ਵਿੱਚ ਨਿਰਧਾਰਤ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ਾਂ ਵਿੱਚ ਪਰਿਭਾਸ਼ਿਤ ਕੀਤੇ ਗਏ ਸਾਫ਼ਟਵੇਅਰ ਵਿਕਸਤ ਕਰਨ ਦੇ ਜੀਵਨ ਚੱਕਰ ਦੇ ਕੰਮ ਕਰਨ ਦੇ ਤਰੀਕਿਆਂ ਨਾਲ ਇਨ੍ਹਾਂ ਨਾਲ ਨੇੜਿਓਂ ਜੁੜੇ ਹੋਏ ਹੁੰਦੇ ਹਨ:

- NCSC ਦਾ ਸੁਰੱਖਿਅਤ ਵਿਕਾਸ ਅਤੇ ਤੈਨਾਤੀ ਮਾਰਗਦਰਸ਼ਨ
- ਨੈਸ਼ਨਲ ਇੰਸਟੀਚਿਊਟ ਆਫ਼ ਸਟੈਂਡਰਡਜ਼ ਐਂਡ ਟੈਕਨਾਲੋਜੀ (NIST) ਸੁਰੱਖਿਅਤ ਸਾਫ਼ਟਵੇਅਰ ਵਿਕਸਤ ਕਰਨ ਲਈ ਫਰੇਮਵਰਕ (SSDF)[6]



1. ਸੁਰੱਖਿਅਤ ਡਿਜ਼ਾਈਨ

ਇਸ ਭਾਗ ਵਿੱਚ ਉਹ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ ਸ਼ਾਮਲ ਹਨ ਜੋ AI ਪ੍ਰਣਾਲੀ ਦੇ ਵਿਕਾਸ ਜੀਵਨ ਚੱਕਰ ਦੇ **ਡਿਜ਼ਾਈਨ** ਪੜਾਅ 'ਤੇ ਲਾਗੂ ਹੁੰਦੇ ਹਨ। ਇਸ ਵਿੱਚ ਪ੍ਰਣਾਲੀ ਅਤੇ ਮਾਡਲ ਡਿਜ਼ਾਈਨ 'ਤੇ ਵਿਚਾਰ ਕਰਨ ਲਈ ਜ਼ੋਖਮਾਂ ਅਤੇ ਖ਼ਤਰੇ ਦੇ ਮਾਡਲਿੰਗ ਦੇ ਨਾਲ-ਨਾਲ ਖ਼ਾਸ ਵਿਸ਼ਿਆਂ ਅਤੇ ਟ੍ਰੇਡ-ਆਫ (ਕੁੱਝ ਚੰਗਾ ਕਰਨ ਲਈ ਕੁੱਝ ਬੁਰਾ ਸਵੀਕਾਰ ਕਰਨ) ਨੂੰ ਸਮਝਣਾ ਵੀ ਸ਼ਾਮਲ ਹੈ।

ਖ਼ਤਰਿਆਂ ਅਤੇ ਜ਼ੋਖਮਾਂ ਬਾਰੇ ਸਟਾਫ਼ ਦੀ ਜਾਗਰੂਕਤਾ ਵਧਾਓ

ਪ੍ਰਣਾਲੀ ਦੇ ਮਾਲਕ ਅਤੇ ਸੀਨੀਅਰ ਆਗੂ, ਸੁਰੱਖਿਅਤ AI ਅਤੇ ਉਹਨਾਂ ਦੀਆਂ ਕਮੀਆਂ ਲਈ ਖ਼ਤਰਿਆਂ ਨੂੰ ਸਮਝਦੇ ਹਨ। ਤੁਹਾਡੇ ਡੇਟਾ ਵਿਗਿਆਨੀ ਅਤੇ ਡਿਵੈਲਪਰ ਸੰਬੰਧਿਤ ਸੁਰੱਖਿਆ ਖ਼ਤਰਿਆਂ ਅਤੇ ਅਸਫ਼ਲਤਾ ਮੋਡਾਂ ਬਾਰੇ ਜਾਗਰੂਕਤਾ ਬਰਕਰਾਰ ਰੱਖਦੇ ਹਨ ਅਤੇ ਜ਼ੋਖਮ ਦੇ ਮਾਲਕਾਂ ਨੂੰ ਜਾਣਕਾਰੀ ਭਰਪੂਰ ਫ਼ੈਸਲੇ ਲੈਣ ਵਿੱਚ ਮੱਦਦ ਕਰਦੇ ਹਨ। ਤੁਸੀਂ ਉਪਭੋਗਤਾਵਾਂ ਨੂੰ AI ਪ੍ਰਣਾਲੀਆਂ ਵੱਲੋਂ ਸਾਹਮਣਾ ਕੀਤੇ ਜਾਂਦੇ ਵਿਲੱਖਣ ਸੁਰੱਖਿਆ ਜ਼ੋਖਮਾਂ ਬਾਰੇ ਮਾਰਗਦਰਸ਼ਨ ਪ੍ਰਦਾਨ ਕਰਦੇ ਹੋ (ਉਦਾਹਰਨ ਲਈ, ਮਿਆਰੀ InfoSec ਸਿਖਲਾਈ ਦੇ ਹਿੱਸੇ ਵਜੋਂ) ਅਤੇ ਡਿਵੈਲਪਰਾਂ ਨੂੰ ਸੁਰੱਖਿਅਤ ਕੋਡਿੰਗ ਤਕਨੀਕਾਂ ਅਤੇ ਸੁਰੱਖਿਅਤ ਅਤੇ ਜ਼ਿੰਮੇਵਾਰ AI ਕੰਮ ਕਰਨ ਦੇ ਤਰੀਕਿਆਂ ਵਿੱਚ ਸਿਖਲਾਈ ਦਿੰਦੇ ਹੋ।

ਆਪਣੀ ਪ੍ਰਣਾਲੀ ਲਈ ਖ਼ਤਰਿਆਂ ਦਾ ਮਾਡਲ ਬਣਾਓ

ਤੁਹਾਡੀ ਜ਼ੋਖਮ ਪ੍ਰਬੰਧਨ ਪ੍ਰਕਿਰਿਆ ਦੇ ਹਿੱਸੇ ਵਜੋਂ, ਤੁਸੀਂ ਆਪਣੀ ਪ੍ਰਣਾਲੀ ਲਈ ਖ਼ਤਰਿਆਂ ਦਾ ਮੁਲਾਂਕਣ ਕਰਨ ਲਈ ਇੱਕ ਸੰਪੂਰਨ ਪ੍ਰਕਿਰਿਆ ਲਾਗੂ ਕਰਦੇ ਹੋ, ਜਿਸ ਵਿੱਚ ਉਸ ਪ੍ਰਣਾਲੀ, ਉਪਭੋਗਤਾਵਾਂ, ਸੰਸਥਾਵਾਂ ਅਤੇ ਵਿਆਪਕ ਸਮਾਜ ਉੱਤੇ ਪੈਣ ਵਾਲੇ ਸੰਭਾਵੀ ਪ੍ਰਭਾਵਾਂ ਨੂੰ ਸਮਝਣਾ ਸ਼ਾਮਲ ਹੁੰਦਾ ਹੈ ਜੇਕਰ AI ਹਿੱਸੇ ਤੱਕ ਅਣਅਧਿਕਾਰਤ ਪਹੁੰਚ ਕੀਤੀ ਜਾਂਦੀ ਹੈ ਜਾਂ ਇਹ ਅਣਕਿਆਸਾ ਵਿਵਹਾਰ ਕਰਦਾ ਹੈ[7]। ਇਸ ਪ੍ਰਕਿਰਿਆ ਵਿੱਚ AI-ਵਿਸ਼ੇਸ਼ ਖ਼ਤਰਿਆਂ ਦੇ ਪ੍ਰਭਾਵ ਦਾ ਮੁਲਾਂਕਣ ਕਰਨਾ[8] ਅਤੇ ਤੁਹਾਡੇ ਫ਼ੈਸਲੇ ਲੈਣ ਦਾ ਦਸਤਾਵੇਜ਼ੀਕਰਨ ਕਰਨਾ ਸ਼ਾਮਲ ਹੈ।

ਤੁਸੀਂ ਸਵੀਕਾਰਦੇ ਹੋ ਕਿ ਤੁਹਾਡੀ ਪ੍ਰਣਾਲੀ ਵਿੱਚ ਵਰਤੇ ਗਏ ਡੇਟੇ ਦੀ ਸੰਵੇਦਨਸ਼ੀਲਤਾ ਅਤੇ ਕਿਸਮਾਂ ਇੱਕ ਹਮਲਾਵਰ ਦੇ ਨਿਸ਼ਾਨੇ ਵਜੋਂ ਇਸਦੇ ਮੁੱਲ ਨੂੰ ਪ੍ਰਭਾਵਿਤ ਕਰ ਸਕਦੀਆਂ ਹਨ। ਤੁਹਾਡੇ ਮੁਲਾਂਕਣ ਨੂੰ ਇਸ ਗੱਲ 'ਤੇ ਵਿਚਾਰ ਕਰਨਾ ਚਾਹੀਦਾ ਹੈ ਕਿ ਕੁੱਝ ਖ਼ਤਰੇ ਵਧ ਸਕਦੇ ਹਨ ਕਿਉਂਕਿ AI ਪ੍ਰਣਾਲੀਆਂ ਨੂੰ ਉੱਚ ਮੁੱਲ ਵਾਲੇ ਟੀਚਿਆਂ ਵਜੋਂ ਦੇਖਿਆ ਜਾਂਦਾ ਹੈ, ਅਤੇ ਜਿਵੇਂ ਕਿ AI ਖੁਦ ਨਵੇਂ, ਸਵੈਚਾਲਿਤ ਹਮਲੇ ਵਾਲੇ ਵੈਕਟਰਾਂ ਨੂੰ ਚਾਲੂ ਕਰਦਾ ਹੈ।

ਸੁਰੱਖਿਆ ਦੇ ਨਾਲ-ਨਾਲ ਕਾਰਜਕੁਸ਼ਲਤਾ ਅਤੇ ਪ੍ਰਦਰਸ਼ਨ ਲਈ ਆਪਣੀਆਂ ਪ੍ਰਣਾਲੀਆਂ ਨੂੰ ਡਿਜ਼ਾਈਨ ਕਰੋ

ਤੁਹਾਨੂੰ ਯਕੀਨ ਹੈ ਕਿ ਹੱਥ ਵਿਚਲੇ ਕੰਮ ਨੂੰ AI ਦੀ ਵਰਤੋਂ ਕਰਕੇ ਸਭ ਤੋਂ ਉਚਿਤ ਢੰਗ ਨਾਲ ਹੱਲ ਕੀਤਾ ਗਿਆ ਹੈ। ਇਹ ਨਿਰਧਾਰਤ ਕਰਨ ਤੋਂ ਬਾਅਦ, ਤੁਸੀਂ ਆਪਣੇ AI-ਵਿਸ਼ੇਸ਼ ਡਿਜ਼ਾਈਨ ਵਿਕਲਪਾਂ ਦੀ ਉਚਿਤਤਾ ਦਾ ਮੁਲਾਂਕਣ ਕਰਦੇ ਹੋ। ਹੋਰ ਵਿਚਾਰਨਯੋਗ ਗੱਲਾਂ ਦੇ ਨਾਲ-ਨਾਲ ਤੁਸੀਂ ਕਾਰਜਸ਼ੀਲਤਾ, ਉਪਭੋਗਤਾ ਅਨੁਭਵ, ਤੈਨਾਤੀ ਮਾਹੌਲ, ਪ੍ਰਦਰਸ਼ਨ, ਕਾਰਗੁਜ਼ਾਰੀ, ਨਿਗਰਾਨੀ, ਨੈਤਿਕ ਅਤੇ ਕਾਨੂੰਨੀ ਲੋੜਾਂ ਦੇ ਨਾਲ-ਨਾਲ ਆਪਣੇ ਖ਼ਤਰੇ ਦੇ ਮਾਡਲ ਅਤੇ ਸੰਬੰਧਿਤ ਸੁਰੱਖਿਆ ਕਮੀਆਂ 'ਤੇ ਵਿਚਾਰ ਕਰਦੇ ਹੋ। ਉਦਾਹਰਨ ਲਈ:

- ਤੁਸੀਂ ਸਪਲਾਈ ਚੇਨ ਸੁਰੱਖਿਆ 'ਤੇ ਵਿਚਾਰ ਕਰਦੇ ਹੋ ਜਦੋਂ ਇਹ ਚੁਣਦੇ ਹੋ ਕਿ ਕੀ ਆਪ ਵਿਕਾਸ ਕਰਨਾ ਹੈ ਜਾਂ ਬਾਹਰੀ ਹਿੱਸਿਆਂ ਦੀ ਵਰਤੋਂ ਕਰਨੀ ਹੈ, ਉਦਾਹਰਨ ਲਈ:
    - ਇੱਕ ਨਵੇਂ ਮਾਡਲ ਨੂੰ ਸਿਖਲਾਈ ਦੇਣ ਲਈ ਤੁਹਾਡੀ ਚੋਣ, ਤੁਹਾਡੀਆਂ ਜ਼ਰੂਰਤਾਂ ਲਈ ਮੌਜੂਦਾ ਮਾਡਲ (ਫਾਈਨ-ਟਿਊਨਿੰਗ ਦੇ ਨਾਲ ਜਾਂ ਬਿਨਾਂ) ਦੀ ਵਰਤੋਂ ਕਰਨ ਜਾਂ ਕਿਸੇ ਬਾਹਰੀ API ਰਾਹੀਂ ਮਾਡਲ ਤੱਕ ਪਹੁੰਚ ਕਰਨ ਦੀ ਚੋਣ ਉਚਿਤ ਹੈ
    - ਕਿਸੇ ਬਾਹਰੀ ਮਾਡਲ ਪ੍ਰਦਾਤਾ ਦੇ ਨਾਲ ਕੰਮ ਕਰਨ ਦੀ ਤੁਹਾਡੀ ਚੋਣ ਵਿੱਚ ਉਸ ਪ੍ਰਦਾਤਾ ਦੇ ਆਪਣੇ ਸੁਰੱਖਿਆ ਪ੍ਰਬੰਧਾਂ ਦਾ ਢੁੱਕਵਾਂ ਮੁਲਾਂਕਣ ਕਰਨਾ ਸ਼ਾਮਲ ਹੁੰਦਾ ਹੈ
    - ਜੇਕਰ ਕਿਸੇ ਬਾਹਰੀ ਲਾਇਬ੍ਰੇਰੀ ਦੀ ਵਰਤੋਂ ਕਰ ਰਹੇ ਹੋ, ਤਾਂ ਤੁਸੀਂ ਢੁੱਕਵਾਂ ਮੁਲਾਂਕਣ ਕਰਦੇ ਹੋ (ਉਦਾਹਰਨ ਲਈ, ਇਹ ਯਕੀਨੀ ਬਣਾਉਣ ਲਈ ਕਿ ਲਾਇਬ੍ਰੇਰੀ ਵਿੱਚ ਉਹ ਨਿਯੰਤਰਣ ਹਨ ਜੋ ਪ੍ਰਣਾਲੀ ਨੂੰ ਕਿਸੇ ਧੱਕੇਸ਼ਾਹੀ ਵਾਲੇ ਕੋਡ ਦੀ ਪਾਲਣਾ ਲਈ ਤੁਰੰਤ ਆਪਣੇ-ਆਪ ਨੂੰ ਖ਼ਤਰੇ ਵਿੱਚ ਪਾਏ ਤੋਂ ਬਗ਼ੈਰ ਗ਼ੈਰ-ਭਰੋਸੇਯੋਗ ਮਾਡਲਾਂ ਨੂੰ ਲੋਡ ਕਰਨ ਤੋਂ ਰੋਕਦੇ ਹਨ[9])
    - ਤੁਸੀਂ ਤੀਜੀ-ਧਿਰ ਦੇ ਮਾਡਲਾਂ ਜਾਂ ਲੜੀਬੱਧ ਵਜ਼ਨਾਂ ਨੂੰ ਆਯਾਤ ਕਰਦੇ ਸਮੇਂ ਸਕੈਨਿੰਗ ਅਤੇ ਆਈਸੋਲੇਸ਼ਨ/ਸੈਂਡਬਾਕਸਿੰਗ ਨੂੰ ਲਾਗੂ ਕਰਦੇ ਹੋ, ਜਿਸ ਨੂੰ ਗ਼ੈਰ-ਭਰੋਸੇਯੋਗ ਤੀਜੀ-ਧਿਰ ਦੇ ਕੋਡ ਮੰਨਿਆ ਜਾਣਾ ਚਾਹੀਦਾ ਹੈ ਅਤੇ ਰਿਮੋਟ ਕੋਡ ਪਾਲਣਾ ਨੂੰ ਲਾਗੂ ਕਰ ਸਕਦਾ ਹੋਣਾ ਚਾਹੀਦਾ ਹੈ।

