# **Fawzooz AI - Artificial Intelligence Management System (AIMS) Manual**

Document ID: AIMS-MAN-001

Version: 1.0

Date: May 19, 2025

Classification: Fawzooz AI Confidential

## **Part 1: AIMS Governance and Foundation**

### **Section 1: Introduction and Overview**

#### **1.1 Purpose and Scope of the AIMS Manual**

This manual establishes the formal Artificial Intelligence Management System (AIMS) for Fawzooz AI. It is designed and implemented in accordance with the requirements of the international standard ISO/IEC 42001:2023 and serves as the central governing document for all Artificial Intelligence (AI) related activities, systems, and processes within the organization.1 The primary purpose of this manual is to provide a comprehensive, centralized reference for the policies, procedures, controls, and objectives that constitute the AIMS. It serves as the definitive blueprint for establishing, implementing, maintaining, and continually improving the AIMS, ensuring that responsibility is embedded into the very fabric of our AI initiatives.1

The AIMS, as detailed herein, ensures that all AI systems developed, provided, or used by Fawzooz AI are aligned with our core strategic objectives, meet all applicable legal, regulatory, and ethical requirements, and proactively address the expectations of our users, clinicians, partners, and other stakeholders.1 It translates high-level principles such as fairness, transparency, and accountability into actionable operational procedures, assigning clear responsibilities and systematizing the management of AI-specific risks and impacts.1 This document provides the framework to guide Fawzooz AI in harnessing the transformative potential of AI responsibly, fostering innovation while building and maintaining trust.

#### **1.2 Fawzooz AI's Commitment to Responsible AI in Mental Wellness**

Fawzooz AI was founded on the principle that technology can be a profound force for good in mental wellness. We recognize the immense responsibility that accompanies the application of AI in such a sensitive and personal domain. Our unwavering commitment is to harness the power of AI to improve lives, expand access to care, and provide effective support, all while upholding the highest possible standards of ethics, clinical safety, privacy, and trustworthiness.1 This commitment is not an adjunct to our business strategy; it is the foundation of our mission and is reflected in every aspect of this AIMS.

We strive to achieve a delicate but critical balance between pioneering innovation and robust governance. This means ensuring our applications are not only technologically advanced and clinically effective but are also demonstrably fair, transparent, and respectful of human dignity, privacy, and autonomy.1 The principles and procedures outlined in this manual represent our pledge to our users, their families, the clinical community, and society at large. We are dedicated to building AI systems that are worthy of the profound trust placed in us, transforming risk into an opportunity to create resilient, trustworthy, and ethically sound AI-driven futures for mental wellness.1

### **Section 2: Context of the Organization (Clause 4)**

#### **2.1 Understanding Fawzooz AI's Internal and External Context**

The establishment of an effective AIMS necessitates a deep and nuanced understanding of the unique environment in which Fawzooz AI operates. This context is shaped by a combination of internal factors that define our organization and external forces that influence our industry.1

Internal Issues:

Our internal landscape is defined by our specialized focus and structure. As both a developer and provider of AI systems for mental wellness, our AIMS must govern the entire lifecycle from conception to delivery.1 Key internal issues include:

* **Organizational Structure:** Fawzooz AI maintains a dedicated, cross-functional governance structure that includes an AI Governance Committee, a Chief Information Security Officer (CISO), and a Chief AI Ethics Officer (CAIO). This structure ensures that security and ethical considerations are integrated at every level of decision-making.
* **Strategic Vision:** Our corporate strategy is to be the global leader in *ethical* mental wellness AI. This strategic imperative drives our investment in robust governance and positions the AIMS not as a compliance cost but as a core competitive differentiator.
* **Technical Capabilities:** Our expertise lies in advanced Natural Language Processing (NLP), sentiment analysis, and predictive analytics tailored for clinical application. The AIMS must govern the responsible use of these powerful technologies.
* **Corporate Culture:** We foster a "responsibility-first" culture that prioritizes clinical safety and user trust over speed to market, a principle that is reinforced through our training and performance management systems.

External Issues:

We operate within a complex and highly regulated global environment. The AIMS is designed to be responsive to these critical external factors:

* **Regulatory Landscape:** Our applications are subject to a multi-jurisdictional legal framework. This requires strict adherence to the Health Insurance Portability and Accountability Act (HIPAA) in the United States, which governs the use of Protected Health Information (PHI) 3; the General Data Protection Regulation (GDPR) in the European Union, with its stringent requirements for consent, data subject rights, and processing of sensitive health data 6; and the data protection laws of the United Arab Emirates (UAE), including the Personal Data Protection Law (PDPL) and the ICT Health Law.9
* **Data Localization:** A critical external factor is the data localization provision within the UAE's ICT Health Law (Federal Law No. 2 of 2019), which mandates that health data related to services provided in the UAE cannot be stored, processed, or transferred outside the country without specific approval from health authorities.11 This legal requirement has profound implications for our global IT architecture. It precludes the use of a single, centralized cloud infrastructure for all users and necessitates a distributed architecture with a dedicated, compliant data center or cloud region within the UAE for our Emirati users. The AIMS must therefore include specific operational controls to enforce this jurisdictional data segregation, making our governance framework a direct driver of core architectural and investment decisions.
* **Societal Expectations:** There is heightened public and user sensitivity regarding privacy, confidentiality, and fairness in mental health technology. Maintaining public trust is our license to operate, and the AIMS is our primary mechanism for earning and keeping that trust.12
* **Market and Technology:** The AI market is characterized by rapid innovation. Our AIMS must balance the competitive pressure to adopt new technologies, such as advanced Large Language Models (LLMs), with the non-negotiable requirement for rigorous clinical validation, safety testing, and ethical vetting before deployment in a mental health context.14

#### **2.2 Identifying Interested Parties and Their Requirements**

The Fawzooz AI AIMS is designed to be responsive to the needs and expectations of our diverse stakeholders. A thorough analysis of these interested parties is foundational to defining appropriate and effective controls.1

* **Patients/Users:** This is our most critical stakeholder group. Their primary expectations are for absolute confidentiality and privacy of their sensitive mental health data, clinical safety and efficacy of the AI tools, and fairness in how the AI interacts with them. They have a right to transparent information about how AI is used in their care and how their data is protected, and they must be able to provide meaningful, informed consent.2 The AIMS must ensure these rights and expectations are met through robust security controls, ethical design principles, and clear communication.
* **Clinicians/Therapists:** These professional users require AI tools that are accurate, reliable, and serve to augment—not replace—their professional judgment. They need absolute assurance of data security to maintain patient confidentiality and clear, validated protocols for using AI-assisted diagnostic insights or crisis intervention alerts. The AIMS must provide this assurance and clarity to foster clinical adoption and trust.2
* **Regulators:** Governmental and regulatory bodies, such as the U.S. Department of Health and Human Services (HHS), European Data Protection Authorities (DPAs), and the UAE's Telecommunications and Digital Government Regulatory Authority (TDRA), demand demonstrable compliance with all applicable laws. They expect robust risk management frameworks, evidence of Data Protection Impact Assessments (DPIAs), secure data handling, and timely breach notification protocols.4 The AIMS is our primary vehicle for demonstrating this compliance.
* **Investors and Leadership:** Our investors and executive team expect responsible innovation that builds sustainable, long-term trust and value. They rely on the AIMS to mitigate the significant reputational, legal, and financial risks associated with operating in the high-stakes mental wellness sector.1
* **The User-Clinician Dynamic:** The stakeholder ecosystem in mental wellness is uniquely complex, as the needs of the patient (user) and the clinician (customer) can sometimes diverge. For instance, a patient may desire absolute privacy, while a clinician may have a legal or ethical duty to report a credible risk of harm to self or others, which requires access to certain data. The AIMS must be designed to manage this dynamic. It achieves this through controls such as granular, role-based access control (RBAC) that limits clinician access to the minimum necessary information, and a sophisticated, transparent informed consent process that clearly explains to the user the specific, limited circumstances under which their data might be shared with a human professional for safety reasons.2

