SPATIAL

AI SYSTEM DESIGN

DESIGN PATTERNS DESIGN PRINCIPLES

ABREVIATED VERSION

DESIGN PATTERNS

POST-HOC AI-INSIGHT PATTERN

Utilizes decoupled architecture to provide explanations from xAI methods without exposing internal model details, enhancing compatibility and security. Facilitates integration of explainable AI (xAI) methods across models, ensuring security and flexibility to achieve better model transparency.

PROPERTY EXPOSURE MINIMIZATION PATTERN

Implements strategies to restrict detailed model outputs, preventing privacy breaches and limiting adversarial attack opportunities.

Limits sensitive data exposure in model outputs to enhance privacy and prevent data breaches.

API GATEWAY PATTERN

Acts as a unified entry point for client requests, routing them to appropriate microservices while maintaining security and efficiency.

Centralizes client requests, simplifies access, and enforces security in microservices to achieve efficient system management.

LOAD BALANCING PATTERN

Uses algorithms to allocate incoming requests among multiple servers, optimizing resource use and maintaining system performance.

Distributes traffic efficiently across servers to ensure high performance and fault tolerance, achieving optimal resource utilization.

UNIFIED AND USER-ORIENTED EXPLANATIONS PATTERN

Defines standardized explanation formats for different user groups, ensuring explanations are clear, consistent, and tailored to users' needs.

Standardizes explanation outputs for better interpretability across user groups to achieve clear and consistent communication.

PRIVACY-FOCUSED ML TRAINING PATTERN

Employs federated learning to train models collaboratively without exchanging raw data, preserving privacy and security.

Enables ML training without sharing raw data, ensuring data privacy and security to achieve collaborative model building.

TRUSTED EXECUTION ENVIRONMENT COMPUTING PATTERN

Utilizes trusted execution environments (TEEs) to run computations securely, with remote attestation to verify integrity and confidentiality.

Provides secure, transparent computing with minimal user impact to achieve confidential computing support.

ADVERSARIAL TRAINING PATTERN

Integrates adversarial examples into the training process, enhancing the model's ability to resist evasion attacks. Trains models with adversarial examples to improve robustness against attacks, achieving higher model security.

LABEL SANITIZATION PATTERN

Removes or modifies incorrect, irrelevant, or sensitive data to improve data quality and protect privacy.

Cleans and corrects data to ensure integrity and compliance with privacy regulations, achieving higher data quality.

ENHANCED INTERPRETABILITY PATTERN

Provides users with interactive, customizable explanations to enhance the clarity and usability of XAI outputs.

Customizes XAI outputs for better user understanding and interaction, achieving greater user engagement.

ITERATIVE RESILIENCE IMPROVEMENT AGAINST EVASION ATTACK PATTERN

Involves regular vulnerability assessments and targeted enhancements to strengthen model resilience against evasion attacks.

Continuously improves ML model resilience through iterative assessments to achieve robust defense mechanisms.

DESIGN PRINCIPLES

ENSURE HIGH-QUALITY AND ACCURATE DATA

Ensure the data used for AI is diverse, unbiased, and well-managed to make AI systems accurate and fair.

This involves collecting data from various sources, regularly cleaning and checking it for errors, and maintaining transparency and compliance with regulations.

The data used for training and evaluating AI models must be representative of the real-world scenario the AI system aims to address.

This requires collecting diverse data that encompasses various demographics, geographies, and relevant contextual factors. Additionally, it is essential to ensure the data is unbiased which could introduce unfairness or inaccuracies into the AI models.

Furthermore, implementing robust data management practices is crucial for maintaining data integrity. This involves data cleansing where inconsistencies and inaccuracies in the data are corrected and may involve data integration from different sources to capture the full context.

All transformations and aggregations of raw data must be documented to ensure transparency and reproducibility. Data must be managed consistently and in compliance with regulations and organizational requirements.

Regular monitoring and assessment of the quality of data is essential to identify and correct issues such as inconsistencies and biases.

GUARANTEE DATA PRIVACY AND SECURITY

Protect data at all stages — when stored, sent, or used. Encrypt data, control access strictly, and isolate data processing to prevent attacks.

Ensure systems and tools are secure, allow users to control their data, and monitor all interactions with the AI system to maintain security and privacy.

This principle mandates privacy and security groundworks for data at rest, in transit, and in use. All the sensitive data must be removed in preprocessing procedures.

Access to the data in multiparty computations must respect the least privileged approach. Data at rest and in transit must be encrypted and all the data processing must happen in controlled and isolated environments to reduce potential attack frames.

The AI system must be resilient to attacks targeted at the model, data, or the system itself. This principle also demands all the development tooling (such as libraries) and deployment platforms to pass security audits.

Users must be in control of their sensitive data and the system must be able to permanently remove personal if required by data owners.

Trained models should be as isolated as possible and all the interactions with the model and the platform monitored.

ACHIEVE ROBUST AI MODEL PERFORMANCE

Ensure AI models are highly accurate, can adapt to new data, and perform well under different conditions.

They should handle varying amounts of data and devices, resist attacks through strong testing, and provide reliable evidence of their performance and use.

This principle mandates the provision of high accuracy in the AI model's predictions, emphasizing that the model must generalize well to unseen data for robustness and optimal performance.