- ਜੇਕਰ ਕਿਸੇ ਬਾਹਰੀ API ਦੀ ਵਰਤੋਂ ਕਰ ਰਹੇ ਹੋ, ਤਾਂ ਤੁਸੀਂ ਉਸ ਡੇਟਾ 'ਤੇ ਉਚਿਤ ਨਿਯੰਤਰਣ ਲਾਗੂ ਕਰੋ ਜੋ ਤੁਹਾਡੀ ਸੰਸਥਾ ਦੇ ਨਿਯੰਤਰਣ ਤੋਂ ਬਾਹਰ ਵਾਲੀਆਂ ਸੇਵਾਵਾਂ ਨੂੰ ਭੇਜੇ ਜਾ ਸਕਦੇ ਹਨ, ਜਿਵੇਂ ਕਿ ਉਪਭੋਗਤਾਵਾਂ ਨੂੰ ਸੰਭਾਵੀ ਤੌਰ 'ਤੇ ਸੰਵੇਦਨਸ਼ੀਲ ਜਾਣਕਾਰੀ ਭੇਜਣ ਤੋਂ ਪਹਿਲਾਂ ਲੌਗ-ਇਨ ਕਰਨ ਅਤੇ ਤਸਦੀਕ ਕਰਨ ਦੀ ਲੋੜ ਹੁੰਦੀ ਹੈ।
- ਤੁਸੀਂ ਡੇਟਾ ਅਤੇ ਮਿਲੀ ਜਾਣਕਾਰੀ ਦੀ ਢੁੱਕਵੀਂ ਜਾਂਚ ਅਤੇ ਸੈਨੀਟਾਈਜ਼ੇਸ਼ਨ (ਸਟੋਰੇਜ ਡਿਵਾਈਸ ਤੋਂ ਡੇਟਾ ਨੂੰ ਜਾਣਬੁੱਝ ਕੇ, ਸਥਾਈ ਤੌਰ 'ਤੇ ਮਿਟਾਉਣਾ ਜਾਂ ਨਸ਼ਟ ਕਰਨਾ) ਕਰਦੇ ਹੋ; ਇਸ ਵਿੱਚ ਜਦੋਂ ਤੁਸੀਂ ਆਪਣੇ ਮਾਡਲ ਵਿੱਚ ਉਪਭੋਗਤਾ ਫੀਡਬੈਕ ਜਾਂ ਨਿਰੰਤਰ ਸਿਖਲਾਈ ਡੇਟਾ ਨੂੰ ਸ਼ਾਮਲ ਕਰਦੇ ਹੋ, ਅਤੇ ਇਹ ਪਛਾਣਦੇ ਹੋ ਕਿ ਸਿਖਲਾਈ ਡੈਟਾ ਪ੍ਰਣਾਲੀ ਦੇ ਵਿਵਹਾਰ ਨੂੰ ਪਰਿਭਾਸ਼ਿਤ ਕਰਦਾ ਹੈ, ਸ਼ਾਮਲ ਹੈ

- ਤੁਸੀਂ AI ਸਾਫਟਵੇਅਰ ਪ੍ਰਣਾਲੀ ਦੇ ਵਿਕਾਸ ਨੂੰ ਮੌਜੂਦਾ ਸੁਰੱਖਿਅਤ ਵਿਕਾਸ ਅਤੇ ਸੰਚਾਲਨ ਦੇ ਵਧੀਆ ਤਰੀਕਿਆਂ ਨਾਲ ਜੋੜਦੇ ਹੋ; AI ਪ੍ਰਣਾਲੀ ਦੇ ਸਾਰੇ ਤੱਤ ਕੋਡਿੰਗ ਕਰਨ ਦੇ ਤਰੀਕਿਆਂ ਅਤੇ ਭਾਸ਼ਾਵਾਂ ਦੀ ਵਰਤੋਂ ਕਰਦੇ ਹੋਏ ਢੁੱਕਵੇਂ ਮਾਹੌਲ ਵਿੱਚ ਲਿਖੇ ਗਏ ਹਨ ਜੋ ਕਿ ਜਿੱਥੇ ਵੀ ਸੰਭਵ ਹੋਵੇ, ਕਮਜ਼ੋਰੀਆਂ ਦੀਆਂ ਪਤਾ ਹੋਣ ਵਾਲੀਆਂ ਸ਼੍ਰੇਣੀਆਂ ਨੂੰ ਘਟਾਉਂਦੇ ਜਾਂ ਖ਼ਤਮ ਕਰਦੇ ਹਨ

- ਜੇਕਰ AI ਭਾਗਾਂ ਨੂੰ ਕਾਰਵਾਈਆਂ ਸ਼ੁਰੂ ਕਰਨ ਦੀ ਲੋੜ ਹੈ, ਉਦਾਹਰਨ ਲਈ ਫਾਈਲਾਂ ਨੂੰ ਸੋਧਣਾ ਜਾਂ ਬਾਹਰੀ ਪ੍ਰਣਾਲੀਆਂ ਲਈ ਆਉਟਪੁੱਟ ਨੂੰ ਨਿਰਦੇਸ਼ਤ ਕਰਨਾ, ਤੁਸੀਂ ਸੰਭਾਵਿਤ ਕਾਰਵਾਈਆਂ ਲਈ ਉਚਿਤ ਪਾਬੰਦੀਆਂ ਲਾਗੂ ਕਰੋ (ਜੇ ਲੋੜ ਹੋਵੇ ਤਾਂ ਇਸ ਵਿੱਚ ਬਾਹਰੀ AI ਅਤੇ ਗ਼ੈਰ-AI ਫ਼ੇਲ-ਸੁਰੱਖਿਅਤ ਢੰਗ ਸ਼ਾਮਲ ਹਨ)

- ਉਪਭੋਗਤਾਵਾਂ ਦੇ ਆਪਸੀ ਤਾਲਮੇਲ ਬਾਰੇ ਫ਼ੈਸਲਿਆਂ ਨੂੰ AI-ਵਿਸ਼ੇਸ਼ ਜ਼ੋਖਮਾਂ ਦੁਆਰਾ ਸੂਚਿਤ ਕੀਤਾ ਜਾਂਦਾ ਹੈ, ਉਦਾਹਰਨ ਲਈ:

  - ਤੁਹਾਡੀ ਪ੍ਰਣਾਲੀ ਸੰਭਾਵੀ ਹਮਲਾਵਰ ਨੂੰ ਵੇਰਵੇ ਦੇ ਬੇਲੋੜੇ ਪੱਧਰਾਂ ਦਾ ਖ਼ੁਲਾਸਾ ਕੀਤੇ ਬਗ਼ੈਰ ਉਪਭੋਗਤਾਵਾਂ ਨੂੰ ਉਪਯੋਗੀ ਆਉਟਪੁੱਟ ਪ੍ਰਦਾਨ ਕਰਦੀ ਹੈ
  - ਜੇਕਰ ਲੋੜ ਹੋਵੇ, ਤਾਂ ਤੁਹਾਡੀ ਪ੍ਰਣਾਲੀ ਮਾਡਲ ਆਉਟਪੁੱਟ ਦੇ ਆਲੇ-ਦੁਆਲੇ ਪ੍ਰਭਾਵਸ਼ਾਲੀ ਸੁਰੱਖਿਆ ਘੇਰਾ ਪ੍ਰਦਾਨ ਕਰਦੀ ਹੈ
  - ਜੇਕਰ ਬਾਹਰੀ ਗਾਹਕਾਂ ਜਾਂ ਸਹਿਯੋਗੀਆਂ ਨੂੰ API ਦੀ ਪੇਸ਼ਕਸ਼ ਕਰਦੇ ਹੋ, ਤਾਂ ਤੁਸੀਂ ਉਚਿਤ ਨਿਯੰਤਰਣ ਲਾਗੂ ਕਰੋ ਜੋ API ਦੁਆਰਾ AI ਪ੍ਰਣਾਲੀ 'ਤੇ ਹਮਲਿਆਂ ਨੂੰ ਘੱਟ ਕਰਦੇ ਹਨ।
  - ਤੁਸੀਂ ਡਿਫੌਲਟ ਰੂਪ ਵਿੱਚ ਪ੍ਰਣਾਲੀ ਵਿੱਚ ਸਭ ਤੋਂ ਸੁਰੱਖਿਅਤ ਸੈਟਿੰਗਾਂ ਨੂੰ ਜੋੜੋ
  - ਤੁਸੀਂ ਪ੍ਰਣਾਲੀ ਦੀ ਕਾਰਜਕੁਸ਼ਲਤਾ ਤੱਕ ਪਹੁੰਚ ਨੂੰ ਸੀਮਿਤ ਕਰਨ ਲਈ ਘੱਟੋ-ਘੱਟ ਅਧਿਕਾਰ ਹੋਣ ਦੇ ਸਿਧਾਂਤ ਲਾਗੂ ਕਰਦੇ ਹੋ
  - ਤੁਸੀਂ ਉਪਭੋਗਤਾਵਾਂ ਨੂੰ ਜ਼ੋਖਮ ਭਰੀਆਂ ਸਮਰੱਥਾਵਾਂ ਬਾਰੇ ਦੱਸਦੇ ਹੋ ਅਤੇ ਉਪਭੋਗਤਾਵਾਂ ਨੂੰ ਉਹਨਾਂ ਦੀ ਵਰਤੋਂ ਕਰਨ ਲਈ ਚੁਣਨ ਦੀ ਲੋੜ ਹੁੰਦੀ ਹੈ; ਤੁਸੀਂ ਮਨਾਹੀ ਵਰਤੋਂ ਵਾਲੇ ਮਾਮਲਿਆਂ ਬਾਰੇ ਦੱਸਦੇ ਹੋ, ਅਤੇ, ਜਿੱਥੇ ਸੰਭਵ ਹੋਵੇ, ਵਰਤੋਂ ਕਰਨ ਵਾਲਿਆਂ ਨੂੰ ਵਿਕਲਪਿਕ ਹੱਲਾਂ ਬਾਰੇ ਸੂਚਿਤ ਕਰਦੇ ਹੋ

ਆਪਣੇ AI ਮਾਡਲ ਦੀ ਚੋਣ ਕਰਦੇ ਸਮੇਂ ਸੁਰੱਖਿਆ ਲਾਭਾਂ ਅਤੇ ਟ੍ਰੇਡ-ਆਫਾਂ (ਚੰਗੇ-ਮਾੜੇ) 'ਤੇ ਵਿਚਾਰ ਕਰੋ

AI ਮਾਡਲ ਦੀ ਤੁਹਾਡੀ ਚੋਣ ਵਿੱਚ ਕਈ ਕਿਸਮ ਦੀਆਂ ਲੋੜਾਂ ਨੂੰ ਸੰਤੁਲਿਤ ਕਰਨਾ ਸ਼ਾਮਲ ਹੋਵੇਗਾ। ਇਸ ਵਿੱਚ ਮਾਡਲ ਆਰਕੀਟੈਕਚਰ, ਸੰਰਚਨਾ, ਸਿਖਲਾਈ ਡੇਟਾ, ਸਿਖਲਾਈ ਐਲਗੋਰਿਦਮ ਅਤੇ ਹਾਈਪਰਪੈਰਾਮੀਟਰਾਂ ਦੀ ਚੋਣ ਕਰਨਾ ਸ਼ਾਮਲ ਹੈ। ਤੁਹਾਡੇ ਫ਼ੈਸਲਿਆਂ ਨੂੰ ਤੁਹਾਡੇ ਖ਼ਤਰੇ ਦੇ ਮਾਡਲ ਦੁਆਰਾ ਸੂਚਿਤ ਕੀਤਾ ਜਾਂਦਾ ਹੈ, ਅਤੇ AI ਸੁਰੱਖਿਆ ਖੋਜ ਦੀ ਤਰੱਕੀ ਅਤੇ ਖ਼ਤਰੇ ਦੀ ਸਮਝ ਵਿਕਸਿਤ ਹੋਣ ਦੇ ਨਾਲ ਨਿਯਮਿਤ ਤੌਰ 'ਤੇ ਮੁੜ-ਮੁਲਾਂਕਣ ਕੀਤਾ ਜਾਂਦਾ ਹੈ।

AI ਮਾਡਲ ਦੀ ਚੋਣ ਕਰਦੇ ਸਮੇਂ, ਤੁਹਾਡੇ ਲਈ ਵਿਚਾਰਨਯੋਗ ਨੁਕਤਿਆਂ ਵਿੱਚ ਸੰਭਾਵਤ ਤੌਰ 'ਤੇ ਹੇਠ ਲਿਖੀਆਂ ਗੱਲਾਂ ਸ਼ਾਮਲ ਹੋਣਗੀਆਂ, ਪਰ ਇਹ ਇਹਨਾਂ ਤੱਕ ਸੀਮਿਤ ਨਹੀਂ ਹਨ:

- ਤੁਹਾਡੇ ਦੁਆਰਾ ਵਰਤੇ ਜਾ ਰਹੇ ਮਾਡਲ ਦੀ ਗੁੰਝਲਤਾ 'ਤੇ, ਜੋ ਕਿ, ਚੁਣਿਆ ਗਿਆ ਆਰਕੀਟੈਕਚਰ ਅਤੇ ਪੈਰਾਮੀਟਰਾਂ ਦੀ ਗਿਣਤੀ; ਤੁਹਾਡੇ ਮਾਡਲ ਦੀ ਚੁਣੀ ਹੋਈ ਆਰਕੀਟੈਕਚਰ ਅਤੇ ਪੈਰਾਮੀਟਰਾਂ ਦੀ ਸੰਖਿਆ, ਹੋਰ ਕਾਰਕਾਂ ਦੇ ਨਾਲ-ਨਾਲ, ਇਸ ਗੱਲ 'ਤੇ ਅਸਰ ਪਵੇਗੀ ਕਿ ਇਸਨੂੰ ਕਿੰਨੇ ਸਿਖਲਾਈ ਡੇਟਾ ਦੀ ਲੋੜ ਹੈ ਅਤੇ ਵਰਤੋਂ ਵਿੱਚ ਹੋਣ ਵੇਲੇ ਇਨਪੁਟ ਡੇਟਾ ਵਿਚਲੀਆਂ ਤਬਦੀਲੀਆਂ ਲਈ ਇਹ ਕਿੰਨਾ ਮਜ਼ਬੂਤ ਹੈ।