#### **2.3 Defining the Scope and Boundaries of the AIMS**

The scope of the Fawzooz AI AIMS is comprehensive, ensuring that all AI-related activities are subject to consistent and robust governance.1

* **Organizational Scope:** The AIMS applies to all departments, functions, personnel, and processes involved in the lifecycle of Fawzooz AI's products and services. This includes, but is not limited to, the AI Research and Development department, Product Management, Engineering, Quality Assurance, IT Operations, Security, Legal, and Clinical Oversight teams.
* **System Scope:** The AIMS governs all AI systems developed and/or provided by Fawzooz AI. This explicitly includes:
  + Our AI-powered conversational agents (chatbots) for delivering cognitive behavioral therapy (CBT) and other therapeutic modalities.17
  + Our predictive analytics models for mood tracking, pattern analysis, and user wellness insights.
  + Our specialized AI model for the early detection of suicide risk from user-provided text.18
* **Lifecycle Scope:** The AIMS covers the entire AI system lifecycle: from initial conception, requirements gathering, and design; through data acquisition and preparation, model development, and validation; to deployment, operational monitoring, maintenance, and eventual decommissioning.1
* **Geographical and Data Scope:** The AIMS has a global geographical boundary, applying to all users regardless of location. It includes specific controls to address jurisdictional requirements, most notably the enforcement of data localization for UAE users in adherence to the ICT Health Law.11 The scope covers all data processed by our AI systems, with a particular focus on the stringent protection of PHI and other sensitive personal information.

### **Section 3: Leadership and Commitment (Clause 5)**

#### **3.1 Leadership's Role in Driving a Responsible AI Culture**

The successful implementation and sustained effectiveness of the AIMS at Fawzooz AI are fundamentally dependent on the visible and unwavering commitment of our top management. Leadership's role transcends passive endorsement; it is an active, driving force that embeds responsible AI into the organization's DNA.1 Top management demonstrates its commitment by:

* **Championing the AI Policy:** Ensuring the AI Policy is established, communicated, and understood as a primary directive for the organization.
* **Strategic Integration:** Ensuring the AIMS and its objectives are fully integrated with and support the overall strategic direction of Fawzooz AI.
* **Resource Allocation:** Authorizing and providing the necessary financial, human (e.g., skilled ethicists, security professionals), and technological resources required for the effective establishment, implementation, maintenance, and continual improvement of the AIMS.1
* **Active Participation:** Engaging directly in the AIMS through activities such as chairing the AI Governance Committee and leading management reviews.
* **Fostering Culture:** Setting the "tone at the top" by consistently prioritizing ethical considerations and clinical safety in all strategic decisions related to AI.

#### **3.2 The Fawzooz AI Policy for Responsible AI**

Policy Statement:

Fawzooz AI is committed to the responsible and ethical development, provision, and use of Artificial Intelligence to improve mental wellness globally. This policy establishes the framework for our AI Management System (AIMS), ensuring that all AI activities align with our core principles, meet stakeholder expectations, and comply with all applicable legal and regulatory requirements.

Core Principles:

Our AI initiatives are guided by the following non-negotiable principles, derived from established bioethics and leading AI ethics frameworks 2:

1. **Beneficence and Non-maleficence:** Our primary goal is to benefit our users and "do no harm." AI systems will be designed and validated to be safe and effective.
2. **Patient Autonomy:** We respect the right of individuals to make informed decisions about their care. We will provide transparency and require meaningful informed consent for the use of AI and the processing of personal data.
3. **Fairness and Equity:** We will actively work to identify and mitigate unfair bias in our AI systems to ensure equitable access and outcomes for all individuals, regardless of their background or demographic characteristics.
4. **Transparency and Explainability:** We will be transparent with users and clinicians about the capabilities and limitations of our AI systems. Where appropriate and feasible, we will provide explanations for AI-driven recommendations.
5. **Privacy and Security:** We will protect the confidentiality, integrity, and availability of all user data, particularly sensitive health information, through robust data governance and state-of-the-art security controls.
6. **Accountability and Human Oversight:** We will maintain clear lines of accountability for our AI systems and ensure that appropriate human oversight is integrated into our processes, especially in high-risk scenarios.

Commitments:

Fawzooz AI formally commits to:

* Satisfying all applicable requirements, including but not limited to ISO/IEC 42001:2023, HIPAA, GDPR, and the laws of the UAE.1
* Providing a framework for setting, monitoring, and achieving specific AI objectives that align with this policy.
* The continual improvement of the suitability, adequacy, and effectiveness of our AIMS.1

This policy applies to all employees, contractors, and partners of Fawzooz AI. It is owned by the CISO and approved by the CEO, and will be reviewed at least annually or upon significant changes to the organization's context.

#### **3.3 AIMS Roles, Responsibilities, and Authorities**

Clear and unambiguous assignment of roles, responsibilities, and authorities is critical to the operational integrity of the AIMS, particularly in our high-stakes environment where accountability is paramount.1

**Key Roles:**

* **Chief Executive Officer (CEO):** Ultimately accountable for the AIMS and for fostering a culture of responsible AI. Approves the AI Policy and chairs the Management Review.
* **Chief Information Security Officer (CISO):** Designated as the "owner" of the AIMS. Responsible for the overall implementation, maintenance, monitoring, and reporting on the performance of the AIMS.
* **Chief AI Ethics Officer (CAIO):** A dedicated, senior role responsible for providing ethical guidance, leading the AI System Impact Assessment process, chairing the AI Ethics Review Board (a subcommittee of the Governance Committee), and developing ethics training.
* **AI Governance Committee:** A cross-functional steering body comprising the CISO, CAIO, General Counsel, Clinical Lead, and Head of AI Development. It is responsible for reviewing and approving high-risk AI projects, AIAs, and the AI Risk Treatment Plan.
* **AI System Owners:** Product Managers or Business Unit Leaders who are accountable for the responsible operation of specific AI systems throughout their lifecycle, including commissioning risk assessments and ensuring alignment with the AI Policy.
* **Clinical Lead:** A licensed mental health professional responsible for providing clinical oversight, validating the safety and efficacy of AI interventions, and defining crisis escalation protocols.
* **AI Development Team:** Responsible for the technical design, development, and testing of AI systems in accordance with the AIMS requirements.