The principle underscores the importance of the model's ability to adapt, ensuring that it maintains its level of performance under various circumstances. It advocates for scalability, asserting that AI models should handle varying data volumes, devices, and services effectively.

Additionally, the principle demands that AI models must be resilient against adversarial ML attacks, including evasion and poisoning, through rigorous testing, threat modelling, and security monitoring processes.

It promotes extensive testing using well-defined performance metrics to ensure consistency in diverse environments.

Lastly, the principle highlights the necessity for AI models to provide objective evidence of fulfilling specific intended use, ensuring reliability and accountability within a given time interval and under defined conditions.

ELIMINATE BIAS, UPHOLD FAIRNESS

Make AI systems fair and unbiased by using diverse data and special methods to prevent discrimination.

Regularly evaluate and adjust the data and algorithms to identify and fix biases, ensuring AI decisions are fair and inclusive for all groups.

The Fairness and Bias Mitigation principle emphasizes the importance of ensuring that AI systems are designed and trained to be fair and unbiased.

Bias can emerge in AI systems due to biased training data or flawed algorithms, leading to discriminatory outcomes. Addressing this principle involves identifying and mitigating biases related to race, gender, ethnicity, religion, sexual orientation, and other protected characteristics.

To implement ethical and fair AI practices, developers adopt a multifaceted approach. Firstly, they prioritize diverse and inclusive datasets, actively seeking perspectives from different demographic groups to diminish biases and promote fair outcomes. Rigorous evaluation methods, including statistical analysis, stakeholder feedback, and third-party audits, are employed to identify and rectify biases within both training data and algorithms.

During algorithm development, techniques such as re-sampling, re-weighting, and adversarial training are applied to mitigate biases, with a focus on designing fairness-aware machine learning algorithms that minimize disparate impact across diverse groups. Additionally, the calibration of model predictions, considering uncertainty and avoiding reliance on predefined thresholds, is emphasized to enhance effective communication of results.

By embedding the values proposed by this principle, AI systems yield more equitable outcomes, as their decisions are not influenced by unfair biases, what increases stakeholders' trust on AI systems and contributes to social harmony reducing existing inequalities and promoting inclusivity.

MANDATE TRANSPARENCY AND EXPLAINABILITY

Provide clear explanations for how AI makes decisions and keep detailed documentation to build trust.

Ensure AI outputs are understandable and relevant to users, and enable human oversight to intervene when necessary. Maintain transparency to enhance reliability and performance.

This principle mandates the provision of clear and comprehensible explanations for AI decisions and operations, a critical factor in building trust and understanding among all users.

The developers must maintain thorough, accessible documentation covering all aspects of AI training, deployment, and operations, ensuring transparency and accountability.

The principle demands that AI systems are designed with user-centric interpretability, ensuring that their outputs and decisions are meaningful and relevant in real-world contexts. This is vital for enhancing user engagement and trust.

Furthermore, the inherent testability and verifiability of AI models are obligatory, underpinning their reliability and performance. Essential to this principle is the incorporation of robust mechanisms for human oversight, including intervention and override capabilities.

OPTIMIZE FOR USER-CENTRIC DESIGN

Design AI systems to be user-friendly and accessible, considering diverse needs and preferences.

Include features like dark mode and user-friendly interfaces, ensure seamless integration, and maintain transparency about AI system impacts.

This design principle encapsulates the essence of user accessibility and usability, fostering a harmonious interaction between users and AI-based systems.

This involves the incorporation of features such as a dark mode for user interface customization, accommodating diverse cultural and linguistic preferences, and ensuring user-friendly interfaces.

The adoption of a microservice architecture contributes to infrastructure resilience, while API specifications facilitate smooth integration with external services. Continuous Integration/Continuous Deployment (CI/CD) practices play a pivotal role in the deployment of microservices, ensuring that updates are seamlessly rolled out with minimal disruption to users.

Transparency regarding the health and psychological impact of AI system usage is crucial, and the design should incorporate mechanisms to communicate these aspects to users openly.

Privacy measures, when correctly implemented, not only safeguard user data but also contribute to user comfort and trust in the AI system. The requirement for fast response times aligns with the overall goal of good performance, enhancing the system’s usability.

COMPLY WITH LEGAL AND ETHICAL STANDARDS

Ensure AI systems follow legal regulations and ethical standards to protect individual rights and minimize risks.

Maintain thorough documentation and transparency, manage risks effectively, and provide human oversight to build trust and accountability in AI systems.

This principle mandates adherence to prevailing legal frameworks for AI systems, specifically the General Data Protection Regulation and the AI act. This principle encompasses processes, documentation, and safeguards to protect individual rights and mitigate potential risks.

The GDPR demands adherence to data protection principles and individuals’ rights, such as providing accessible information and responding to data access and removal requests. The AI Act addresses other aspects of AI systems, encompassing risk management, testing processes, data governance, transparency, and human oversight.

This principle mandates comprehensive documentation, including system architecture and accountability details, fostering transparency and traceability. Furthermore, the principle underscores the importance of accountability in AI systems, necessitating clear documentation of the ML solution’s purpose, inherent risks, and non-functional requirements.

By embedding these legal and regulatory considerations into the design process, this principle establishes a responsible foundation, fostering trust, transparency, and the protection of fundamental human rights in the development and deployment of AI systems.