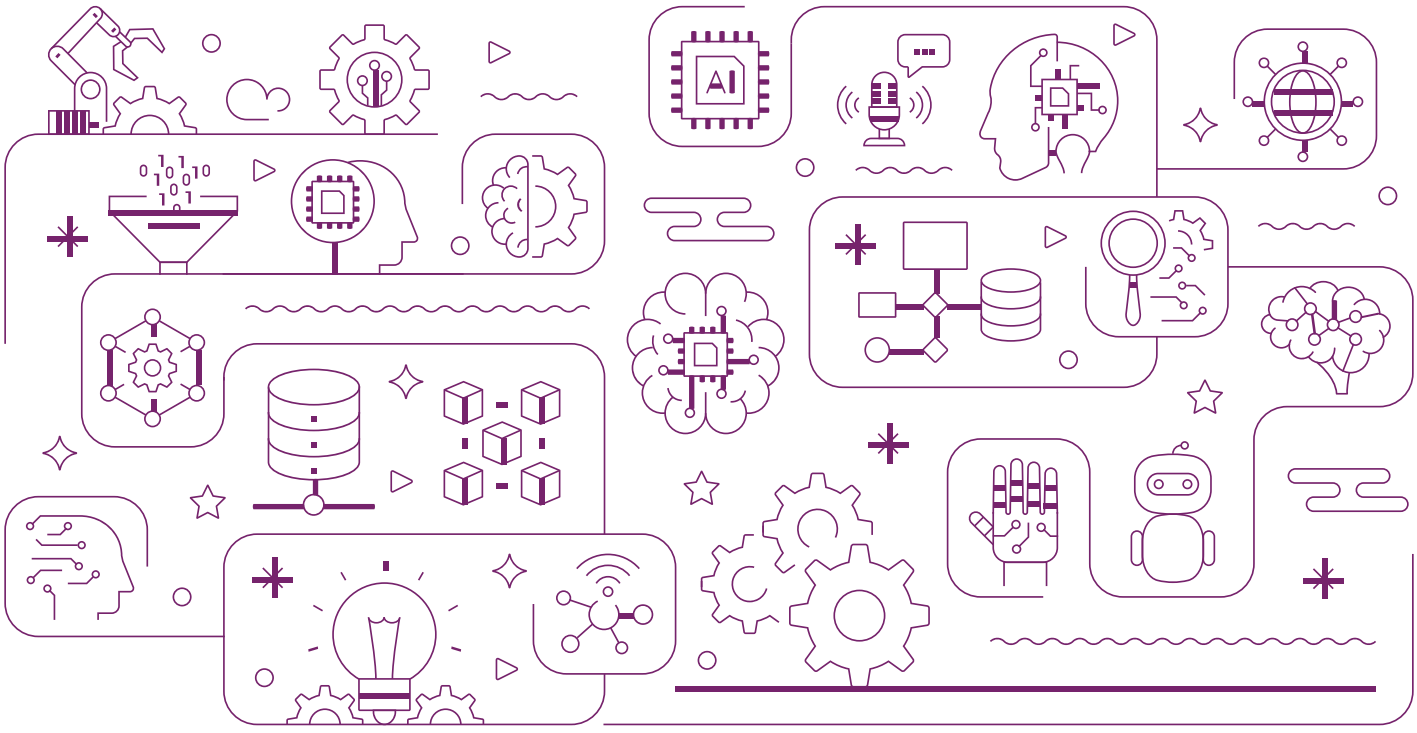
- ਤੁਹਾਡੇ ਵਰਤੋਂ ਲਈ ਮਾਡਲ ਦੇ ਢੁੱਕਵੇਂਪਨ ਅਤੇ/ਜਾਂ ਇਸਨੂੰ ਤੁਹਾਡੀ ਖ਼ਾਸ ਲੋੜ ਅਨੁਸਾਰ ਢਾਲਣ ਦੀ ਸੰਭਾਵਨਾ 'ਤੇ (ਉਦਾਹਰਨ ਲਈ ਫਾਈਨ-ਟਿਊਨਿੰਗ (ਛੋਟੇ-ਛੋਟੇ ਬਦਲਾਵਾਂ ਦੁਆਰਾ))

- ਤੁਹਾਡੇ ਮਾਡਲ ਦੇ ਆਉਟਪੁੱਟ ਨੂੰ ਇਕਸਾਰ ਕਰਨ, ਅਰਥ ਕੱਢਣ ਅਤੇ ਵਿਆਖਿਆ ਕਰਨ ਦੀ ਯੋਗਤਾ 'ਤੇ (ਉਦਾਹਰਨ ਲਈ ਡੀਬਗਿੰਗ (ਗ਼ਲਤੀਆਂ ਜਾਂ ਤਰੁੱਟੀਆਂ ਖ਼ਤਮ ਕਰਨਾ) , ਆਡਿਟ ਜਾਂ ਕਾਨੂੰਨ ਦੀ ਪਾਲਣਾ); ਅਰਥ ਕੱਢਣ ਵਿੱਚ ਵਧੇਰੇ ਮੁਸ਼ਕਲ ਵੱਡੇ ਅਤੇ ਗੁੰਝਲਦਾਰ ਮਾਡਲਾਂ ਨਾਲੋਂ ਸਰਲ, ਵਧੇਰੇ ਪਾਰਦਰਸ਼ੀ ਮਾਡਲਾਂ ਦੀ ਵਰਤੋਂ ਕਰਨ ਦੇ ਲਾਭ ਹੋ ਸਕਦੇ ਹਨ

- ਸਿਖਲਾਈ ਡੈਟਾਸੈੱਟ(ਟਾਂ) ਦੀਆਂ ਵਿਸ਼ੇਸ਼ਤਾਵਾਂ, ਆਕਾਰ, ਇਕਸਾਰਤਾ, ਗੁਣਵੱਤਾ, ਸੰਵੇਦਨਸ਼ੀਲਤਾ, ਉਮਰ, ਢੁੱਕਵੇਂਪਨ ਅਤੇ ਵਿਭਿੰਨਤਾ ਸਮੇਤ

- ਮਾਡਲ ਹਾਰਡਨਿੰਗ (ਜਿਵੇਂ ਕਿ ਵਿਰੋਧਮਈ ਸਿਖਲਾਈ), ਨਿਯਮਤੀਕਰਨ ਅਤੇ/ਜਾਂ ਗੁਪਤਤਾ ਵਧਾਉਣ ਵਾਲੀਆਂ ਤਕਨੀਕਾਂ ਦੀ ਵਰਤੋਂ ਕਰਨ ਦਾ ਮੁੱਲ
- ਮਾਡਲ ਜਾਂ ਫਾਊਂਡੇਸ਼ਨ ਮਾਡਲ, ਸਿਖਲਾਈ ਡੇਟਾ ਅਤੇ ਸੰਬੰਧਿਤ ਟੂਲਾਂ ਸਮੇਤ ਭਾਗਾਂ ਦੀ ਉਤਪਤੀ ਦਾ ਸਥਾਨ ਅਤੇ ਸਪਲਾਈ ਚੇਨ

ਇਹਨਾਂ ਵਿੱਚੋਂ ਕਿੰਨੇ ਕਾਰਕ ਸੁਰੱਖਿਆ ਨਤੀਜਿਆਂ ਨੂੰ ਪ੍ਰਭਾਵਿਤ ਕਰਦੇ ਹਨ ਇਸ ਬਾਰੇ ਹੋਰ ਜਾਣਕਾਰੀ ਲਈ, NCSC ਦੇ 'ਮਸ਼ੀਨ ਲਰਨਿੰਗ ਦੀ ਸੁਰੱਖਿਆ ਲਈ ਸਿਧਾਂਤ' ਵੇਖੋ, ਖ਼ਾਸ ਤੌਰ 'ਤੇ ਸੁਰੱਖਿਆ ਲਈ ਡਿਜ਼ਾਈਨ (ਮਾਡਲ ਆਰਕੀਟੈਕਚਰ)।



2. ਸੁਰੱਖਿਅਤ ਵਿਕਾਸ

ਇਸ ਭਾਗ ਵਿੱਚ ਅਜਿਹੇ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ ਸ਼ਾਮਲ ਹਨ ਜੋ AI ਪ੍ਰਣਾਲੀ ਵਿਕਸਤ ਕਰਨ ਦੇ ਜੀਵਨ ਚੱਕਰ ਦੇ **ਵਿਕਾਸ** ਪੜਾਅ 'ਤੇ ਲਾਗੂ ਹੁੰਦੇ ਹਨ, ਜਿਸ ਵਿੱਚ ਸਪਲਾਈ ਚੇਨ ਸੁਰੱਖਿਆ, ਦਸਤਾਵੇਜ਼, ਅਤੇ ਸੰਪਤੀ ਅਤੇ ਤਕਨੀਕੀ ਕਰਜ਼ਾ ਪ੍ਰਬੰਧਨ ਸ਼ਾਮਲ ਹਨ।

ਆਪਣੀ ਸਪਲਾਈ ਚੇਨ ਨੂੰ ਸੁਰੱਖਿਅਤ ਕਰੋ

ਤੁਸੀਂ ਇੱਕ ਪ੍ਰਣਾਲੀ ਦੇ ਜੀਵਨ ਚੱਕਰ ਵਿੱਚ ਆਪਣੀਆਂ AI ਸਪਲਾਈ ਚੇਨਾਂ ਦੀ ਸੁਰੱਖਿਆ ਦਾ ਮੁਲਾਂਕਣ ਅਤੇ ਨਿਗਰਾਨੀ ਕਰਦੇ ਹੋ, ਅਤੇ ਸਪਲਾਇਰਾਂ ਨੂੰ ਵੀ ਉਹਨਾਂ ਮਾਪਦੰਡਾਂ ਦੀ ਪਾਲਣਾ ਕਰਨ ਦੀ ਲੋੜ ਹੁੰਦੀ ਹੈ ਜੋ ਤੁਹਾਡੀ ਆਪਣੀ ਸੰਸਥਾ ਦੂਜੇ ਸਾਫ਼ਟਵੇਅਰ 'ਤੇ ਲਾਗੂ ਕਰਦੀ ਹੈ। ਜੇਕਰ ਸਪਲਾਇਰ ਤੁਹਾਡੀ ਸੰਸਥਾ ਦੇ ਮਾਪਦੰਡਾਂ ਦੀ ਪਾਲਣਾ ਨਹੀਂ ਕਰ ਸਕਦੇ, ਤਾਂ ਤੁਸੀਂ ਆਪਣੀਆਂ ਮੌਜੂਦਾ ਜ਼ੋਖਮ ਪ੍ਰਬੰਧਨ ਨੀਤੀਆਂ ਦੇ ਅਨੁਸਾਰ ਕੰਮ ਕਰਦੇ ਹੋ।

ਜਿੱਥੇ ਸੰਸਥਾ-ਅੰਦਰ ਉਤਪਾਦਨ ਨਹੀਂ ਕੀਤਾ ਜਾਂਦਾ, ਤੁਸੀਂ ਆਪਣੀਆਂ ਪ੍ਰਣਾਲੀਆਂ ਵਿੱਚ ਮਜ਼ਬੂਤ ਸੁਰੱਖਿਆ ਨੂੰ ਯਕੀਨੀ ਬਣਾਉਣ ਲਈ ਪ੍ਰਮਾਣਿਤ ਵਪਾਰਕ, ਓਪਨ ਸੋਰਸ, ਅਤੇ ਹੋਰ ਤੀਜੀ-ਧਿਰ ਦੇ ਡਿਵੈਲਪਰਾਂ ਤੋਂ ਚੰਗੀ ਤਰ੍ਹਾਂ ਸੁਰੱਖਿਅਤ ਅਤੇ ਚੰਗੀ ਤਰ੍ਹਾਂ ਦਸਤਾਵੇਜ਼ ਕੀਤੇ ਹੋਏ ਹਾਰਡਵੇਅਰ ਅਤੇ ਸਾਫ਼ਟਵੇਅਰ ਭਾਗਾਂ (ਉਦਾਹਰਨ ਲਈ, ਮਾਡਲ, ਡੇਟਾ, ਸਾਫ਼ਟਵੇਅਰ ਲਾਇਬ੍ਰੇਰੀਆਂ, ਮੋਡੀਊਲ, ਮਿਡਲਵੇਅਰ, ਫਰੇਮਵਰਕ, ਅਤੇ ਬਾਹਰੀ API) ਪ੍ਰਾਪਤ ਕਰਦੇ ਹੋ ਅਤੇ ਉਹਨਾਂ ਦਾ ਰੱਖ-ਰਖਾਵ ਕਰਦੇ ਹੋ।

ਜੇਕਰ ਸੁਰੱਖਿਆ ਮਾਪਦੰਡ ਪੂਰੇ ਨਹੀਂ ਕੀਤੇ ਜਾਂਦੇ ਹਨ, ਤਾਂ ਤੁਸੀਂ ਮਿਸ਼ਨ-ਨਾਜ਼ੁਕ ਪ੍ਰਣਾਲੀਆਂ ਲਈ ਵਿਕਲਪਿਕ ਹੱਲਾਂ ਵਿੱਚ ਅਸਫ਼ਲ ਰਹਿਣ ਲਈ ਤਿਆਰ ਹੋ। ਤੁਸੀਂ ਸਪਲਾਈ ਚੇਨ ਅਤੇ ਸਾਫ਼ਟਵੇਅਰ ਦੇ ਜੀਵਨ ਚੱਕਰਾਂ ਦੇ ਪ੍ਰਮਾਣੀਕਰਨ ਨੂੰ ਟਰੈਕ ਕਰਨ ਲਈ NCSC ਦੀ ਸਪਲਾਈ ਚੇਨ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ਾਂ ਅਤੇ ਫਰੇਮਵਰਕ ਜਿਵੇਂ ਕਿ ਸਾਫ਼ਟਵੇਅਰ ਆਰਟੀਫੈਕਟਸ (SLSA)[10] ਲਈ ਸਪਲਾਈ ਚੇਨ ਪੱਧਰਾਂ ਵਰਗੇ ਸਰੋਤਾਂ ਦੀ ਵਰਤੋਂ ਕਰਦੇ ਹੋ।