To ensure absolute clarity, responsibilities are formally documented in the following RACI (Responsible, Accountable, Consulted, Informed) matrix. This level of detail is essential for effective governance and rapid, coordinated action, especially during a crisis. For example, in the event of an AI-related incident, the matrix clearly defines that the CISO is ultimately *Accountable* for the response, while the AI Development Team is *Responsible* for the technical fix and the Legal team must be *Consulted* on disclosure obligations. This prevents ambiguity and delay when time is critical.

| Activity/Deliverable | CEO | CISO | CAIO | AI Governance Committee | Legal Counsel | Clinical Lead | AI Dev Team Lead |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **AI Policy Approval & Review** | A | R | C | C | C | C | I |
| **AI Risk Assessment Execution** | I | A | C | C | C | C | R |
| **AI Risk Treatment Plan Approval** | I | C | C | A | C | C | I |
| **AI System Impact Assessment** | I | C | A | R | C | R | C |
| **AI Incident Response** | A | R | C | I | C | C | R |
| **AIMS Internal Audit** | I | A | I | I | I | I | R |
| **AIMS Management Review** | A | R | C | C | C | C | C |
| **AI Ethics Training Program** | I | C | A | C | I | R | R |

*Table 1: Fawzooz AI AIMS RACI Matrix*

## **Part 2: Strategic Planning for Responsible AI**

### **Section 4: Planning for the AIMS (Clause 6)**

#### **4.1 Actions to Address AI Risks and Opportunities**

In accordance with Clause 6.1 of ISO/IEC 42001:2023, Fawzooz AI will implement and maintain a formal, documented process to identify, assess, and treat risks and opportunities related to the AIMS and our AI systems. The objective of this process is to provide reasonable assurance that the AIMS can achieve its intended outcomes, prevent or reduce undesired effects, and achieve continual improvement.1 This strategic planning is co-developed and executed by a cross-functional team that includes not only technical experts but also clinical and ethical professionals, ensuring a holistic view of risk.

#### **4.2 The Fawzooz AI Risk Assessment Process for Mental Wellness Applications**

Our AI risk assessment methodology is specifically tailored to the unique context of mental wellness, extending far beyond standard IT or technical risks to address the profound clinical and ethical challenges inherent in our domain.1 A "system failure" in our context is not merely a technical bug; it represents a potential clinical crisis. Therefore, our risk assessment process is uniquely adapted.

Risk Identification:

The process begins with the systematic identification of potential risks across the entire AI lifecycle. Key risk categories include:

* **Clinical Safety Risks:** The risk of direct harm to users resulting from AI system failures. This includes inaccurate diagnosis or assessment 14, incorrect or inappropriate treatment recommendations 19, the failure of our AI to detect and escalate a crisis situation such as suicidal ideation 18, and the risk of "client abandonment" where a user in distress does not receive a timely and appropriate response.15
* **Ethical and Bias Risks:** The risk of our AI systems perpetuating or amplifying societal biases. This includes algorithmic bias leading to disparate performance or outcomes for certain demographic, racial, or ethnic groups 2; a lack of cultural sensitivity in language or therapeutic approach 12; the potential for AI interactions to erode patient autonomy or foster an unhealthy over-reliance on the application.14
* **Privacy and Security Risks:** The risk of unauthorized access to, or breach of, highly sensitive Protected Health Information (PHI). This includes the risk of re-identification of data that has been pseudonymized or anonymized, and failing to protect data in accordance with stringent legal standards.2
* **Legal and Compliance Risks:** The risk of non-compliance with the complex web of regulations governing our operations, including HIPAA, GDPR, and UAE data laws, which could result in severe financial penalties, operational sanctions, and loss of market access.3
* **Adversarial Use Risks:** The risk of "foreseeable misuse," where users may attempt to "jailbreak" or adversarially manipulate our AI chatbots to elicit harmful, inappropriate, or non-therapeutic content.

Risk Analysis and Evaluation:

Once identified, each risk is analyzed to determine its potential impact and likelihood. Our risk evaluation is distinguished by its use of a tailored AI Risk Matrix that incorporates dimensions beyond financial or operational loss. The "Impact" criteria in our matrix include specific, defined scales for "Clinical Harm," "Psychological Distress," and "Erosion of Trust," ensuring that the most significant human-centric risks are prioritized. The evaluation compares the calculated risk level against our pre-defined risk acceptance criteria, which are set by the AI Governance Committee with a very low tolerance for risks that could lead to user harm.

#### **4.3 The Fawzooz AI System Impact Assessment Process**

The AI System Impact Assessment (AIA) is a mandatory planning activity for all new AI systems and prior to any major changes to existing systems. It is a formal, documented process designed to proactively consider the potential consequences of our AI systems on individuals, groups, and society at large.1 The findings of the AIA are a primary input into the AI Risk Assessment process.

Our AIA methodology systematically evaluates potential impacts across several domains, with a special focus on vulnerable populations 1:

* **Impact on Fundamental Rights:** Assessing potential effects on dignity, autonomy, privacy, and non-discrimination.
* **Impact on Fairness:** Analyzing the potential for biased outcomes and ensuring equitable treatment.
* **Impact on Safety:** Evaluating risks of psychological or other forms of harm.
* **Impact on Societal Well-being:** Considering broader effects on mental health discourse, access to care, and community well-being.

The AIA process involves consultation with internal experts (clinical, legal, ethical) and, where appropriate, external stakeholders to ensure a comprehensive and multi-faceted perspective.

#### **4.4 AI Objectives and Planning to Achieve Them**

To ensure our AIMS is effective and drives tangible improvements, Fawzooz AI establishes SMART (Specific, Measurable, Achievable, Relevant, Time-bound) objectives at relevant functions and levels.1 These objectives translate the high-level principles of our AI Policy into concrete, trackable targets.

**Example AIMS Objectives for 2025-2026:**

* **Objective AIMS-OBJ-01 (Fairness):** "To reduce demographic bias in our depression-symptom prediction model by 15% within 12 months, as measured by the disparate impact ratio across defined racial and gender subgroups, to ensure more equitable performance." 12
* **Objective AIMS-OBJ-02 (Safety):** "To achieve a minimum 95% accuracy rate (recall) in identifying high-risk keywords and phrases within our suicide risk detection module, while maintaining a false positive escalation rate below 5%, by the end of Q4 2025. This will be validated against a curated, clinically-reviewed dataset." 18
* **Objective AIMS-OBJ-03 (Compliance):** "To ensure 100% of new AI systems processing personal data from EU or UAE users are designed and deployed on GDPR-compliant and in-country cloud infrastructure, respectively, from their initial launch date, as verified by pre-deployment compliance audits." 7
* **Objective AIMS-OBJ-04 (Competence):** "To train 100% of our clinical oversight staff in 'AI Literacy for Clinicians' by the end of Q2 2025, with competence measured by a 90% pass rate on a post-training assessment covering AI limitations and interpretation of AI-generated insights." 16

For each objective, a detailed plan will be created outlining the necessary actions, required resources, responsible personnel, and timelines for achievement. Progress towards these objectives will be monitored and reported as part of the AIMS performance evaluation.

