AIDX TECH Profile

AIDX TECH focuses on providing customers with high-quality AI model testing services, offering a comprehensive testing platform that covers aspects such as robustness, fairness, interpretability, and security of models. In addition, AIDX TECH also provides testing for AI-generated content (AIGC) and large language models (LLM), helping customers to fully understand the performance and potential risks of their AI models.

Course participants

Participants	Why attend this course
Government agencies/ AI auditors/ Regulation	Government officials and regulatory staff seek Trusted AI training to regulate and manage AI development effectively, aligning policies with societal interests. This training empowers them to oversee AI applications and develop corresponding regulations responsibly.
Industries and organizations	Senior executives and decision-makers require guidance on integrating AI into business operations while upholding ethical standards, regulations, and corporate values. AI developers, data scientists, and engineers need expertise in designing systems that comply with ethical and regulatory guidelines.
Education/University/ Research institutions/ Training centres	Offer Trusted AI training to students, faculty, and researchers to deepen their understanding of AI's ethical and legal dimensions. This training fosters a culture of ethical reflection and social awareness in AI research and teaching, guiding participants to integrate these principles into their work and practical applications.

Speaker Introduction

We invite experts relevant to the field based on client needs. Our current special guests include experts from Singaporean government agencies, AI-related associations, representatives of standards committee, and university professors, lecturers, testing organization experts, and senior industrial AI engineers.

Course Details

AIDX TECH offers an optional 30-minute introductory course. Based on client needs, we offer customized course content, with both scheduling and instructor selection adaptable to the specific requirements of the client's industry.

	Security Management and Governance of Trusted AI Technologies	Application, Security, and Regulation of Trusted AI Technologies
Target Audience	Corporate Executives, Managers, Government Officials, Regulatory Agency Members	Engineers, Programmers, Product Managers, Risk Analysts, Professionals from specialized institutions, and individuals with a foundation in AI from AI research institutes
Certification	Certified Trustworthy AI manager	Certified Professional Trustworthy AI Engineer
Course Duration	8-hr	16-hr
Content Overview	**Introduction to AI Basic Concepts:** – Fundamentals and significance of trustworthy AI. – Relationship between AI safety, ethics, and regulations. **Laws and Regulations:** – Overview of domestic and international AI regulations. – Introduction of EU AI ACT – Introduction of US NIST AI Risk Management Framework – Analysis of compliance issues. – Policy formulation and implementation. **Global AI system lifecycle standards:** – Overview AI standards – Introduction of ISO SC 42 - Artificial intelligence – Overview of AI trustworthiness (ISO/IEC 24027, ISO/IEC 23894) – Highlight of AI Governance (ISO/IEC 38507) **AI Data and Model Security:** – Data privacy and model security importance. – Legal frameworks and privacy protection. – Risk management for security. **Transparency and Explainability:** – AI algorithm transparency. – Model explainability **Fairness and Bias:** – Understanding biases in AI. – Application of fairness principles. – Strategies for reducing bias. **Supervision and Audit:** – Designing mechanisms for compliance. – Implementing supervision and audit. – Best practices for tools and techniques.	**Foundations of Trustworthy AI** – Review of machine learning and deep learning – Overview of Trusted AI principles – Common AI application scenarios and case studies **Data Quality and Preprocessing** – AI data cleaning, handling missing values, and outlier detection – Data quality for trustworthy AI models **Model Selection, Application and Evaluation** – Machine learning and deep learning models – Large language model (LLM) – AI model fine tuning – Hardware introduction (e.g. GPU) – Evaluation metrics and techniques **Robustness and Data Security of AI Models** – Robustness metrics – Privacy protection technologies: data encryption, differential privacy, secure multi-party computation – Attack and defence (e.g. Adversarial attacks, Backdoor attack, Data poisoning) – AI model and data copyright **Explanability Technologies** – Local and global explanation methods – Feature importance analysis and model interpretability techniques – Explainability vs. Accuracy Trade-off **Fairness and Bias Identification** – Fairness metrics – Discrimination identification – Sensitive attribute recognition – Bias mitigation **AI Standards and Oversight Mechanisms** – Designing oversight and audit mechanisms – Application of oversight tools (model monitoring, anomaly detection, automated testing) **Case Studies and Practical Projects** – Analysis of real-world Trusted AI cases – Hands-on projects (e.g. Attack an AI system)

Contact Us

Email: sales@aidxtech.com

WhatsApp: +65 9392 7685 (Gaven Guo)