ਆਪਣੀਆਂ ਸੰਪਤੀਆਂ ਦੀ ਪਛਾਣ ਕਰੋ, ਟਰੈਕ ਕਰੋ ਅਤੇ ਸੁਰੱਖਿਅਤ ਕਰੋ

ਤੁਸੀਂ ਸੰਸਥਾ ਲਈ ਮਾਡਲ, ਡੇਟਾ (ਉਪਭੋਗਤਾ ਫੀਡਬੈਕ ਸਮੇਤ), ਪ੍ਰੋਂਪਟ, ਸਾਫ਼ਟਵੇਅਰ, ਦਸਤਾਵੇਜ਼, ਲੌਗਜ਼ ਅਤੇ ਮੁਲਾਂਕਣਾਂ ਸਮੇਤ (ਸੰਭਾਵੀ ਤੌਰ 'ਤੇ ਅਸੁਰੱਖਿਅਤ ਸਮਰੱਥਾਵਾਂ ਅਤੇ ਅਸਫ਼ਲਤਾ ਮੋਡਾਂ ਬਾਰੇ ਜਾਣਕਾਰੀ ਸਮੇਤ) ਤੁਹਾਡੀਆਂ AI-ਸੰਬੰਧੀ ਸੰਪਤੀਆਂ ਦੇ ਮੁੱਲ ਨੂੰ ਸਮਝਦੇ ਹੋ, ਇਹ ਪਛਾਣਦੇ ਹੋਏ ਕਿ ਉਹ ਮਹੱਤਵਪੂਰਨ ਨਿਵੇਸ਼ ਨੂੰ ਕਿੱਥੇ ਦਰਸਾਉਂਦੀਆਂ ਹਨ ਅਤੇ ਕਿੱਥੇ ਉਹਨਾਂ ਤੱਕ ਪਹੁੰਚ ਇੱਕ ਹਮਲਾਵਰ ਨੂੰ ਸਮਰੱਥ ਬਣਾਉਂਦੀ ਹੈ। ਤੁਸੀਂ ਲੌਗਜ਼ ਨੂੰ ਸੰਵੇਦਨਸ਼ੀਲ ਡੇਟਾ ਮੰਨਦੇ ਹੋ ਅਤੇ ਉਹਨਾਂ ਦੀ ਗੁਪਤਤਾ, ਅਖੰਡਤਾ ਅਤੇ ਉਪਲਬਧਤਾ ਦੀ ਸੁਰੱਖਿਆ ਲਈ ਨਿਯੰਤਰਣ ਲਾਗੂ ਕਰਦੇ ਹੋ।

ਤੁਸੀਂ ਜਾਣਦੇ ਹੋ ਕਿ ਤੁਹਾਡੀਆਂ ਸੰਪਤੀਆਂ ਕਿੱਥੇ ਰੱਖੀਆਂ ਗਈਆਂ ਹਨ ਅਤੇ ਕਿਸੇ ਵੀ ਸਬੰਧਿਤ ਜ਼ੋਖਮ ਦਾ ਮੁਲਾਂਕਣ ਕੀਤਾ ਅਤੇ ਸਵੀਕਾਰ ਕੀਤਾ ਹੈ। ਤੁਹਾਡੇ ਕੋਲ ਤੁਹਾਡੀਆਂ ਸੰਪਤੀਆਂ ਨੂੰ ਟਰੈਕ ਕਰਨ, ਪ੍ਰਮਾਣਿਤ ਕਰਨ, ਸੰਸਕਰਣ ਨਿਯੰਤਰਣ ਅਤੇ ਸੁਰੱਖਿਅਤ ਕਰਨ ਲਈ ਪ੍ਰਕਿਰਿਆਵਾਂ ਅਤੇ ਟੂਲ ਹਨ, ਅਤੇ ਅਣਅਧਿਕਾਰਤ ਪਹੁੰਚ ਕੀਤੇ ਜਾਣ ਦੀ ਹਾਲਾਤ ਵਿੱਚ ਇੱਕ ਚੰਗੀ ਜਾਣੀ-ਪਛਾਣੀ ਸਥਿਤੀ ਵਿੱਚ ਰੀਸਟੋਰ ਕਰ ਸਕਦੇ ਹੋ।

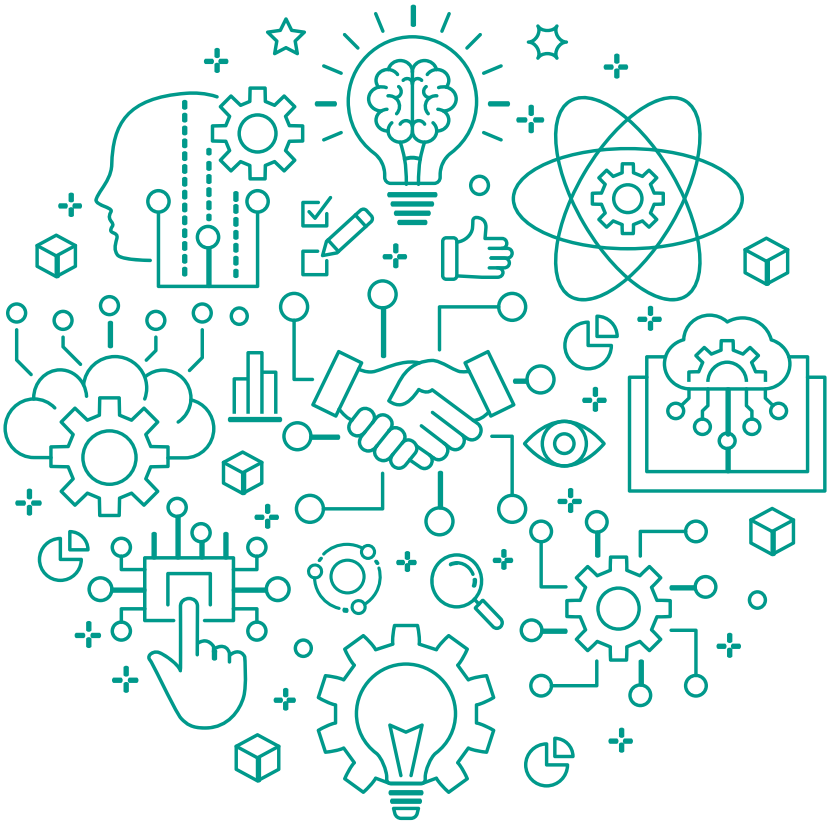
AI ਪ੍ਰਣਾਲੀਆਂ ਕਿਸ ਡੇਟਾ ਤੱਕ ਪਹੁੰਚ ਕਰ ਸਕਦੀਆਂ ਹਨ, ਅਤੇ AI ਦੁਆਰਾ ਤਿਆਰ ਕੀਤੀ ਸਮੱਗਰੀ ਨੂੰ ਇਸਦੀ ਸੰਵੇਦਨਸ਼ੀਲਤਾ (ਅਤੇ ਇਨਪੁਟਸ ਦੀ ਸੰਵੇਦਨਸ਼ੀਲਤਾ ਜੋ ਇਸਨੂੰ ਬਣਾਉਣ ਵਿੱਚ ਵਰਤੇ ਗਏ ਸਨ) ਲਈ ਪ੍ਰਬੰਧਿਤ ਕਰਨ ਲਈ ਤੁਹਾਡੇ ਕੋਲ ਪ੍ਰਕਿਰਿਆਵਾਂ ਅਤੇ ਨਿਯੰਤਰਣ ਲਾਗੂ ਹਨ।

ਤੁਹਾਡੇ ਡੇਟਾ, ਮਾਡਲਾਂ ਅਤੇ ਪ੍ਰੋਂਪਟਾਂ ਦੇ ਦਸਤਾਵੇਜ਼ ਬਣਾਓ

ਤੁਸੀਂ ਕਿਸੇ ਵੀ ਮਾਡਲਾਂ, ਡੇਟਾਸੈਟਾਂ ਅਤੇ ਮੈਟਾ-ਜਾਂ ਸਿਸਟਮ-ਪ੍ਰੋਂਪਟਾਂ ਦੀ ਰਚਨਾ, ਸੰਚਾਲਨ ਅਤੇ ਜੀਵਨ ਚੱਕਰ ਪ੍ਰਬੰਧਨ ਦਾ ਦਸਤਾਵੇਜ਼ ਬਣਾ ਰਹੇ ਹੋ। ਤੁਹਾਡੇ ਦਸਤਾਵੇਜ਼ਾਂ ਵਿੱਚ ਸੁਰੱਖਿਆ-ਸੰਬੰਧਿਤ ਜਾਣਕਾਰੀ ਜਿਵੇਂ ਕਿ ਸਿਖਲਾਈ ਡੇਟਾ ਦੇ ਸਰੋਤ (ਫਾਈਨ-ਟਿਊਨਿੰਗ ਡੇਟਾ ਅਤੇ ਮਨੁੱਖੀ ਜਾਂ ਹੋਰ ਸੰਚਾਲਨ ਸੰਬੰਧੀ ਫੀਡਬੈਕ ਸਮੇਤ), ਉਦੇਸ਼ਿਤ ਸਕੋਪ ਅਤੇ ਸੀਮਾਵਾਂ, ਸੁਰੱਖਿਆ ਘੇਰਾ, ਕ੍ਰਿਪਟੋਗ੍ਰਾਫਿਕ ਹੈਸ਼ ਜਾਂ ਦਸਤਖ਼ਤ, ਬਰਕਰਾਰ ਰੱਖਣ ਦਾ ਸਮਾਂ, ਸਮੀਖਿਆ ਕਰਨ ਦੀ ਸੁਝਾਈ ਗਈ ਬਾਰੰਬਾਰਤਾ ਅਤੇ ਸੰਭਾਵੀ ਅਸਫ਼ਲਤਾ ਮੋਡ ਸ਼ਾਮਲ ਹੁੰਦੀ ਹੈ। ਅਜਿਹਾ ਕਰਨ ਵਿੱਚ ਮੱਦਦ ਕਰਨ ਲਈ ਉਪਯੋਗੀ ਢਾਂਚੇ ਵਿੱਚ ਮਾਡਲ ਕਾਰਡ, ਡੇਟਾ ਕਾਰਡ ਅਤੇ ਸਾਫ਼ਟਵੇਅਰ ਬਿਲਜ਼ ਆਫ਼ ਮਟੀਰੀਅਲਜ਼ (SBOMs) ਸ਼ਾਮਲ ਹੁੰਦੇ ਹਨ। ਵਿਆਪਕ ਦਸਤਾਵੇਜ਼ਾਂ ਦਾ ਉਤਪਾਦਨ ਪਾਰਦਰਸ਼ਤਾ ਅਤੇ ਜਵਾਬਦੇਹੀ ਦਾ ਸਮਰਥਨ ਕਰਦਾ ਹੈ[11]।

ਆਪਣੇ ਤਕਨੀਕੀ ਕਰਜ਼ੇ ਦਾ ਪ੍ਰਬੰਧਨ ਕਰੋ

ਜਿਵੇਂ ਕਿ ਕਿਸੇ ਵੀ ਸਾਫ਼ਟਵੇਅਰ ਪ੍ਰਣਾਲੀ ਨਾਲ ਹੁੰਦਾ ਹੈ, ਤੁਸੀਂ ਇੱਕ AI ਪ੍ਰਣਾਲੀ ਦੇ ਜੀਵਨ ਚੱਕਰ ਦੌਰਾਨ ਆਪਣੇ 'ਤਕਨੀਕੀ ਕਰਜ਼ੇ' ਦੀ ਪਛਾਣ, ਟ੍ਰੈਕ ਅਤੇ ਪ੍ਰਬੰਧਨ ਕਰਦੇ ਹੋ (ਤਕਨੀਕੀ ਕਰਜ਼ਾ ਉਹ ਹੁੰਦਾ ਹੈ ਜਿੱਥੇ ਲੰਬੇ ਸਮੇਂ ਦੇ ਲਾਭਾਂ ਨੂੰ ਛੱਡਣ ਦੇ ਬਦਲੇ, ਥੋੜ੍ਹੇ ਸਮੇਂ ਵਿੱਚ ਨਤੀਜੇ ਪ੍ਰਾਪਤ ਕਰਨ ਲਈ ਕੰਮ ਕਰਨ ਦੇ ਸਭ ਤੋਂ ਵਧੀਆ ਤਰੀਕਿਆਂ ਦੀ ਬਜਾਏ ਘੱਟ ਵਧੀਆ ਤਰੀਕੇ ਵਾਲੇ ਇੰਜੀਨੀਅਰਿੰਗ ਫ਼ੈਸਲੇ ਲਏ ਜਾਂਦੇ ਹਨ)। ਵਿੱਤੀ ਕਰਜ਼ੇ ਦੀ ਤਰ੍ਹਾਂ, ਤਕਨੀਕੀ ਕਰਜ਼ਾ ਕੁਦਰਤੀ ਤੌਰ 'ਤੇ ਬੁਰਾ ਨਹੀਂ ਹੈ, ਪਰ ਵਿਕਾਸ ਦੇ ਸ਼ੁਰੂਆਤੀ ਪੜਾਵਾਂ ਤੋਂ ਪ੍ਰਬੰਧਿਤ ਕੀਤਾ ਜਾਣਾ ਚਾਹੀਦਾ ਹੈ[12]। ਤੁਸੀਂ ਸਵੀਕਾਰਦੇ ਹੋ ਕਿ ਅਜਿਹਾ ਕਰਨਾ ਮਿਆਰੀ ਸਾਫ਼ਟਵੇਅਰ ਨਾਲੋਂ ਇੱਕ AI ਮਾਮਲੇ ਵਿੱਚ ਵਧੇਰੇ ਚੁਣੌਤੀਪੂਰਨ ਹੋ ਸਕਦਾ ਹੈ, ਅਤੇ ਇਹ ਕਿ ਤੇਜ਼ ਵਿਕਾਸ ਚੱਕਰ ਅਤੇ ਚੰਗੀ ਤਰ੍ਹਾਂ ਸਥਾਪਿਤ ਪ੍ਰੋਟੋਕੋਲ ਅਤੇ ਇੰਟਰਫੇਸਾਂ ਦੀ ਘਾਟ ਕਾਰਨ ਤੁਹਾਡੇ ਤਕਨੀਕੀ ਕਰਜ਼ੇ ਦੇ ਪੱਧਰ ਉੱਚੇ ਹੋਣ ਦੀ ਸੰਭਾਵਨਾ ਹੈ। ਤੁਸੀਂ ਇਹ ਯਕੀਨੀ ਬਣਾਉਂਦੇ ਹੋ ਕਿ ਤੁਹਾਡੀਆਂ ਜੀਵਨ ਚੱਕਰ ਯੋਜਨਾਵਾਂ (AI ਪ੍ਰਣਾਲੀਆਂ ਨੂੰ ਬੰਦ ਕਰਨ ਦੀਆਂ ਪ੍ਰਕਿਰਿਆਵਾਂ ਸਮੇਤ) ਭਵਿੱਖੀ ਅਜਿਹੀਆਂ ਪ੍ਰਣਾਲੀਆਂ ਦੇ ਜ਼ੋਖਮਾਂ ਦਾ ਮੁਲਾਂਕਣ, ਸਵੀਕਾਰ ਕਰਦੀਆਂ ਅਤੇ ਘੱਟ ਕਰਦੀਆਂ ਹਨ।



3. ਸੁਰੱਖਿਅਤ ਤੈਨਾਤੀ

ਇਸ ਭਾਗ ਵਿੱਚ ਅਜਿਹੇ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ ਸ਼ਾਮਲ ਹਨ ਜੋ AI ਪ੍ਰਣਾਲੀ ਵਿਕਸਤ ਕਰਨ ਦੇ ਜੀਵਨ ਚੱਕਰ ਦੇ **ਤੈਨਾਤੀ** ਪੜਾਅ 'ਤੇ ਲਾਗੂ ਹੁੰਦੇ ਹਨ, ਜਿਸ ਵਿੱਚ ਬੁਨਿਆਦੀ ਢਾਂਚੇ ਅਤੇ ਮਾਡਲਾਂ ਨੂੰ ਅਣਅਧਿਕਾਰਤ ਪਹੁੰਚ, ਖ਼ਤਰੇ ਜਾਂ ਨੁਕਸਾਨ ਤੋਂ ਬਚਾਉਣਾ, ਘਟਨਾ ਪ੍ਰਬੰਧਨ ਦੀਆਂ ਪ੍ਰਕਿਰਿਆਵਾਂ ਦਾ ਵਿਕਾਸ ਕਰਨਾ, ਅਤੇ ਜ਼ਿੰਮੇਵਾਰੀ ਨਾਲ ਜਾਰੀ ਕਰਨਾ ਸ਼ਾਮਿਲ ਹਨ।