### **Section 5: Support for the AIMS (Clause 7)**

#### **5.1 Provision of Resources**

Fawzooz AI's leadership is committed to providing the necessary and sufficient resources for the establishment, implementation, maintenance, and continual improvement of the AIMS. This includes allocating an adequate budget for governance tools and platforms, and dedicating skilled human resources, including protected time for personnel from security, ethics, legal, clinical, and engineering teams to participate in AIMS activities such as risk assessments and audits.1

#### **5.2 Competence, Training, and AI Literacy for Staff and Clinicians**

Given the specialized nature of our work, competence is a critical support function for the AIMS. Our competence program is tailored to the specific roles and responsibilities within the organization, going far beyond generic AI training to address the unique challenges of mental wellness technology.1

* **Competence Program for Technical Staff:** All data scientists, AI engineers, and software developers are required to complete a mandatory annual training curriculum. This program covers:
  + **Data Ethics and Bias Mitigation:** Training on how to identify and mitigate bias in clinical datasets and NLP models.12
  + **Privacy-Enhancing Technologies (PETs):** Practical training on implementing techniques like federated learning and differential privacy.
  + **Secure AI Development Lifecycle (ADLC):** Education on securing the entire AI pipeline, from data ingestion to model deployment, against AI-specific attacks.
  + **The Ethical Landscape of Mental Health:** An overview of the specific ethical challenges and responsibilities associated with developing technology for vulnerable populations.
* **Competence Program for Clinicians and Clinical Oversight Staff:** Our team of licensed therapists and clinical supervisors undergoes a specialized "AI Literacy" training program designed to ensure they can use our tools safely and effectively. This program is crucial for mitigating risks of misinterpretation and over-reliance.16 The curriculum includes:
  + **Understanding AI Capabilities and Limitations:** Clear education on what our AI can and cannot do, emphasizing that AI tools are for augmentation, not replacement, of professional judgment.2
  + **Interpreting AI-Generated Insights:** Training on how to critically evaluate AI-generated summaries, risk scores, or pattern analyses.
  + **Human Oversight Protocols:** Detailed instruction on their role in the human-in-the-loop process, especially the crisis intervention and escalation protocols.
* **Evidence of Competence:** Fawzooz AI will maintain documented information as evidence of competence. This includes training completion records, results from post-training knowledge assessments, and performance reviews that evaluate the application of these competencies in practice.1

#### **5.3 Awareness Programs for AI Risks and Responsibilities**

All Fawzooz AI personnel, regardless of their role, will participate in a mandatory annual AIMS awareness program.1 The program is designed to ensure that every employee understands:

* The principles and commitments of the Fawzooz AI Policy for Responsible AI.
* Their specific contribution to the effectiveness of the AIMS and the benefits of responsible AI practices.
* The severe implications of non-conformance with AIMS requirements, which include not only regulatory penalties and reputational damage but, most importantly, the potential for significant harm to our users.1

#### **5.4 Communication Protocols**

Fawzooz AI will establish and maintain clear, consistent, and transparent communication processes relevant to the AIMS.

* **Internal Communication:** Processes for communicating AIMS updates, policy changes, and incident reports to all relevant personnel.
* **External Communication:** Protocols for communicating with users and clinicians about how AI is used in their care, the data being processed, and their rights. This commitment to transparency is a cornerstone of building trust and obtaining informed consent.2 This also includes pre-defined plans for communicating with regulators and affected individuals in the event of an AI-related incident or data breach.

#### **5.5 Managing Documented Information**

All documented information required by the AIMS and by ISO/IEC 42001:2023 will be controlled in accordance with Clause 7.5 of the standard. This ensures that documentation is accurate, up-to-date, and available to those who need it.1 Our process for managing documented information includes robust controls for:

* **Creation, Review, and Approval:** Ensuring documents are accurate and fit for purpose.
* **Version Control:** Meticulous versioning of all documents, especially policies, procedures, and technical specifications. This is critically important for AI assets; we will use specialized tools to version datasets, model training code, and the trained models themselves to ensure full reproducibility and auditability of our AI systems.1
* **Access Control:** Ensuring that sensitive documentation is protected and accessible only to authorized personnel.
* **Retention and Disposition:** Managing the lifecycle of documents in accordance with legal, regulatory, and business requirements.

## **Part 3: Operationalizing Responsible AI**

### **Section 6: Operation of the AIMS (Clause 8)**

#### **6.1 Operational Planning and Control for the AI Lifecycle**

This section details the operational heart of the AIMS, outlining how the principles of responsible AI are translated into concrete, mandatory practices embedded within our day-to-day workflows. At Fawzooz AI, responsibility is not an afterthought or a final review gate; it is an integral component of every phase of the AI system lifecycle, from the first line of code to the final decommissioning of a system.1 Our operational controls are designed to be proactive, preventative, and auditable.

#### **6.2 Implementing AI Risk and Impact Assessments**

AI Risk Assessments and AI System Impact Assessments are not just strategic planning exercises; they are recurring operational activities.

* **Triggering Events:** An AIA and a corresponding AI Risk Assessment must be initiated for:
  1. All new AI systems or products before development begins.
  2. Any proposed major change to an existing AI system, such as retraining a model on a new dataset, significantly altering its algorithm, or expanding its intended use.
  3. At planned periodic intervals (e.g., annually) for high-risk systems to re-evaluate their impact in light of new data and evolving societal norms.
* **Operational Integration:** The outputs of these assessments—the identified risks and required mitigation measures—are formally documented and fed directly into the project's development backlog and the AI Risk Treatment Plan, ensuring they are actioned and tracked.1

#### **6.3 Managing the AI System Lifecycle Responsibly**

Fawzooz AI has implemented specific controls and procedures at each stage of the AI lifecycle to enforce our commitment to responsible AI.1

