Product Guideline

Kinsmen AI
AI Governance
& Impact Assessment

Date: May 30, 2024

KINSMEN AI

Table of Contents

Introduction

This is a structured guide covering topics relating to an AI Impact Assessment for the Kinsmen AI application. This guide covers essential areas such as safety, governance, and the overall impact of AI within the application.

Kinsmen AI leverages advanced AI technologies including Microsoft Cognitive AI, Visual Search services, and Large Language Models (LLMs) such as OpenAI's GPT-4o. This guide is designed to help customers understand the safety, governance, and overall impact of integrating Kinsmen AI into their work environment.

1. Purpose and Scope

This assessment illustrates how the deployment and usage of Kinsmen AI is safe, ethical, and compliant with regulatory standards. It covers the impact on data privacy, security, ethical considerations, and governance related to the deployment of Kinsmen AI.

2. Data Privacy and Security

2.1 Data Segregation

Azure CosmosDB	Each customer has a segregated CosmosDB instance ensuring data isolation and security.
Data Encryption	Data is encrypted both in transit and at rest to protect sensitive information.

2.2 Access Control

Authentication	Robust authentication mechanisms (e.g., multi-factor authentication) are in place, managed through Microsoft Entra ID
Authorization	Role-based access control (RBAC) ensures that only authorized personnel can access specific data and functionality.

2.3 Data Handling

Ingestion	Only authorized files and data sources are ingested
Processing	All data processing adheres to compliance standards, including GDPR and CCPA

3. AI Model Governance

3.1 Model Transparency

Source Indication	The chat system clearly indicates whether a response is generated from internal data or the external LLM (GPT-4o)
Explainability	AI model decisions are documented, and explanations are provided to users where possible.

3.2 Bias Mitigation

Bias Detection	Regular audits are conducted to detect and mitigate biases in AI model
Fairness	Models are trained on diverse data sets to ensure fairness and inclusivity

4. Ethical Considerations

4.1 User Consent and Responsible Usage

Informed Consent	Users are informed about data usage and AI functionality before interacting with the system.
Opt-Out Mechanism	Users have the option to opt out of data collection and AI interactions.
Human-in-the-Loop	Critical decisions are validated by human experts to ensure accuracy and accountability
Usage Guidelines	Clear guidelines are provided to users on responsible AI usage.

5. Risk Management

5.1 Risk Identification

Threat Modeling	Potential threats and vulnerabilities are identified through regular threat modeling exercises.
Impact Analysis	The impact of identified risks is analyzed and documented.

5.2 Mitigation Strategies

Incident Response	A well-defined incident response plan is in place to address any AI-related incidents promptly.
Continuous Monitoring	AI models and systems are continuously monitored for performance and compliance.

6. Compliance and Audits

Legal Adherence	The system complies with all relevant legal and regulatory requirements.
Data Protection Impact Assessments (DPIAs)	DPIAs are conducted regularly to assess data protection measures
Audit Trails	Where possible, logs are maintained for AI activities for audit purposes
Regular Audits	Internal audits are conducted to ensure adherence to governance policies and procedures.

7. User Training and Support

AI Literacy	Training programs are provided to enhance users' understanding of AI technologies and their implications
Usage Training	Hands-on training sessions are offered to help users effectively utilize Kinsmen AI.
Helpdesk	A dedicated helpdesk is available for users to seek assistance.
Documentation	Comprehensive documentation is provided, covering all aspects of Kinsmen AI.

Conclusion

The successful deployment and usage of Kinsmen AI requires a thorough understanding of its impact on data privacy, security, and ethical considerations. This guide provides a comprehensive framework to help customers navigate these critical areas, ensuring a safe and responsible AI experience.