ਆਪਣੇ ਬੁਨਿਆਦੀ ਢਾਂਚੇ ਨੂੰ ਸੁਰੱਖਿਅਤ ਕਰਨਾ

ਤੁਸੀਂ ਆਪਣੀ ਪ੍ਰਣਾਲੀ ਦੇ ਜੀਵਨ ਚੱਕਰ ਦੇ ਹਰ ਹਿੱਸੇ ਵਿੱਚ ਵਰਤੇ ਜਾਣ ਵਾਲੇ ਬੁਨਿਆਦੀ ਢਾਂਚੇ ਲਈ ਚੰਗੇ ਬੁਨਿਆਦੀ ਢਾਂਚੇ ਦੇ ਸੁਰੱਖਿਆ ਸਿਧਾਂਤ ਲਾਗੂ ਕਰਦੇ ਹੋ। ਤੁਸੀਂ ਆਪਣੇ API, ਮਾਡਲਾਂ ਅਤੇ ਡੇਟਾ, ਅਤੇ ਉਹਨਾਂ ਦੀ ਸਿਖਲਾਈ ਅਤੇ ਪ੍ਰੋਸੈਸਿੰਗ ਪਾਈਪਲਾਈਨਾਂ, ਖੋਜ ਅਤੇ ਵਿਕਾਸ ਦੇ ਨਾਲ-ਨਾਲ ਤੈਨਾਤੀ ਵਿੱਚ ਉਚਿਤ ਪਹੁੰਚ ਨਿਯੰਤਰਣ ਲਾਗੂ ਕਰਦੇ ਹੋ। ਇਸ ਵਿੱਚ ਸੰਵੇਦਨਸ਼ੀਲ ਕੋਡ ਜਾਂ ਡੇਟਾ ਰੱਖਣ ਵਾਲੇ ਵਾਤਾਵਰਣਾਂ ਨੂੰ ਉਚਿਤ ਤਰੀਕੇ ਨਾਲ ਵੱਖ-ਵੱਖ ਕਰਨਾ ਸ਼ਾਮਲ ਹੈ। ਇਹ ਮਿਆਰੀ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਹਮਲਿਆਂ ਨੂੰ ਘਟਾਉਣ ਵਿੱਚ ਵੀ ਮੱਦਦ ਕਰੇਗਾ ਜਿਸਦਾ ਉਦੇਸ਼ ਮਾਡਲ ਨੂੰ ਚੋਰੀ ਕਰਨਾ ਜਾਂ ਇਸਦੇ ਪ੍ਰਦਰਸ਼ਨ ਨੂੰ ਨੁਕਸਾਨ ਪਹੁੰਚਾਉਣਾ ਹੈ।

ਆਪਣੇ ਮਾਡਲ ਦੀ ਲਗਾਤਾਰ ਸੁਰੱਖਿਆ ਕਰੋ

ਹਮਲਾਵਰ ਕਿਸੇ ਮਾਡਲ ਤੱਕ ਸਿੱਧੇ ਤੌਰ 'ਤੇ ਪਹੁੰਚ ਕਰਕੇ (ਮਾਡਲ ਦੇ ਵਜ਼ਨਾਂ ਨੂੰ ਹਾਸਲ ਕਰਕੇ) ਜਾਂ ਅਸਿੱਧੇ ਤੌਰ 'ਤੇ (ਕਿਸੇ ਐਪਲੀਕੇਸ਼ਨ ਜਾਂ ਸੇਵਾ ਰਾਹੀਂ ਮਾਡਲ ਦੀ ਪੁੱਛਗਿੱਛ ਕਰਕੇ) ਮਾਡਲ [13] ਜਾਂ ਜਿਸ ਡੇਟਾ 'ਤੇ ਇਸ ਮਾਡਲ ਨੂੰ ਸਿਖਲਾਈ ਦਿੱਤੀ ਗਈ ਸੀ[14] ਉਸਦੀ ਕਾਰਜਕੁਸ਼ਲਤਾ ਨੂੰ ਪੁਨਰਗਠਨ ਕਰਨ ਦੇ ਯੋਗ ਹੋ ਸਕਦੇ ਹਨ। ਹਮਲਾਵਰ ਮਾਡਲਾਂ, ਡੇਟਾ ਜਾਂ ਪ੍ਰੋਂਪਟ (ਤੁਰੰਤ ਕਾਰਕਾਂ) ਨਾਲ ਸਿਖਲਾਈ ਦੇ ਦੌਰਾਨ ਜਾਂ ਬਾਅਦ ਵਿੱਚ ਵੀ ਛੇੜਛਾੜ ਕਰ ਸਕਦੇ ਹਨ, ਆਉਟਪੁੱਟ ਨੂੰ ਗ਼ੈਰ-ਭਰੋਸੇਮੰਦ ਬਣਾ ਸਕਦੇ ਹਨ।

ਤੁਸੀਂ ਮਾਡਲ ਅਤੇ ਡੇਟਾ ਨੂੰ ਹੇਠ ਲਿਖੀਆਂ ਗੱਲਾਂ ਦੁਆਰਾ ਸਿੱਧੀ ਅਤੇ ਅਸਿੱਧੀ ਪਹੁੰਚ ਤੋਂ ਤਰਤੀਬਵਾਰ ਸੁਰੱਖਿਅਤ ਕਰਦੇ ਹੋ:

- ਮਿਆਰੀ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਦੇ ਸਭ ਤੋਂ ਵਧੀਆ ਕੰਮ ਕਰਨ ਦੇ ਤਰੀਕਿਆਂ ਨੂੰ ਲਾਗੂ ਕਰਕੇ
- ਗੁਪਤ ਜਾਣਕਾਰੀ ਤੱਕ ਪਹੁੰਚ ਕਰਨ, ਸੋਧਣ ਅਤੇ ਬਾਹਰ ਕੱਢਣ ਦੀਆਂ ਕੋਸ਼ਿਸ਼ਾਂ ਨੂੰ ਖੋਜਣ ਅਤੇ ਰੋਕਣ ਲਈ ਪੁੱਛਗਿੱਛ ਇੰਟਰਫੇਸ 'ਤੇ ਨਿਯੰਤਰਣ ਲਾਗੂ ਕਰਕੇ

ਇਹ ਯਕੀਨੀ ਬਣਾਉਣ ਲਈ ਕਿ ਖ਼ਪਤ ਪ੍ਰਣਾਲੀਆਂ ਮਾਡਲਾਂ ਨੂੰ ਪ੍ਰਮਾਣਿਤ ਕਰ ਸਕਦੀਆਂ ਹਨ, ਤੁਸੀਂ ਮਾਡਲ ਦੇ ਸਿਖਲਾਈ ਪ੍ਰਾਪਤ ਹੁੰਦੇ ਹੀ ਕ੍ਰਿਪਟੋਗ੍ਰਾਫਿਕ ਹੈਸ਼ਾਂ ਅਤੇ/ਜਾਂ ਮਾਡਲ ਫਾਈਲਾਂ (ਉਦਾਹਰਨ ਲਈ, ਮਾਡਲ ਦੇ ਵਜ਼ਨਾਂ) ਅਤੇ ਡੇਟਾ ਸੈਟਾਂ (ਚੈੱਕ-ਪੁਆਇੰਟਾਂ ਸਮੇਤ) ਦੇ ਦਸਤਖ਼ਤਾਂ ਦੀ ਗਣਨਾ ਅਤੇ ਸਾਂਝਾ ਕਰੋ। ਜਿਵੇਂ ਕਿ ਹਮੇਸ਼ਾ ਕ੍ਰਿਪਟੋਗ੍ਰਾਫੀ ਦੇ ਨਾਲ ਹੁੰਦਾ ਹੈ, ਵਧੀਆ ਕੁੰਜੀ ਪ੍ਰਬੰਧਨ ਜ਼ਰੂਰੀ ਹੈ[15]।

ਤੁਹਾਡੀ ਗੁਪਤਤਾ ਜ਼ੋਖਮ ਘਟਾਉਣ ਦੀ ਪਹੁੰਚ ਵਰਤੋਂ ਵਾਲੇ ਕੇਸ ਅਤੇ ਖ਼ਤਰੇ ਦੇ ਮਾਡਲ 'ਤੇ ਕਾਫ਼ੀ ਨਿਰਭਰ ਕਰੇਗੀ। ਕੁੱਝ ਐਪਲੀਕੇਸ਼ਨਾਂ, ਉਦਾਹਰਨ ਲਈ ਜਿਨ੍ਹਾਂ ਵਿੱਚ ਬਹੁਤ ਸੰਵੇਦਨਸ਼ੀਲ ਡੇਟਾ ਸ਼ਾਮਲ ਹੁੰਦਾ ਹੈ, ਨੂੰ ਸਿਧਾਂਤਕ ਗਾਰੰਟੀਆਂ ਦੀ ਲੋੜ ਹੋ ਸਕਦੀ ਹੈ ਜੋ ਲਾਗੂ ਕਰਨਾ ਔਖਾ ਜਾਂ ਮਹਿੰਗਾ ਹੋ ਸਕਦਾ ਹੈ। ਜੇਕਰ ਉਚਿਤ ਹੋਵੇ, ਗੁਪਤਤਾ-ਵਧਾਉਣ ਵਾਲੀਆਂ ਤਕਨਾਲੋਜੀਆਂ (ਜਿਵੇਂ ਕਿ ਵਿਭਿੰਨਤਾਂ ਗੁਪਤਤਾ ਜਾਂ ਹੋਮੋਮੋਰਫਿਕ ਇਨਕ੍ਰਿਪਸ਼ਨ) ਦੀ ਵਰਤੋਂ ਮਾਡਲਾਂ ਅਤੇ ਆਉਟਪੁੱਟਾਂ ਤੱਕ ਪਹੁੰਚ ਵਾਲੇ ਉਪਭੋਗਤਾਵਾਂ, ਵਰਤੋਂ ਕਰਤਾਵਾਂ ਅਤੇ ਹਮਲਾਵਰਾਂ ਨਾਲ ਜੁੜੇ ਜ਼ੋਖਮ ਦੇ ਪੱਧਰਾਂ ਦੀ ਪੜਚੋਲ ਕਰਨ ਜਾਂ ਭਰੋਸਾ ਦੇਣ ਲਈ ਕੀਤੀ ਜਾ ਸਕਦੀ ਹੈ।

ਘਟਨਾ ਪ੍ਰਬੰਧਨ ਪ੍ਰਕਿਰਿਆਵਾਂ ਵਿਕਸਿਤ ਕਰਨਾ

ਤੁਹਾਡੀਆਂ AI ਪ੍ਰਣਾਲੀਆਂ ਨੂੰ ਪ੍ਰਭਾਵਿਤ ਕਰਨ ਵਾਲੀਆਂ ਸੁਰੱਖਿਆ ਘਟਨਾਵਾਂ ਦੀ ਅਟੱਲਤਾ ਤੁਹਾਡੀ ਘਟਨਾ ਪ੍ਰਤੀ ਜਵਾਬੀ ਕਾਰਵਾਈ, ਵਾਧਾ ਅਤੇ ਉਪਾਅ ਯੋਜਨਾਵਾਂ ਵਿੱਚ ਝਲਕਦੀ ਹੈ। ਤੁਹਾਡੀਆਂ ਯੋਜਨਾਵਾਂ ਵੱਖੋ-ਵੱਖਰੀਆਂ ਸੰਭਾਵਨਾਵਾਂ ਨੂੰ ਦਰਸਾਉਂਦੀਆਂ ਹਨ ਅਤੇ ਪ੍ਰਣਾਲੀ ਅਤੇ ਵਿਆਪਕ ਖੋਜ ਦੇ ਵਿਕਸਿਤ ਹੋਣ 'ਤੇ ਨਿਯਮਿਤ ਤੌਰ 'ਤੇ ਮੁੜ ਮੁਲਾਂਕਣ ਕੀਤੀਆਂ ਜਾਂਦੀਆਂ ਹਨ। ਤੁਸੀਂ ਔਫ਼ਲਾਈਨ ਬੈਕਅੱਪ ਵਿੱਚ ਕੰਪਨੀ ਦੇ ਮਹੱਤਵਪੂਰਨ ਡਿਜ਼ੀਟਲ ਸਰੋਤਾਂ ਨੂੰ ਸਟੋਰ ਕਰਦੇ ਹੋ। ਜਵਾਬ ਦੇਣ ਵਾਲਿਆਂ ਨੂੰ AI-ਸੰਬੰਧਿਤ ਘਟਨਾਵਾਂ ਦਾ ਮੁਲਾਂਕਣ ਕਰਨ ਅਤੇ ਹੱਲ ਕਰਨ ਲਈ ਸਿਖਲਾਈ ਦਿੱਤੀ ਗਈ ਹੈ। ਤੁਸੀਂ ਉੱਚ-ਗੁਣਵੱਤਾ ਆਡਿਟ ਲੌਗ ਅਤੇ ਹੋਰ ਸੁਰੱਖਿਆ ਵਿਸ਼ੇਸ਼ਤਾਵਾਂ ਜਾਂ ਜਾਣਕਾਰੀ ਗਾਹਕਾਂ ਅਤੇ ਉਪਭੋਗਤਾਵਾਂ ਨੂੰ ਉਹਨਾਂ ਦੀਆਂ ਘਟਨਾ ਪ੍ਰਤੀਕਿਰਿਆ ਪ੍ਰਕਿਰਿਆਵਾਂ ਨੂੰ ਸਮਰੱਥ ਕਰਨ ਲਈ ਬਗ਼ੈਰ ਕਿਸੇ ਵਾਧੂ ਖ਼ਰਚੇ ਦੇ ਪ੍ਰਦਾਨ ਕਰਦੇ ਹੋ।

AI ਨੂੰ ਜ਼ਿੰਮੇਵਾਰੀ ਨਾਲ ਜਾਰੀ ਕਰਨਾ

ਤੁਸੀਂ ਮਾਡਲਾਂ, ਐਪਲੀਕੇਸ਼ਨਾਂ ਜਾਂ ਪ੍ਰਣਾਲੀਆਂ ਨੂੰ ਉਚਿਤ ਅਤੇ ਪ੍ਰਭਾਵੀ ਸੁਰੱਖਿਆ ਮੁਲਾਂਕਣ ਜਿਵੇਂ ਕਿ ਬੈਂਚਮਾਰਕਿੰਗ ਅਤੇ ਰੈੱਡ ਟੀਮਿੰਗ (ਨਾਲ ਹੀ ਨਾਲ ਹੋਰ ਟੈਸਟ ਜੋ ਇਹਨਾਂ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ਾਂ, ਜਿਵੇਂ ਕਿ ਸੁਰੱਖਿਆ ਜਾਂ ਨਿਰਪੱਖਤਾ ਦੇ ਦਾਇਰੇ ਤੋਂ ਬਾਹਰ ਹਨ) ਦੇ ਅਧੀਨ ਕਰਨ ਤੋਂ ਬਾਅਦ ਹੀ ਜਾਰੀ ਕਰਦੇ ਹੋ, ਅਤੇ ਤੁਸੀਂ ਆਪਣੇ ਉਪਭੋਗਤਾਵਾਂ ਨੂੰ ਜਾਣੀਆਂ-ਪਛਾਣੀਆਂ ਸੀਮਾਵਾਂ ਜਾਂ ਸੰਭਾਵੀ ਅਸਫ਼ਲਤਾ ਮੋਡਾਂ ਬਾਰੇ ਸਪੱਸ਼ਟ ਹੋ। ਓਪਨ-ਸੋਰਸ ਸੁਰੱਖਿਆ ਟੈਸਟਿੰਗ ਲਾਇਬ੍ਰੇਰੀਆਂ ਦੇ ਵੇਰਵੇ ਇਸ ਦਸਤਾਵੇਜ਼ ਦੇ ਅੰਤ ਵਿੱਚ ਅੱਗੇ ਪੜ੍ਹਨ ਵਾਲੇ ਭਾਗ ਵਿੱਚ ਦਿੱਤੇ ਗਏ ਹਨ।