* **Design Phase:**
  + **Ethics-by-Design:** Every new AI project begins with a mandatory ethics and responsibility review, guided by the CAIO.
  + **Human Oversight by Design:** The design phase must explicitly define the mechanisms for human oversight. This involves identifying what decisions can be fully automated, which require human-in-the-loop review, and which require human-on-the-loop intervention capabilities. For our clinical tools, interfaces are designed to allow clinicians to easily review, validate, and, where necessary, override AI-generated suggestions.24
* **Data Management Phase:**
  + **Data Governance:** We enforce a strict data governance framework overseen by our Data Governance Lead.
  + **Data Minimization:** In adherence with GDPR and the HIPAA Minimum Necessary Standard, our systems are designed to collect and process only the PHI that is absolutely essential for the specific, stated therapeutic purpose.4
  + **Anonymization and Pseudonymization:** Robust techniques are applied to de-identify data wherever possible, particularly for data used in model training and analytics, to protect user privacy.7
  + **Data Quality and Bias Checks:** Before being used for training, all datasets undergo automated and manual checks for quality, completeness, and potential sources of demographic or societal bias.12
* **Development Phase:**
  + **Secure Development:** Our developers follow a Secure AI Development Lifecycle (Secure ADLC), which includes secure coding practices, vulnerability scanning of AI libraries, and protection of the model training environment.
  + **Fairness-Aware Algorithms:** Where applicable, development teams are required to implement and test fairness-aware machine learning techniques to actively mitigate identified biases.12
* **Verification & Validation Phase:**
  + **Comprehensive Testing:** Our Quality Assurance process includes a dedicated "Responsibility Testing" phase. This goes beyond testing for functional accuracy and includes:
    - **Fairness Testing:** Evaluating model performance across different demographic subgroups to detect and measure disparate impact.
    - **Robustness Testing:** Subjecting models to adversarial inputs and out-of-distribution data to test their resilience and prevent unexpected failures.
    - **Safety Testing:** Simulating edge-case scenarios to ensure the AI behaves safely and predictably.1
* **Deployment Phase:**
  + **Phased Rollouts:** High-impact AI systems are deployed using a phased or canary rollout approach, initially releasing to a limited user base to monitor real-world performance in a controlled environment.
  + **Pre-Deployment Sign-off:** No AI system can be deployed to production without a formal pre-deployment review and approval from the AI Governance Committee, which verifies that all high-risk issues from the AIA and risk assessment have been mitigated.1
* **Monitoring Phase:**
  + **Continuous Monitoring:** All production AI systems are subject to continuous, automated monitoring for performance degradation (model drift), shifts in input data distributions (data drift), and the emergence of new biases over time.5 Alerts are configured to notify the AI System Owner and development team of any significant deviations from established baselines.

#### **6.4 Crisis Intervention and Escalation Protocols for High-Risk Scenarios**

This set of operational controls is arguably the most critical for Fawzooz AI, as it directly addresses the highest-risk scenarios in mental wellness applications. The principle of "human oversight" is not abstract; it is operationalized into a concrete, tiered intervention protocol that ensures clinical safety.

* **The Role of the Suicide Risk Detection AI:** Our AI model that analyzes user text for suicide risk is explicitly defined as an **early warning and triage tool**, not a diagnostic or therapeutic tool.18 Its sole purpose is to flag potential risk for immediate human review.
* **The Crisis Escalation Protocol:** This non-negotiable protocol is triggered automatically when the AI model identifies a user at a "high" or "critical" risk level.
  1. **Automated Alert:** The system immediately generates a high-priority alert.
  2. **Human Handoff:** The alert is routed directly to our dedicated, 24/7 **Human Crisis Intervention Team**, which is staffed by licensed and trained mental health professionals. The user interaction is seamlessly transferred from the AI to the human clinician.
  3. **AI Disengagement:** The protocol explicitly **forbids the AI from continuing the crisis conversation**. Once the alert is triggered, the AI's role is complete. It may provide a transitional message like, "It sounds like you're going through a difficult time. I'm connecting you with one of our human clinicians right now who can help."
  4. **Clinical Response:** The human clinician takes over the interaction, engaging the user directly and following established clinical safety plans and crisis intervention procedures.
* **Risk Mitigation:** This protocol is our primary control for mitigating the severe risks of "client abandonment" during a crisis and the potential for an AI to provide an inappropriate or harmful response.15 It ensures that in the most critical moments, a qualified human is always in the loop, making the final judgment and providing the necessary care.18 This tiered approach—where AI assists in detection but humans are responsible for intervention—is fundamental to our operational philosophy.

### **Section 7: Performance Evaluation (Clause 9)**

#### **7.1 Monitoring, Measurement, Analysis, and Evaluation of the AIMS**

To ensure the AIMS is effective and achieving its goals, Fawzooz AI will establish and maintain a systematic process for monitoring, measuring, analyzing, and evaluating its performance.1 This data-driven approach provides objective evidence of our AI governance capabilities.

We will define and monitor Key Performance Indicators (KPIs) across several domains:

* **AIMS Process Metrics:** Timeliness of risk assessments, completion rates of required training, number of nonconformities identified and closed.
* **Control Effectiveness Metrics:** Results of periodic tests of security controls, effectiveness ratings of corrective actions.
* **AI System Responsibility Metrics:**
  + **Fairness:** Regular measurement of bias metrics (e.g., disparate impact, equal opportunity difference) for key models.
  + **Safety:** The accuracy and false positive/negative rates of our suicide risk detection model.
  + **Privacy:** Number of data subject access requests processed within the required timeframe.
* **Incident Metrics:** Number of AI-related incidents, mean time to detect (MTTD), and mean time to resolve (MTTR).

The results of these monitoring and measurement activities will be analyzed for trends and reported to relevant management. They serve as a critical input to the management review process.1

#### **7.2 Internal Audit Program**

Fawzooz AI will establish, implement, and maintain an internal audit program to conduct audits at planned intervals. The purpose of these audits is to provide independent and objective assurance that the AIMS conforms to the requirements of ISO/IEC 42001:2023 and to our own internal policies and procedures, and that it is effectively implemented and maintained.1

The audit program will be managed by the CISO's office. Auditors will be selected based on their competence and will be independent of the specific function or process being audited to ensure impartiality. The audit scope, criteria, and findings will be formally documented in an audit report, which will be distributed to relevant management. Any identified nonconformities will be entered into our corrective action process.1

#### **7.3 Management Review**

Top management at Fawzooz AI will formally review the AIMS at planned intervals (at least annually) to ensure its continuing suitability, adequacy, and effectiveness.1 This review is a critical governance checkpoint and demonstrates leadership's ongoing commitment.

The management review will consider a range of inputs, including:

* The results of performance monitoring and measurement.
* Findings from internal and external audits.
* The status of nonconformities and corrective actions.
* Changes in the external and internal context (e.g., new regulations, new AI risks).
* Feedback from interested parties.
* The adequacy of resources allocated to the AIMS.

The outputs of the management review will be formally documented and will include decisions on any necessary changes to the AIMS and opportunities for continual improvement. These outputs will be translated into actionable tasks with assigned responsibilities and timelines.1

### **Section 8: Improvement (Clause 10)**

#### **8.1 Nonconformity and Corrective Action**

Fawzooz AI will implement a formal process to manage nonconformities with AIMS requirements. When a nonconformity is identified (e.g., through an audit, incident, or user feedback), the following process will be initiated 1:

1. **Containment and Correction:** Immediate action will be taken to control and correct the nonconformity and to deal with its consequences.
2. **Root Cause Analysis (RCA):** A thorough investigation will be conducted to determine the underlying root cause(s) of the nonconformity to prevent its recurrence.
3. **Corrective Action:** Actions will be implemented to eliminate the identified root cause(s). This may involve changes to processes, controls, training, or the AI system itself.
4. **Effectiveness Review:** The effectiveness of the corrective action will be reviewed to ensure it has successfully resolved the issue.