ਵਰਤੋਂ ਕਰਨ ਵਾਲਿਆਂ ਲਈ ਸਹੀ ਚੀਜ਼ਾਂ ਕਰਨਾ ਆਸਾਨ ਬਣਾਉਣਾ

ਤੁਸੀਂ ਇਸ ਗੱਲ ਨੂੰ ਸਵੀਕਾਰਦੇ ਹੋ ਕਿ ਹਰੇਕ ਨਵੀਂ ਸੈਟਿੰਗ ਜਾਂ ਕੌਂਫਿਗਰੇਸ਼ਨ ਵਿਕਲਪ ਦਾ ਮੁਲਾਂਕਣ ਇਸ ਵੱਲੋਂ ਕੀਤੇ ਜਾਣ ਵਾਲੇ ਵਪਾਰਕ ਲਾਭ, ਅਤੇ ਇਸ ਦੁਆਰਾ ਪੇਸ਼ ਕੀਤੇ ਜਾਣ ਵਾਲੇ ਕਿਸੇ ਵੀ ਸੁਰੱਖਿਆ ਜ਼ੋਖਮ ਦੇ ਨਾਲ ਜੋੜ ਕੇ ਕੀਤਾ ਜਾਣਾ ਹੈ। ਆਦਰਸ਼ਕ ਤੌਰ 'ਤੇ, ਸਭ ਤੋਂ ਸੁਰੱਖਿਅਤ ਸੈਟਿੰਗ ਨੂੰ ਪ੍ਰਣਾਲੀ ਵਿੱਚ ਉਪਲਬਧ ਇੱਕੋ ਇੱਕ ਵਿਕਲਪ ਵਜੋਂ ਜੋੜਿਆ ਜਾਵੇਗਾ। ਜਦੋਂ ਕੌਂਫਿਗਰੇਸ਼ਨ ਜ਼ਰੂਰੀ ਹੁੰਦੀ ਹੈ, ਤਾਂ ਡਿਫੌਲਟ ਵਿਕਲਪ ਆਮ ਖ਼ਤਰਿਆਂ ਪ੍ਰਤੀ ਵਿਆਪਕ ਤੌਰ 'ਤੇ ਸੁਰੱਖਿਅਤ ਹੋਣਾ ਚਾਹੀਦਾ ਹੈ (ਜੋ ਕਿ, ਮੂਲ ਰੂਪ ਵਿੱਚ ਸੁਰੱਖਿਅਤ ਹੈ)। ਤੁਸੀਂ ਖ਼ਤਰਨਾਕ ਤਰੀਕਿਆਂ ਨਾਲ ਆਪਣੀ ਪ੍ਰਣਾਲੀ ਦੀ ਵਰਤੋਂ ਜਾਂ ਤੈਨਾਤੀ ਨੂੰ ਰੋਕਣ ਲਈ ਨਿਯੰਤਰਣ ਲਾਗੂ ਕਰਦੇ ਹੋ।

ਤੁਸੀਂ ਉਪਭੋਗਤਾਵਾਂ ਦਾ ਆਪਣੇ ਮਾਡਲ ਜਾਂ ਪ੍ਰਣਾਲੀ ਦੀ ਉਚਿਤ ਵਰਤੋਂ ਲਈ ਮਾਰਗਦਰਸ਼ਨ ਕਰਦੇ ਹੋ, ਜਿਸ ਵਿੱਚ ਕਮੀਆਂ ਨੂੰ ਉਜਾਗਰ ਕਰਨਾ ਅਤੇ ਸੰਭਾਵੀ ਅਸਫ਼ਲਤਾ ਮੋਡ ਸ਼ਾਮਲ ਹਨ। ਤੁਸੀਂ ਉਪਭੋਗਤਾਵਾਂ ਨੂੰ ਸਪੱਸ਼ਟ ਤੌਰ 'ਤੇ ਦੱਸਦੇ ਹੋ ਕਿ ਉਹ ਸੁਰੱਖਿਆ ਦੇ ਕਿਹੜੇ ਪਹਿਲੂਆਂ ਲਈ ਜ਼ਿੰਮੇਵਾਰ ਹਨ, ਅਤੇ ਇਸ ਬਾਰੇ ਪਾਰਦਰਸ਼ੀ ਹੋ ਕਿ ਉਹਨਾਂ ਦਾ ਡੇਟਾ ਕਿੱਥੇ (ਅਤੇ ਕਿਵੇਂ) ਵਰਤਿਆ ਜਾ ਸਕਦਾ ਹੈ, ਐਕਸੈਸ ਜਾਂ ਸਟੋਰ ਕੀਤਾ ਜਾ ਸਕਦਾ ਹੈ (ਉਦਾਹਰਨ ਲਈ, ਜੇਕਰ ਇਹ ਡੇਟਾ ਮਾਡਲ ਦੀ ਮੁੜ ਸਿਖਲਾਈ ਲਈ ਵਰਤਿਆ ਜਾਂਦਾ ਹੈ, ਜਾਂ ਕਰਮਚਾਰੀਆਂ ਜਾਂ ਭਾਈਵਾਲਾਂ ਦੁਆਰਾ ਇਸਦੀ ਸਮੀਖਿਆ ਕੀਤੀ ਜਾਂਦੀ ਹੈ।)।

4. ਸੁਰੱਖਿਅਤ ਸੰਚਾਲਨ ਅਤੇ ਰੱਖ-ਰਖਾਅ

ਇਸ ਭਾਗ ਵਿੱਚ ਅਜਿਹੇ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ ਹਨ ਜੋ AI ਪ੍ਰਣਾਲੀ ਵਿਕਸਤ ਕਰਨ ਦੇ ਜੀਵਨ ਚੱਕਰ ਦੇ **ਸੁਰੱਖਿਅਤ ਸੰਚਾਲਨ ਅਤੇ ਰੱਖ-ਰਖਾਅ** ਦੇ ਪੜਾਅ 'ਤੇ ਲਾਗੂ ਹੁੰਦੇ ਹਨ। ਇਹ ਲੌਗਿੰਗ (ਰੇਜ਼ਨਾਮਚੇ ਵਿਚ ਦਰਜ ਕਰਨ) ਅਤੇ ਨਿਗਰਾਨੀ, ਅੱਪਡੇਟ ਪ੍ਰਬੰਧਨ ਅਤੇ ਜਾਣਕਾਰੀ ਸਾਂਝਾ ਕਰਨ ਸਮੇਤ, ਇਸ ਪ੍ਰਣਾਲੀ ਦੇ ਲਾਗੂ ਹੋਣ ਤੋਂ ਬਾਅਦ ਖ਼ਾਸ ਤੌਰ 'ਤੇ ਚੁੱਕਵੀਆਂ ਕਾਰਵਾਈਆਂ ਬਾਰੇ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ ਪ੍ਰਦਾਨ ਕਰਦਾ ਹੈ।

ਆਪਣੀ ਪ੍ਰਣਾਲੀ ਦੇ ਵਿਵਹਾਰ ਦੀ ਨਿਗਰਾਨੀ ਕਰਨਾ

ਤੁਸੀਂ ਆਪਣੇ ਮਾਡਲ ਅਤੇ ਪ੍ਰਣਾਲੀ ਦੇ ਆਉਟਪੁੱਟ ਅਤੇ ਕਾਰਗੁਜ਼ਾਰੀ ਨੂੰ ਇਸ ਤਰ੍ਹਾਂ ਮਾਪਦੇ ਹੋ ਕਿ ਤੁਸੀਂ ਸੁਰੱਖਿਆ ਨੂੰ ਪ੍ਰਭਾਵਿਤ ਕਰਨ ਵਾਲੇ ਵਿਵਹਾਰ ਵਿੱਚ ਅਚਾਨਕ ਅਤੇ ਹੌਲੀ-ਹੌਲੀ ਆਉਂਦੀਆਂ ਤਬਦੀਲੀਆਂ ਨੂੰ ਦੇਖ ਸਕਦੇ ਹੋ। ਤੁਸੀਂ ਸੰਭਾਵੀ ਘੁਸਪੈਠ ਅਤੇ ਅਣਅਧਿਕਾਰਤ ਪਹੁੰਚ ਦੇ ਨਾਲ-ਨਾਲ ਕੁਦਰਤੀ ਡੇਟਾ ਰੁਝਾਨ ਦਾ ਲੇਖਾ-ਜੋਖਾ ਅਤੇ ਪਛਾਣ ਕਰ ਸਕਦੇ ਹੋ।

ਆਪਣੀ ਪ੍ਰਣਾਲੀ ਦੇ ਇਨਪੁਟਸ ਦੀ ਨਿਗਰਾਨੀ ਕਰਨਾ

ਗੁਪਤਤਾ ਅਤੇ ਡੇਟਾ ਸੁਰੱਖਿਆ ਲੋੜਾਂ ਦੇ ਅਨੁਸਾਰ, ਤੁਸੀਂ ਅਣਅਧਿਕਾਰਤ ਜਾਂ ਦੁਰਵਰਤੋਂ ਦੇ ਮਾਮਲੇ ਵਿੱਚ ਕਾਨੂੰਨ ਪਾਲਣਾ ਦੀਆਂ ਜ਼ਿੰਮੇਵਾਰੀਆਂ, ਆਡਿਟ, ਜਾਂਚ ਅਤੇ ਉਪਾਅ ਨੂੰ ਸਮਰੱਥ ਬਣਾਉਣ ਲਈ ਆਪਣੀ ਪ੍ਰਣਾਲੀ (ਜਿਵੇਂ ਕਿ ਅਨੁਮਾਨ ਬੇਨਤੀਆਂ, ਸਵਾਲ ਜਾਂ ਪ੍ਰੋਂਪਟ) ਵਿੱਚ ਇਨਪੁਟਸ ਦੀ ਨਿਗਰਾਨੀ ਅਤੇ ਲੌਗ ਕਰਦੇ ਹੋ। ਇਸ ਵਿੱਚ ਡਿਸਟਰੀਬਿਊਸ਼ਨ ਤੋਂ ਬਾਹਰ ਅਤੇ/ਜਾਂ ਵਿਰੋਧਮਈ ਇਨਪੁਟਸ ਦਾ ਸਾਫ਼-ਸਾਫ਼ ਪਤਾ ਲਗਾਉਣਾ ਸ਼ਾਮਲ ਹੋ ਸਕਦਾ ਹੈ, ਜਿਸ ਵਿੱਚ ਉਹ ਵੀ ਸ਼ਾਮਲ ਹਨ ਜੋ ਡੇਟਾ ਤਿਆਰ ਕਰਨ ਦੇ ਕਦਮਾਂ ਦਾ ਸ਼ੋਸ਼ਣ ਕਰਨਾ ਚਾਹੁੰਦੇ ਹਨ (ਜਿਵੇਂ ਕਿ ਚਿੱਤਰਾਂ ਲਈ ਕ੍ਰੌਪਿੰਗ ਅਤੇ ਰੀਸਾਈਜ਼)।

ਅੱਪਡੇਟਾਂ ਲਈ 'ਡਿਜ਼ਾਇਨ ਦੁਆਰਾ ਸੁਰੱਖਿਅਤ' ਪਹੁੰਚ ਦਾ ਪਾਲਣ ਕਰਨਾ

ਤੁਸੀਂ ਹਰੇਕ ਉਤਪਾਦ ਵਿੱਚ ਮੂਲ ਰੂਪ ਵਿੱਚ ਸਵੈਚਲਿਤ ਅੱਪਡੇਟ ਸ਼ਾਮਲ ਕਰਦੇ ਹੋ ਅਤੇ ਉਹਨਾਂ ਨੂੰ ਵਿਤਰਨ ਕਰਨ ਲਈ ਸੁਰੱਖਿਅਤ, ਮਾਡਿਊਲਰ ਅੱਪਡੇਟ ਪ੍ਰਕਿਰਿਆਵਾਂ ਦੀ ਵਰਤੋਂ ਕਰਦੇ ਹੋ। ਤੁਹਾਡੀਆਂ ਅੱਪਡੇਟ ਪ੍ਰਕਿਰਿਆਵਾਂ (ਟੈਸਟਿੰਗ ਅਤੇ ਮੁਲਾਂਕਣ ਪ੍ਰਣਾਲੀਆਂ ਸਮੇਤ) ਇਸ ਤੱਥ ਨੂੰ ਦਰਸਾਉਂਦੀਆਂ ਹਨ ਕਿ ਡੇਟਾ, ਮਾਡਲਾਂ ਜਾਂ ਪ੍ਰੋਂਪਟਾਂ ਵਿੱਚ ਤਬਦੀਲੀਆਂ ਪ੍ਰਣਾਲੀ ਦੇ ਵਿਵਹਾਰ ਵਿੱਚ ਤਬਦੀਲੀਆਂ ਲਿਆ ਸਕਦੀਆਂ ਹਨ (ਉਦਾਹਰਨ ਲਈ, ਤੁਸੀਂ ਵੱਡੇ ਅੱਪਡੇਟਾਂ ਨੂੰ ਨਵੇਂ ਸੰਸਕਰਣਾਂ ਵਾਂਗ ਵਰਤਦੇ ਹੋ)। ਤੁਸੀਂ ਮੁਲਾਂਕਣ ਕਰਨ ਅਤੇ ਮਾਡਲ ਤਬਦੀਲੀਆਂ ਦਾ ਜਵਾਬ ਦੇਣ ਲਈ ਉਪਭੋਗਤਾਵਾਂ ਦਾ ਸਮਰਥਨ ਕਰਦੇ ਹੋ (ਉਦਾਹਰਨ ਲਈ ਪੁਰਵਦਰਸ਼ਨ ਪਹੁੰਚ ਅਤੇ ਸੰਸਕਰਣ ਵਾਲੇ API ਪ੍ਰਦਾਨ ਕਰਕੇ)।