All nonconformities and associated corrective actions will be documented in a central register, which will be reviewed by management.1

#### **8.2 Continual Improvement of the AIMS**

Continual improvement is a core principle of the Fawzooz AI AIMS. We are committed to continually improving the suitability, adequacy, and effectiveness of our AI governance framework.1 Our continual improvement strategy is driven by the Plan-Do-Check-Act (PDCA) cycle and leverages outputs from all aspects of the AIMS:

* **Performance evaluation results** highlight areas where objectives are not being met.
* **Internal audits and management reviews** identify weaknesses and opportunities for enhancement.
* **The nonconformity and corrective action process** provides direct feedback on process failures.
* **Proactive horizon scanning** for new technologies, risks, and regulations ensures the AIMS remains forward-looking and adaptive.

By embracing a culture of learning and continuous improvement, Fawzooz AI will ensure that our AIMS evolves to meet the challenges of the dynamic AI landscape and consistently supports our mission of providing responsible and trustworthy AI for mental wellness.

### **Part 4: Appendices**

#### **Appendix A: AIMS Documentation Templates**

This appendix provides standardized templates to be used for key AIMS documented information, ensuring consistency and completeness. These include templates for the AIMS Scope Document, AI Policy, AI Objectives Record, Nonconformity Report, and Management Review Decisions/Actions Log. 1

**Nonconformity Report Template** 1

| NC ID | Clause/Process/Area | Description of Nonconformity | Evidence | Severity (Optional) | Recommended Corrective Action Due Date (if suggested by auditee) |
| --- | --- | --- | --- | --- | --- |
| NC-001 | ISO 27001: A.8.2.2 Data Remanence | User data from decommissioned AI training environments was not securely erased, leading to potential data exposure risks. | Audit log review showing incomplete wipe procedures. Manual inspection of old server drives revealing residual data. | High | June 28, 2025 |
| NC-002 | Internal Policy: AI Model Documentation | The "Customer Support AI" model lacked comprehensive documentation regarding its training data sources, model architecture, and decision-making logic. | Review of internal documentation repository; "Customer Support AI" model folder missing required sections. | Medium | July 15, 2025 |
| NC-003 | ISO 42001: A.4.5 AI system verification and validation | The recent update to the "Fraud Detection AI" system was deployed without undergoing the mandatory re-validation against the latest fairness and bias metrics. | Deployment checklist incomplete for the recent update (v3.1). No record of re-validation tests in the test management system. | Critical | June 14, 2025 |
| NC-004 | Procedure: Incident Response for AI Systems | The team failed to follow the escalation protocol for a critical AI system outage, resulting in delayed resolution and increased impact on service availability. | Incident Report #2025-05-10 detailing communication breakdown and deviations from defined escalation paths. | High | July 5, 2025 |

**Management Review Decisions and Actions Agreed Template** 1

| Output Area | Decision/Action Agreed | Responsible Person/Team | Target Completion Date | Resources Allocated (if specific) |
| --- | --- | --- | --- | --- |
| Continual Improvement Opportunities | e.g., Initiate project to enhance AI model validation processes. | Head of AI Development |  |  |
| Changes to the AIMS (e.g., AI Policy, AI Objectives, Scope, Controls, Processes) | e.g., Revise AI Policy to address new ethical guideline. Update AI risk criteria. | CISO, AI Governance Lead |  |  |
| Resource Needs Identified | e.g., Approve budget for specialized AI ethics training. | Head of HR, CFO |  |  |
| Other Decisions | e.g., Commission a third-party review of AI security controls. | CISO |  |  |

#### **Appendix B: AI Risk Assessment and Treatment Plan Templates**

This appendix contains the detailed templates for conducting and documenting AI Risk Assessments. It includes the AI Risk Register template and the AI Risk Treatment Plan template, which guides the documentation of mitigation actions for identified risks. 1

**AI Risk Register Template** 1

| Risk ID | Date Identified | Risk Source/Category | Risk Description | AI System/Component Affected | Threat/Vulnerability (if applicable) | Potential Impact Areas | Impact Description | Impact Severity Rating | Likelihood of Occurrence Rating | Overall Risk Level (Calculated) | Existing Controls | Effectiveness of Existing Controls | Risk Treatment Decision | Proposed Treatment Action(s) / Control(s) | Control Reference (e.g., Annex A, SoA ID) | Risk Owner | Treatment Due Date | Treatment Status | Residual Risk Level (Post-Treatment) | Date of Last Review | Next Review Date | Comments/Notes |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AI-RISK-001 | 2025-05-10 | Model-related / Ethical | The AI-powered hiring tool might exhibit gender bias in candidate shortlisting due to historical data imbalances, leading to discriminatory hiring practices. | Candidate Shortlisting AI (Version 2.0) | Biased training data / Algorithmic bias in feature weighting. | Individuals (rights, well-being), Organizational (reputational, legal/compliance, financial) | Unfair candidate selection, legal challenges, damage to company reputation, reduced workforce diversity. | High | Medium | High | Data anonymization; Basic bias detection during model training. | Moderate (detects some bias but not all, especially subtle forms) | Modify | 1. Implement advanced bias detection & mitigation techniques (e.g., debiasing algorithms, fairness metrics). 2. Integrate human-in-the-loop review for high-risk candidate decisions. 3. Conduct regular bias audits with external experts. | A.4.3 (Data for AI systems), A.9.2 (Transparency), SoA-CTRL-A.4.3.01 (Bias Mitigation Policy) | Head of HR & AI Development Lead | 2025-09-30 | In Progress | Medium | 2025-06-05 | 2025-07-05 | Project initiated for 'Fairness in AI Hiring Tool'. External audit scheduled for Q4 2025. |
| AI-RISK-002 | 2025-05-15 | Data-related / Security (AI-specific) | A vulnerability in the AI model's input validation could be exploited by an attacker to inject malicious data (e.g., via prompt injection), leading to data exfiltration or unintended AI behavior. | Customer Service Chatbot AI (v1.5) | Prompt Injection / Lack of robust input sanitization. | Organizational (operational, reputational, financial), Individuals (privacy) | Disclosure of sensitive customer information, chatbot providing incorrect/malicious advice, service disruption, data breach penalties. | Critical | Medium | High | Web Application Firewall (WAF); Rate limiting on API calls. | Low (WAF not specifically trained for AI input attacks; rate limiting is a reactive control) | Modify | 1. Implement AI-specific input sanitization and validation rules for natural language inputs. 2. Utilize machine learning-based anomaly detection for unusual input patterns. 3. Conduct regular penetration testing specifically targeting prompt injection vulnerabilities. | A.4.2 (AI system requirements), A.8.4 (Data security), SoA-CTRL-A.8.4.03 (AI Security Testing Protocol) | CISO & AI Operations Lead | 2025-08-31 | Planned | Medium | 2025-06-08 | 2025-07-08 | Initial analysis of prompt injection techniques completed. Vendor security patch for underlying NLP library pending. |

**AI Risk Treatment Plan Template** 1

| Field | Description/Instructions |
| --- | --- |
| Risk ID | Unique identifier from the AI Risk Assessment Report/Register. |
| AI System/Component | Specific AI system or component the risk pertains to. |
| Risk Description | Brief description of the identified AI risk. |
| Evaluated Risk Level | The level of risk determined in the AI risk assessment (e.g., Critical, High, Medium). |
| Selected Risk Treatment Option(s) | Specify the chosen option(s) (e.g., Modify, Retain/Accept, Avoid, Share). |
| Treatment Action(s) / Control(s) to be Implemented | Detailed description of the actions or specific controls to be implemented. - For ""Modify"" option, list specific controls. - Reference relevant ISO 42001:2023 Annex A controls. - Note any additional controls designed by the organization. - Reference Statement of Applicability (SoA) ID where control is documented. |
| Justification for Treatment Option/Control Selection | Rationale for choosing the specific treatment and controls. |
| Responsible Person/Team | Individual or team accountable for implementing the treatment action/control. |
| Resources Required | Necessary resources (e.g., budget, personnel, technology). |
| Implementation Timeline/Deadline | Start date, key milestones, and completion date for the treatment action. |
| Method for Verifying Effectiveness | How the effectiveness of the implemented control/action will be monitored and verified. |
| Target Residual Risk Level | The expected level of risk after the treatment is implemented. |
| Status | (e.g., Planned, In Progress, Implemented, Verified, Closed). |
| Date of Status Update |  |

#### **Appendix C: AI System Impact Assessment Template**

This appendix provides the comprehensive worksheet template for conducting and documenting AI System Impact Assessments (AIAs). This template ensures a systematic evaluation of potential impacts on individuals, groups, and society, covering domains such as fairness, privacy, safety, and human rights. 1

**AI System Impact Assessment Template** 1

| Impact Area | Potential Impact Description (Positive/Negative) | Affected Parties | Likelihood | Severity | Existing Controls/Mitigations |
| --- | --- | --- | --- | --- | --- |
| Human Rights & Dignity | e.g., Impact on autonomy, freedom of expression |  |  |  |  |
| Fairness & Non-discrimination | e.g., Potential for biased outcomes |  |  |  |  |
| Privacy & Data Protection | e.g., Risk of PII exposure, re-identification |  |  |  |  |
| Transparency & Explainability | e.g., Difficulty in understanding decisions |  |  |  |  |
| Safety & Security (Physical & Psychological) | e.g., Risk of malfunction causing harm |  |  |  |  |
| Societal Impacts (e.g., Economic, Employment, Democratic Processes) | e.g., Job displacement, spread of misinformation |  |  |  |  |
| Environmental Sustainability | e.g., Energy consumption of model training/operation |  |  |  |  |
| Other (Specify) |  |  |  |  |  |

#### **Appendix D: AI Incident Response Plan (AIRP) and Incident Log Templates**

This appendix contains the official AI Incident Response Plan (AIRP) template for Fawzooz AI, which outlines the phases of preparation, identification, containment, eradication, recovery, and lessons learned for AI-specific incidents. It also includes the standardized AI Incident Log/Report template for documenting every incident.1

#### **Appendix E: Statement of Applicability (SoA) for AI Controls**

This appendix provides the template for the Statement of Applicability (SoA). The SoA is a critical document that lists all necessary AI controls selected from ISO/IEC 42001:2023 Annex A and any additional controls, with justification for their inclusion based on our AI risk assessment. It also documents the justification for excluding any Annex A controls that are not applicable. 1

**Statement of Applicability (SoA) - ISO 42001:2023 Annex A Controls Template** 1

| Annex A Control Ref. | Control Title (from Annex A) | Included (Yes/No) | Justification for Inclusion (Why this control is necessary. Link to identified AI risks, organizational objectives, legal/regulatory requirements, etc.) | Justification for Exclusion (If 'No', why this control is not applicable or necessary, e.g., risk assessment outcome, not required by external factors) | Reference to Implementation (e.g., Policy/Procedure ID, specific process, tool, or system where the control is implemented) | Implementation Status (e.g., Implemented, Partially Implemented, Planned) |
| --- | --- | --- | --- | --- | --- | --- |
| A.2.2 | AI policy |  |  |  |  |  |
| A.2.3 | Alignment with other organizational policies |  |  |  |  |  |
| A.2.4 | Review of the AI policy |  |  |  |  |  |
| A.3.2 | AI roles and responsibilities |  |  |  |  |  |
| A.3.3 | Competence |  |  |  |  |  |
| A.3.4 | Awareness |  |  |  |  |  |
| A.3.5 | Communication |  |  |  |  |  |
| A.3.6 | Documentation |  |  |  |  |  |
| A.4.2 | AI system requirements |  |  |  |  |  |
| A.4.3 | Data for AI systems |  |  |  |  |  |
| A.4.4 | AI system development and training |  |  |  |  |  |
| A.4.5 | AI system verification and validation |  |  |  |  |  |
| A.4.6 | AI system deployment |  |  |  |  |  |
| A.4.7 | AI system operation |  |  |  |  |  |
| A.4.8 | AI system maintenance and monitoring |  |  |  |  |  |
| A.4.9 | AI system modification |  |  |  |  |  |
| A.4.10 | AI system decommissioning |  |  |  |  |  |
| A.5.2 | AI risk assessment |  |  |  |  |  |
| A.5.3 | AI risk treatment |  |  |  |  |  |
| A.6.2 | Legal and regulatory requirements |  |  |  |  |  |
| A.6.3 | Contractual obligations |  |  |  |  |  |
| A.7.2 | Records |  |  |  |  |  |
| A.7.3 | Internal audit |  |  |  |  |  |
| A.7.4 | Management review |  |  |  |  |  |
| A.8.2 | Data quality |  |  |  |  |  |
| A.8.3 | Data privacy and protection |  |  |  |  |  |
| A.8.4 | Data security |  |  |  |  |  |
| A.9.2 | Transparency and explainability |  |  |  |  |  |
| A.9.3 | Interpretability |  |  |  |  |  |
| A.9.4 | Predictability |  |  |  |  |  |
| A.9.5 | Control and override capabilities |  |  |  |  |  |
| A.9.6 | Human oversight |  |  |  |  |  |
| A.9.7 | Monitoring of AI systems |  |  |  |  |  |
| A.10.2 | Users |  |  |  |  |  |
| A.10.3 | Third parties |  |  |  |  |  |
| A.10.4 | Customers |  |  |  |  |  |

**Additional Organizational Controls Template** 1

| Org. Control Ref. | Control Title/Description | Included (Yes/No) | Justification for Inclusion | Justification for Exclusion (if applicable for internally considered controls) | Reference to Implementation | Implementation Status |
| --- | --- | --- | --- | --- | --- | --- |
| ORG-AI-CTRL-001 |  | Yes | Addresses specific organizational ethical concerns for high-risk AI. |  |  | Implemented |
| ORG-AI-CTRL-002 |  | Yes | Ensures a systematic evaluation of potential societal and ethical impacts before deployment. |  |  | In Progress |
| ORG-AI-CTRL-003 | [Example: AI Model Interpretability Guidelines] | Yes | Provides guidance for AI developers to ensure models are understandable and explainable to relevant stakeholders. |  |  | Planned |
| ORG-AI-CTRL-004 |  | Yes | Establishes rules and processes for the responsible collection, storage, and use of data for AI systems. |  |  | Implemented |
| ORG-AI-CTRL-005 | [Example: Continuous AI Performance Monitoring Protocol] | Yes | Defines procedures for ongoing monitoring of AI system performance, bias, and drift post-deployment. |  |  | In Progress |
| ORG-AI-CTRL-006 |  | Yes | Specifies the use of particular tools to enhance the transparency and explainability of AI models. |  |  | Planned |
| ORG-AI-CTRL-007 |  | Yes | Outlines procedures for responding to and mitigating adverse events related to AI systems. |  |  | Planned |
| ORG-AI-CTRL-008 |  | Yes | Ensures relevant stakeholders are consulted during the design and development phases of AI systems. |  |  | In Progress |
| ORG-AI-CTRL-009 |  | Yes | Ensures all personnel involved in AI development and deployment are aware of ethical considerations and best practices. |  |  | Implemented |
| ORG-AI-CTRL-010 |  | Yes | Provides specific tools and methodologies for identifying and reducing bias in AI models and data. |  |  | In Progress |