ਸਿੱਖੇ ਗਏ ਸਬਕਾਂ ਨੂੰ ਇਕੱਠਾ ਕਰੋ ਅਤੇ ਸਾਂਝਾ ਕਰੋ

ਤੁਸੀਂ ਜਾਣਕਾਰੀ ਸਾਂਝੀ ਕਰਨ ਵਾਲੇ ਭਾਈਚਾਰਿਆਂ ਵਿੱਚ ਹਿੱਸਾ ਲੈਂਦੇ ਪਾਉਂਦੇ ਹੋ, ਉਦਯੋਗ, ਅਕਾਦਮਿਕ ਅਤੇ ਸਰਕਾਰਾਂ ਦੇ ਗਲੋਬਲ ਈਕੋਸਿਸਟਮ ਵਿੱਚ ਸਹਿਯੋਗ ਕਰਦੇ ਹੋਏ ਕੰਮ ਕਰਨ ਦੇ ਸਭ ਤੋਂ ਵਧੀਆ ਤਰੀਕਿਆਂ ਨੂੰ ਸਾਂਝਾ ਕਰਦੇ ਹੋ। ਤੁਸੀਂ ਪ੍ਰਣਾਲੀ ਸੁਰੱਖਿਆ ਦੇ ਸੰਬੰਧ ਵਿੱਚ ਫੀਡਬੈਕ ਦੇਣ ਲਈ ਸੰਚਾਰ ਦੀਆਂ ਲਾਈਨਾਂ ਖੁੱਲ੍ਹੀਆਂ ਰੱਖਦੇ ਹੋ, ਤੁਹਾਡੀ ਸੰਸਥਾ ਲਈ ਅੰਦਰੂਨੀ ਅਤੇ ਬਾਹਰੀ ਤੌਰ 'ਤੇ, ਜਿਸ ਵਿੱਚ ਸੁਰੱਖਿਆ ਖੋਜਕਰਤਾਵਾਂ ਨੂੰ ਕਮਜ਼ੋਰੀਆਂ ਦੀ ਖੋਜ ਕਰਨ ਅਤੇ ਰਿਪੋਰਟ ਕਰਨ ਲਈ ਸਹਿਮਤੀ ਪ੍ਰਦਾਨ ਕਰਨ ਸ਼ਾਮਿਲ ਹੈ। ਲੋੜ ਪੈਣ 'ਤੇ, ਤੁਸੀਂ ਮੁੱਦਿਆਂ ਨੂੰ ਵਿਆਪਕ ਭਾਈਚਾਰੇ ਤੱਕ ਪਹੁੰਚਾਉਂਦੇ ਹੋ, ਉਦਾਹਰਨ ਲਈ ਵਿਸਤ੍ਰਿਤ ਅਤੇ ਸੰਪੂਰਨ ਆਮ ਕਮਜ਼ੋਰੀ ਦੀ ਗਣਨਾ ਸਮੇਤ, ਕਮਜ਼ੋਰੀ ਦੇ ਖੁਲਾਸੇ ਦਾ ਜਵਾਬ ਦੇਣ ਵਾਲੇ ਬੁਲੇਟਿਨ ਪ੍ਰਕਾਸ਼ਿਤ ਕਰਨਾ। ਤੁਸੀਂ ਸਮੱਸਿਆਵਾਂ ਨੂੰ ਜਲਦੀ ਅਤੇ ਉਚਿਤ ਢੰਗ ਨਾਲ ਘਟਾਉਣ ਅਤੇ ਹੱਲ ਕਰਨ ਲਈ ਕਾਰਵਾਈ ਕਰਦੇ ਹੋ।

ਅੱਗੇ ਪੜ੍ਹਨਾ

AI ਵਿਕਾਸ

ਮਸ਼ੀਨ ਲਰਨਿੰਗ ਦੀ ਸੁਰੱਖਿਆ ਲਈ ਸਿਧਾਂਤ
ML ਭਾਗ ਨਾਲ ਕਿਸੇ ਪ੍ਰਣਾਲੀ ਨੂੰ ਵਿਕਸਤ ਕਰਨ, ਤੈਨਾਤ ਕਰਨ ਜਾਂ ਚਲਾਉਣ ਲਈ NCSC ਦੇ ਵਿਸਥਾਰ ਪੂਰਵਕ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼।

ਡਿਜ਼ਾਇਨ ਦੁਆਰਾ ਸੁਰੱਖਿਅਤ - ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਜ਼ੋਖਮ ਦੇ ਸੰਤੁਲਨ ਨੂੰ ਬਦਲ ਰਿਹਾ ਹੈ: ਡਿਜ਼ਾਈਨ ਦੁਆਰਾ ਸੁਰੱਖਿਅਤ ਸਾਫ਼ਟਵੇਅਰ ਲਈ ਸਿਧਾਂਤ ਅਤੇ ਪਹੁੰਚਾਂ
CISA, NCSC ਅਤੇ ਹੋਰ ਏਜੰਸੀਆਂ ਦੁਆਰਾ ਸਹਿ-ਲਿਖਤ, ਇਹ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ ਦੱਸਦੀਆਂ ਹਨ ਕਿ ਕਿਵੇਂ AI ਸਮੇਤ ਸਾਫ਼ਟਵੇਅਰ ਪ੍ਰਣਾਲੀਆਂ ਦੇ ਨਿਰਮਾਤਾਵਾਂ ਨੂੰ ਉਤਪਾਦ ਵਿਕਾਸ ਦੇ ਡਿਜ਼ਾਇਨ ਪੜਾਅ ਵਿੱਚ ਹੀ ਸੁਰੱਖਿਆ ਦਾ ਨਿਰਮਾਣ ਕਰਨ ਲਈ ਕਦਮ ਚੁੱਕਣੇ ਚਾਹੀਦੇ ਹਨ, ਅਤੇ ਓਹੀ ਉਤਪਾਦ ਅੱਗੇ ਭੇਜਣੇ ਚਾਹੀਦੇ ਹਨ ਜੋ ਡੱਬੇ ਤੋਂ ਬਾਹਰ ਕੱਢਣ ਵੇਲੇ ਸੁਰੱਖਿਅਤ ਹੁੰਦੇ ਹਨ।

ਸੰਖੇਪ ਵਿੱਚ AI ਸੁਰੱਖਿਆ ਚਿੰਤਾਵਾਂ
ਸੂਚਨਾ ਸੁਰੱਖਿਆ ਲਈ ਜਰਮਨੀ ਦੇ ਸੰਘੀ ਦਫ਼ਤਰ (BSI) ਦੁਆਰਾ ਤਿਆਰ ਕੀਤਾ ਗਿਆ, ਇਹ ਦਸਤਾਵੇਜ਼ ਮਸ਼ੀਨ ਲਰਨਿੰਗ ਪ੍ਰਣਾਲੀਆਂ 'ਤੇ ਸੰਭਾਵਿਤ ਹਮਲਿਆਂ ਅਤੇ ਉਨ੍ਹਾਂ ਹਮਲਿਆਂ ਦੇ ਵਿਰੁੱਧ ਸੰਭਾਵੀ ਬਚਾਅ ਪ੍ਰਤੀ ਜਾਣ-ਪਛਾਣ ਪ੍ਰਦਾਨ ਕਰਦਾ ਹੈ।

ਐਡਵਾਂਸਡ AI ਸਿਸਟਮ ਵਿਕਸਿਤ ਕਰਨ ਵਾਲੀਆਂ ਸੰਸਥਾਵਾਂ ਲਈ ਹੀਰੋਸ਼ੀਮਾ ਪ੍ਰਕਿਰਿਆ ਅੰਤਰਰਾਸ਼ਟਰੀ ਮਾਰਗਦਰਸ਼ਕ ਸਿਧਾਂਤ ਅਤੇ ਐਡਵਾਂਸਡ AI ਸਿਸਟਮ ਵਿਕਸਿਤ ਕਰਨ ਵਾਲੀਆਂ ਸੰਸਥਾਵਾਂ ਲਈ ਹੀਰੋਸ਼ੀਮਾ ਪ੍ਰਕਿਰਿਆ ਅੰਤਰਰਾਸ਼ਟਰੀ ਆਚਾਰ-ਸੰਹਿਤਾ ਇਹ ਦਸਤਾਵੇਜ਼ , G7 Hiroshima AI ਪ੍ਰਕਿਰਿਆ ਦੇ ਹਿੱਸੇ ਵਜੋਂ ਤਿਆਰ ਕੀਤਾ ਗਿਆ ਹੈ, ਦੁਨੀਆਂ ਭਰ ਵਿੱਚ ਸੁਰੱਖਿਅਤ, ਮਹਿਫ਼ੂਜ਼ ਅਤੇ ਭਰੋਸੇਮੰਦ AI ਨੂੰ ਉਤਸ਼ਾਹਿਤ ਕਰਨ ਦੇ ਉਦੇਸ਼ ਨਾਲ ਸਭ ਤੋਂ ਐਡਵਾਂਸ ਬੁਨਿਆਦੀ ਮਾਡਲ ਅਤੇ ਜਨਰੇਟਿਵ AI ਪ੍ਰਣਾਲੀਆਂ ਸਮੇਤ ਸਭ ਤੋਂ ਐਡਵਾਂਸ AI ਪ੍ਰਣਾਲੀਆਂ ਨੂੰ ਵਿਕਸਤ ਕਰਨ ਵਾਲੀਆਂ ਸੰਸਥਾਵਾਂ ਲਈ ਮਾਰਗਦਰਸ਼ਨ ਪ੍ਰਦਾਨ ਕਰਦਾ ਹੈ।

AI ਪ੍ਰਮਾਣੀਕਰਨ
ਸਿੰਗਾਪੁਰ ਦਾ AI ਗਵਰਨੈਂਸ ਟੈਸਟਿੰਗ ਫਰੇਮਵਰਕ ਅਤੇ ਸਾਫ਼ਟਵੇਅਰ ਟੂਲਕਿੱਟ ਜੋ ਮਿਆਰੀ ਟੈਸਟਾਂ ਰਾਹੀਂ ਅੰਤਰਰਾਸ਼ਟਰੀ ਤੌਰ 'ਤੇ ਮਾਨਤਾ ਪ੍ਰਾਪਤ ਸਿਧਾਂਤਾਂ ਦੇ ਗੁੱਟ ਪ੍ਰਤੀ AI ਪ੍ਰਣਾਲੀਆਂ ਦੀ ਕਾਰਗੁਜ਼ਾਰੀ ਨੂੰ ਪ੍ਰਮਾਣਿਤ ਕਰਦੀ ਹੈ।

AI ਲਈ ਚੰਗੀ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਦੇ ਕੰਮ ਕਰਨ ਦੇ ਤਰੀਕਿਆਂ ਲਈ ਬੁਹ-ਪਰਤੀ ਫਰੇਮਵਰਕ — ENISA (europa.eu)
ਕੌਮੀ ਸਮਰੱਥ ਅਥਾਰਟੀਆਂ ਅਤੇ AI ਹਿੱਤਧਾਰਕਾਂ ਨੂੰ ਉਹਨਾਂ ਕਦਮਾਂ ਬਾਰੇ ਮਾਰਗਦਰਸ਼ਨ ਦੇਣ ਲਈ ਇੱਕ ਢਾਂਚਾ ਜੋ ਉਹਨਾਂ ਨੂੰ ਉਹਨਾਂ ਦੀਆਂ AI ਪ੍ਰਣਾਲੀਆਂ, ਸੰਚਾਲਨ ਅਤੇ ਪ੍ਰਕਿਰਿਆਵਾਂ ਨੂੰ ਸੁਰੱਖਿਅਤ ਕਰਨ ਲਈ ਅਪਣਾਉਣ ਦੀ ਲੋੜ ਹੈ।

ISO 5338: AI ਪ੍ਰਣਾਲੀ ਦੀਆਂ ਜੀਵਨ ਚੱਕਰ ਪ੍ਰਕਿਰਿਆਵਾਂ (ਸਮੀਖਿਆ ਅਧੀਨ)
ਮਸ਼ੀਨ ਲਰਨਿੰਗ ਅਤੇ ਹਿਊਰਿਸਟਿਕ ਪ੍ਰਣਾਲੀਆਂ 'ਤੇ ਆਧਾਰਿਤ AI ਪ੍ਰਣਾਲੀਆਂ ਦੇ ਜੀਵਨ ਚੱਕਰ ਦਾ ਵਰਣਨ ਕਰਨ ਲਈ ਪ੍ਰਕਿਰਿਆਵਾਂ ਅਤੇ ਸੰਬੰਧਿਤ ਸੰਕਲਪਾਂ ਦਾ ਇੱਕ ਸਮੂਹ।

AI ਕਲਾਊਡ ਸੇਵਾ ਪਾਲਣਾ ਮਾਪਦੰਡ ਕੈਟਾਲਾਗ (AIC4)
BSI ਦਾ AI ਕਲਾਊਡ ਸੇਵਾ ਪਾਲਣਾ ਮਾਪਦੰਡ ਕੈਟਾਲਾਗ AI-ਵਿਸ਼ੇਸ਼ ਮਾਪਦੰਡ ਪ੍ਰਦਾਨ ਕਰਦਾ ਹੈ, ਜੋ ਇੱਕ AI ਸੇਵਾ ਦੇ ਜੀਵਨ ਚੱਕਰ ਵਿੱਚ ਇਸਦੀ ਸੁਰੱਖਿਆ ਦੇ ਮੁਲਾਂਕਣ ਨੂੰ ਸਮਰੱਥ ਬਣਾਉਂਦਾ ਹੈ।

NIST IR 8269 (ਖਰੜਾ) 'ਵਿਰੋਧਮਈ ਮਸ਼ੀਨ ਲਰਨਿੰਗ' ਦਾ ਇੱਕ ਵਰਗੀਕਰਨ ਅਤੇ ਪਰਿਭਾਸ਼ਾ
ਮਸ਼ੀਨ ਲਰਨਿੰਗ ਅਤੇ ਹਿਊਰਿਸਟਿਕ ਪ੍ਰਣਾਲੀਆਂ 'ਤੇ ਆਧਾਰਿਤ AI ਪ੍ਰਣਾਲੀਆਂ ਦੇ ਜੀਵਨ ਚੱਕਰ ਦਾ ਵਰਣਨ ਕਰਨ ਲਈ ਪ੍ਰਕਿਰਿਆਵਾਂ ਅਤੇ ਸੰਬੰਧਿਤ ਸੰਕਲਪਾਂ ਦਾ ਇੱਕ ਸਮੂਹ।

MITRE ATLAS
ਮਸ਼ੀਨ ਲਰਨਿੰਗ (ML) ਪ੍ਰਣਾਲੀਆਂ ਲਈ ਵਿਰੋਧਮਈ ਰਣਨੀਤੀਆਂ, ਤਕਨੀਕਾਂ, ਅਤੇ ਕੇਸ ਸਟੱਡੀਆਂ ਦਾ ਇੱਕ ਗਿਆਨ ਅਧਾਰ, ਜੋ ਕਿ MITER ATT&CK ਫਰੇਮਵਰਕ ਦੇ ਬਾਅਦ ਤਿਆਰ ਕੀਤਾ ਗਿਆ ਹੈ ਅਤੇ ਲਿੰਕ ਕੀਤਾ ਗਿਆ ਹੈ।

ਵਿਨਾਸ਼ਕਾਰੀ AI ਜ਼ੋਖਮਾਂ ਦੀ ਇੱਕ ਸੰਖੇਪ ਜਾਣਕਾਰੀ (2023)
ਸੈਂਟਰ ਫਾਰ AI ਸੇਫ਼ਟੀ ਦੁਆਰਾ ਤਿਆਰ ਕੀਤਾ ਗਿਆ, ਇਹ ਦਸਤਾਵੇਜ਼ AI ਦੁਆਰਾ ਪੈਦਾ ਹੋਣ ਵਾਲੇ ਜ਼ੋਖਮਾਂ ਦੇ ਖੇਤਰਾਂ ਨੂੰ ਨਿਰਧਾਰਤ ਕਰਦਾ ਹੈ।

ਵੱਡੇ ਭਾਸ਼ਾ ਮਾਡਲ (LLM): ਉਦਯੋਗ ਅਤੇ ਅਥਾਰਟੀਆਂ ਲਈ ਮੌਕੇ ਅਤੇ ਜ਼ੋਖਮ
BSI ਦੁਆਰਾ ਉਹਨਾਂ ਕੰਪਨੀਆਂ, ਅਥਾਰਟੀਆਂ ਅਤੇ ਡਿਵੈਲਪਰਾਂ ਲਈ ਤਿਆਰ ਕੀਤਾ ਗਿਆ ਦਸਤਾਵੇਜ਼ ਜੋ LLM ਨੂੰ ਵਿਕਸਤ ਕਰਨ, ਤਾਇਨਾਤ ਕਰਨ ਅਤੇ/ਜਾਂ ਵਰਤਣ ਦੇ ਮੌਕਿਆਂ ਅਤੇ ਜ਼ੋਖਮਾਂ ਬਾਰੇ ਹੋਰ ਜਾਣਨਾ ਚਾਹੁੰਦੇ ਹਨ।

ਉਪਭੋਗਤਾਵਾਂ ਦੀ AI ਮਾਡਲਾਂ ਦੀ ਸੁਰੱਖਿਆ ਜਾਂਚ ਵਿੱਚ ਮੱਦਦ ਕਰਨ ਲਈ ਓਪਨ-ਸੋਰਸ (ਖੁੱਲ੍ਹੇ-ਸਰੋਤਾਂ ਵਾਲੇ) ਪ੍ਰੋਜੈਕਟਾਂ ਵਿੱਚ ਸ਼ਾਮਲ ਹਨ:

- Adversarial Robustness Toolbox (ਐਡਵਰਸੇਰੀਅਲ ਰੋਬੱਸਟਨੈੱਸ ਟੂਲਬਾਕਸ) (IBM)
- CleverHans (ਕਲੈਵਰ ਹੈਨਜ਼) (ਟੋਰਾਂਟੋ ਯੂਨੀਵਰਸਿਟੀ)
- TextAttack (ਟੈਕਸਟ ਅਟੈਕ) (ਵਰਜੀਨੀਆ ਯੂਨੀਵਰਸਿਟੀ)
- Prompt Bench (ਪ੍ਰੋਂਪਟ ਬੈਂਚ) (Microsoft)
- Counterfit (ਕਾਊਂਟਰਫਿੱਟ) (Microsoft)
- AI Verify (AI ਵੈਰੀਫਾਈ) (ਇਨਫ਼ੋਕਾਮ ਮੀਡੀਆ ਡਿਵੈਲਪਮੈਂਟ ਅਥਾਰਟੀ, ਸਿੰਗਾਪੁਰ)

ਸਾਈਬਰ ਸੁਰੱਖਿਆ

CISA ਦੇ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਕਾਰਗੁਜ਼ਾਰੀ ਟੀਚੇ
ਸੁਰੱਖਿਆਵਾਂ ਦਾ ਇੱਕ ਸਾਂਝਾ ਸਮੂਹ ਜੋ ਸਾਰੀਆਂ ਮਹੱਤਵਪੂਰਨ ਬੁਨਿਆਦੀ ਢਾਂਚਾ ਸੰਸਥਾਵਾਂ ਨੂੰ ਜਾਣੇ-ਪਛਾਣੇ ਜ਼ੋਖਮਾਂ ਅਤੇ ਵਿਰੋਧੀ ਤਕਨੀਕਾਂ ਦੀ ਸੰਭਾਵਨਾ ਅਤੇ ਪ੍ਰਭਾਵ ਨੂੰ ਅਰਥਪੂਰਨ ਤੌਰ 'ਤੇ ਘਟਾਉਣ ਲਈ ਲਾਗੂ ਕਰਨਾ ਚਾਹੀਦਾ ਹੈ।

NCSC CAF ਫਰੇਮਵਰਕ
ਸਾਈਬਰ ਅਸੈਸਮੈਂਟ ਫਰੇਮਵਰਕ (CAF) ਮਹੱਤਵਪੂਰਨ ਸੇਵਾਵਾਂ ਅਤੇ ਗਤੀਵਿਧੀਆਂ ਲਈ ਜ਼ਿੰਮੇਵਾਰ ਸੰਸਥਾਵਾਂ ਲਈ ਮਾਰਗਦਰਸ਼ਨ ਪ੍ਰਦਾਨ ਕਰਦਾ ਹੈ।

MITRE ਦਾ ਸਪਲਾਈ ਚੇਨ ਸੁਰੱਖਿਆ ਫਰੇਮਵਰਕ
ਸਪਲਾਈ ਚੇਨ ਦੇ ਅੰਦਰ ਸਪਲਾਇਰਾਂ ਅਤੇ ਸੇਵਾ ਪ੍ਰਦਾਤਾਵਾਂ ਦਾ ਮੁਲਾਂਕਣ ਕਰਨ ਲਈ ਇੱਕ ਢਾਂਚਾ ਹੈ।

ਜ਼ੋਖਮ ਪ੍ਰਬੰਧਨ

NIST AI ਜ਼ੋਖਮ ਪ੍ਰਬੰਧਨ ਫਰੇਮਵਰਕ (AI RMF)
AI RMF ਇਹ ਦੱਸਦਾ ਹੈ ਕਿ AI ਨਾਲ ਵਿਲੱਖਣ ਤੌਰ 'ਤੇ ਜੁੜੇ ਵਿਅਕਤੀਆਂ, ਸੰਸਥਾਵਾਂ ਅਤੇ ਸਮਾਜ ਲਈ ਸਮਾਜਿਕ-ਤਕਨੀਕੀ ਜ਼ੋਖਮਾਂ ਦਾ ਪ੍ਰਬੰਧਨ ਕਿਵੇਂ ਕਰਨਾ ਹੈ।

ISO 27001: ਸੂਚਨਾ ਸੁਰੱਖਿਆ, ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਅਤੇ ਗੁਪਤਤਾ ਸੁਰੱਖਿਆ
ਇਹ ਮਿਆਰ ਸੰਸਥਾਵਾਂ ਨੂੰ ਸੂਚਨਾ ਸੁਰੱਖਿਆ ਪ੍ਰਬੰਧਨ ਪ੍ਰਣਾਲੀ ਦੀ ਸਥਾਪਨਾ ਕਰਨ, ਲਾਗੂ ਕਰਨ ਅਤੇ ਰੱਖ-ਰਖਾਅ ਬਾਰੇ ਮਾਰਗਦਰਸ਼ਨ ਪ੍ਰਦਾਨ ਕਰਦਾ ਹੈ।

ISO 31000: ਜ਼ੋਖਮ ਪ੍ਰਬੰਧਨ
ਇੱਕ ਅੰਤਰਰਾਸ਼ਟਰੀ ਮਿਆਰ ਜੋ ਸੰਸਥਾਵਾਂ ਨੂੰ ਸੰਸਥਾਵਾਂ ਵਿੱਚ ਜ਼ੋਖਮ ਪ੍ਰਬੰਧਨ ਲਈ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ ਅਤੇ ਸਿਧਾਂਤ ਪ੍ਰਦਾਨ ਕਰਦਾ ਹੈ।

NCSC ਜ਼ੋਖਮ ਪ੍ਰਬੰਧਨ ਮਾਰਗਦਰਸ਼ਨ
ਇਹ ਮਾਰਗਦਰਸ਼ਨ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਜ਼ੋਖਮ ਪ੍ਰੈਕਟੀਸ਼ਨਰਾਂ ਨੂੰ ਉਹਨਾਂ ਦੀਆਂ ਸੰਸਥਾਵਾਂ ਨੂੰ ਪ੍ਰਭਾਵਿਤ ਕਰਨ ਵਾਲੇ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਜ਼ੋਖਮਾਂ ਨੂੰ ਬਿਹਤਰ ਢੰਗ ਨਾਲ ਸਮਝਣ ਅਤੇ ਪ੍ਰਬੰਧਨ ਕਰਨ ਵਿੱਚ ਮੱਦਦ ਕਰਦਾ ਹੈ।

ਨੋਟਸ

1. ਇੱਥੇ ਇਸਨੂੰ ਇੱਕ ਅਜਿਹੇ ਵਿਅਕਤੀ, ਜਨਤਕ ਅਥਾਰਟੀ, ਏਜੰਸੀ ਜਾਂ ਹੋਰ ਸੰਸਥਾ ਵਜੋਂ ਪਰਿਭਾਸ਼ਿਤ ਕੀਤਾ ਗਿਆ ਹੈ ਜੋ AI ਪ੍ਰਣਾਲੀ ਵਿਕਸਤ ਕਰਦਾ ਹੈ (ਜਾਂ ਜਿਸ ਕੋਲ ਇੱਕ AI ਪ੍ਰਣਾਲੀ ਵਿਕਸਤ ਕੀਤੀ ਹੋਈ ਹੈ) ਅਤੇ ਉਸ ਪ੍ਰਣਾਲੀ ਨੂੰ ਮਾਰਕੀਟ ਵਿੱਚ ਪੇਸ਼ ਕਰਦਾ ਹੈ ਜਾਂ ਇਸਨੂੰ ਆਪਣੇ ਨਾਮ ਜਾਂ ਟ੍ਰੇਡਮਾਰਕ ਦੇ ਅਧੀਨ ਸੇਵਾ ਲਈ ਜਾਰੀ ਕਰਦਾ ਹੈ

2. ਡਿਜ਼ਾਈਨ ਦੁਆਰਾ ਸੁਰੱਖਿਅਤ ਬਾਰੇ ਹੋਰ ਜਾਣਕਾਰੀ ਲਈ, CISA ਦਾ ਡਿਜ਼ਾਇਨ ਦੁਆਰਾ ਸੁਰੱਖਿਅਤ ਵੈੱਬ ਪੇਜ ਅਤੇ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਜ਼ੋਖਮ ਦੇ ਸੰਤੁਲਨ ਨੂੰ ਬਦਲ ਰਿਹਾ ਹੈ: ਡਿਜ਼ਾਈਨ ਸਾਫ਼ਟਵੇਅਰ ਦੁਆਰਾ ਸੁਰੱਖਿਅਤ ਕਰਨ ਲਈ ਸਿਧਾਂਤ ਅਤੇ ਪਹੁੰਚਾਂ

3. ML AI ਤੋਂ ਬਗ਼ੈਰ ਪਹੁੰਚਾਂ ਜਿਵੇਂ ਕਿ ਨਿਯਮ-ਅਧਾਰਿਤ ਪ੍ਰਣਾਲੀਆਂ ਦੇ ਉਲਟ

4. CEPS ਨੇ ਆਪਣੇ ਪ੍ਰਕਾਸ਼ਨ 'EU ਦੇ ਆਰਟੀਫੀਸ਼ੀਅਲ ਇੰਟੈਲੀਜੈਂਸ ਐਕਟ ਨਾਲ AI ਵੈਲਿਊ ਚੇਨ ਦਾ ਮੇਲ ਕਰਨਾ' ਵਿੱਚ ਸੱਤ ਵੱਖ-ਵੱਖ ਕਿਸਮਾਂ ਦੇ AI ਵਿਕਾਸ ਪਰਸਪਰ ਪ੍ਰਭਾਵਾਂ ਦਾ ਵਰਣਨ ਕੀਤਾ ਹੈ।

5. ISO/IEC 22989:2022(en) ਇਸਨੂੰ 'ਇੱਕ ਕਾਰਜਸ਼ੀਲ ਤੱਤ ਜੋ ਇੱਕ AI ਪ੍ਰਣਾਲੀ ਬਣਾਉਂਦਾ ਹੈ' ਵਜੋਂ ਪਰਿਭਾਸ਼ਿਤ ਕਰਦਾ ਹੈ

6. NIST ਨੂੰ ਆਰਟੀਫੀਸ਼ੀਅਲ ਇੰਟੈਲੀਜੈਂਸ (AI) ਦੇ ਸੁਰੱਖਿਅਤ, ਮਹਿਫੂਜ਼, ਅਤੇ ਭਰੋਸੇਮੰਦ ਵਿਕਾਸ ਅਤੇ ਵਰਤੋਂ ਨੂੰ ਅੱਗੇ ਵਧਾਉਣ ਲਈ ਦਿਸ਼ਾ-ਨਿਰਦੇਸ਼ ਤਿਆਰ ਕਰਨ (ਅਤੇ ਹੋਰ ਕਾਰਵਾਈਆਂ ਕਰਨ) ਦਾ ਕੰਮ ਸੌਂਪਿਆ ਗਿਆ ਹੈ। 30 ਅਕਤੂਬਰ, 2023 ਦੇ ਕਾਰਜਕਾਰੀ ਆਦੇਸ਼ ਦੇ ਤਹਿਤ NIST ਦੀਆਂ ਜ਼ਿੰਮੇਵਾਰੀਆਂ ਦੇਖੋ

7. ਖ਼ਤਰੇ ਵਾਲੀ ਮਾਡਲਿੰਗ ਬਾਰੇ ਹੋਰ ਜਾਣਕਾਰੀ OWASP ਫਾਉਂਡੇਸ਼ਨ ਤੋਂ ਉਪਲਬਧ ਹੈ

8. MITRE ATLAS ਵਿਰੋਧਮਈ ਮਸ਼ੀਨ ਲਰਨਿੰਗ 101 ਦੇਖੋ

9. GitHub: ਖ਼ਤਰਨਾਕ Lambda (ਲਾਂਬਡਾ) ਪਰਤ ਦੀ ਵਰਤੋਂ ਕਰਦੇ ਹੋਏ ਟੈਂਸਰਫਲੋਅ ਲਈ RCE PoC

10. SLSA: 'ਕਿਸੇ ਵੀ ਸਾਫ਼ਟਵੇਅਰ ਸਪਲਾਈ ਚੇਨ 'ਤੇ ਕਲਾਤਮਕ ਅਖੰਡਤਾ ਦੀ ਸੁਰੱਖਿਆ'

11. METI (ਜਾਪਾਨੀ ਆਰਥਿਕਤਾ, ਵਪਾਰ ਅਤੇ ਉਦਯੋਗ ਮੰਤਰਾਲਾ, 2023), 'ਸਾਫ਼ਟਵੇਅਰ ਪ੍ਰਬੰਧਨ ਲਈ ਸਾਫ਼ਟਵੇਅਰ ਬਿਲ ਆਫ਼ ਮਟੀਰੀਅਲਜ਼ (SBOM) ਦੀ ਸ਼ੁਰੂਆਤ ਬਾਰੇ ਗਾਈਡ'

12. ਗੂਗਲ ਖੋਜ: ਮਸ਼ੀਨ ਲਰਨਿੰਗ: ਤਕਨੀਕੀ ਕਰਜ਼ੇ ਦਾ ਵੱਧ ਵਿਆਜ਼ ਵਾਲਾ ਕ੍ਰੈਡਿਟ ਕਾਰਡ

13. Tramèr et al 2016, ਪੂਰਵ-ਅਨੁਮਾਨ API ਦੁਆਰਾ ਮਸ਼ੀਨ ਲਰਨਿੰਗ ਮਾਡਲਾਂ ਦੀ ਚੋਰੀ

14. ਬੋਏਨਿਸ਼, 2020, ਮਸ਼ੀਨ ਲਰਨਿੰਗ ਗੁਪਤਤਾ ਪ੍ਰਤੀ ਹਮਲੇ (ਭਾਗ 1): IBM-ART ਫਰੇਮਵਰਕ ਨਾਲ ਮਾਡਲ ਨੂੰ ਉਲਟਣ ਵਾਲੇ ਹਮਲੇ

15. ਰਾਸ਼ਟਰੀ ਸਾਈਬਰ ਸੁਰੱਖਿਆ ਕੇਂਦਰ, 2020, ਨਿੱਜੀ ਤੌਰ 'ਤੇ ਆਯੋਜਿਤ ਕੀਤਾ ਜਾਂਦਾ Public Key Infrastructure (ਪਬਲਿਕ ਕੀ ਇਨਫ੍ਰਾਸਟ੍ਰਕਚਰ, PKI) ਡਿਜ਼ਾਇਨ ਕੀਤਾ ਅਤੇ ਬਣਾਇਆ