#### **Appendix F: Mapping to Other Standards and Frameworks**

This appendix contains detailed compliance mapping tables that serve as a unified control framework for Fawzooz AI. These tables are critical tools for our compliance and audit functions, demonstrating how the implementation of a single AIMS control can satisfy requirements from multiple, overlapping regulatory regimes and align with other established frameworks. This integrated approach provides significant efficiency, reduces redundant compliance efforts, and demonstrates a holistic and mature approach to global AI governance. 1

**Unified Compliance Mapping (Abbreviated Example)**

| ISO/IEC 42001:2023 Control | Corresponding HIPAA Rule | Corresponding GDPR Article | Corresponding UAE Law Article |
| --- | --- | --- | --- |
| **A.5.4 Assessing impact on individuals** | Risk Analysis (§ 164.308(a)(1)(ii)(A)) | Art. 35: Data Protection Impact Assessment (DPIA) | PDPL Art. 21: DPIA |
| **A.7.2 Data security** | Security Rule (§ 164.312 Technical Safeguards) | Art. 32: Security of processing | PDPL Art. 6: Data Security |
| **A.7.3 Data quality** | Privacy Rule (§ 164.508(c) Integrity) | Art. 5(1)(d): Accuracy | PDPL Art. 4: Accuracy |
| **A.7.6 Data acquisition** | Privacy Rule (§ 164.502(a) Minimum Necessary) | Art. 5(1)(c): Data minimisation | PDPL Art. 5: Purpose Limitation |
| **A.8.2 Information for users** | Privacy Rule (§ 164.520 Notice of Privacy Practices) | Art. 13 & 14: Information to be provided | PDPL Art. 13: Right to Information |
| **A.8.4 Communication of incidents** | Breach Notification Rule (§ 164.400-414) | Art. 33 & 34: Notification of a personal data breach | PDPL Art. 9: Breach Notification |
| **B.9.3 Meaningful human oversight** | (Implicit in professional standards of care) | Art. 22: Automated individual decision-making | (Implicit in consent/control principles) |

*Table 2: Unified Compliance Mapping (Abbreviated Example)*

**Mapping ISO/IEC 42001:2023 to ISO/IEC 27001:2022** 1

| ISO/IEC 42001:2023 Clause/Annex A Control | Relevant ISO/IEC 27001:2022 Clause/Annex A Control | Notes on Overlap, Differences, and Integration Points |
| --- | --- | --- |
| e.g., Clause 5.2 AI Policy | e.g., Clause 5.2 Policy | Both require top-management-established policies. AI Policy will have AI-specific commitments. Integration can ensure consistency in overarching governance principles. |
| e.g., Clause 6.1.2 AI risk assessment | e.g., Clause 6.1.2 Information security risk assessment | Both mandate risk assessment. ISO 42001 focuses on AI-specific risks (bias, ethical, societal) in addition to security. Processes can be integrated, with AI-specific criteria and considerations added. |
| e.g., A.7.2 Data for development and enhancement of AI system | e.g., A.8.2.3 Handling of assets, A.5.1.4 Information classification | Overlap in need for data management. A.7.2 is AI-specific regarding data for model development. ISMS controls for data handling provide a baseline. |
| ... (and so on for other clauses/controls)... | ... | ... |

**Mapping ISO/IEC 42001:2023 Concepts to NIST AI RMF Functions** 1

| ISO/IEC 42001:2023 Element (Clause/Process/Control Theme) | NIST AI RMF Core Function (Govern, Map, Measure, Manage) | Notes on Alignment and How ISO 42001 Implements RMF Concepts |
| --- | --- | --- |
| e.g., Clauses 4, 5, 6.2 (Context, Leadership, Objectives, Policy) | GOVERN | ISO 42001 establishes the overarching governance structure, policies, and objectives for AI, aligning directly with the ""Govern"" function's intent to cultivate a risk-aware culture. |
| e.g., Clause 6.1.2 AI risk assessment, 6.1.4 AI system impact assessment | MAP | These ISO 42001 processes directly support the ""Map"" function by identifying AI system contexts, risks, and potential impacts. |
| e.g., Clause 9.1 Monitoring, measurement, analysis and evaluation | MEASURE | ISO 42001's requirements for monitoring and measuring AIMS performance and AI system characteristics align with the ""Measure"" function's focus on tracking and assessing AI risks and impacts. |
| e.g., Clause 6.1.3 AI risk treatment, Clause 8 Operation, Annex A controls | MANAGE | The risk treatment processes and operational controls within ISO 42001 provide the mechanisms to ""Manage"" identified AI risks, aligning with the RMF's objective to allocate resources to treated risks. |
| ... (and so on for other elements)... | ... | ... |

#### **Appendix G: Glossary of Terms**

This appendix provides definitions for key terms and concepts used within this AIMS manual and in the field of AI governance, ensuring a common vocabulary and understanding across Fawzooz AI. It includes terms from ISO/IEC 42001:2023 and others relevant to our specific domain.1

#### Works cited

1. ISO 42001 AI Management System (AIMS) Workbook.xlsx
2. Ethical Guidance for AI in the Professional Practice of Health Service Psychology, accessed on July 24, 2025, <https://www.apa.org/topics/artificial-intelligence-machine-learning/ethical-guidance-professional-practice.pdf>
3. yourhealthmagazine.net, accessed on July 24, 2025, <https://yourhealthmagazine.net/article/health-technology/hipaa-in-the-age-of-ai-what-mental-health-providers-need-to-know-about-secure-documentation/#:~:text=HIPAA's%20framework%20revolves%20around%20three,for%20any%20technology%20handling%20PHI.>
4. AI in Healthcare; What it means for HIPAA - Accountable HQ, accessed on July 24, 2025, <https://www.accountablehq.com/post/ai-and-hipaa>
5. How mental health professionals can be HIPAA compliant using mobile apps - Paubox, accessed on July 24, 2025, <https://www.paubox.com/blog/how-mental-health-professionals-can-be-hipaa-compliant-using-mobile-apps>
6. The Intersection of GDPR and AI and 6 Compliance Best Practices ..., accessed on July 24, 2025, <https://www.exabeam.com/explainers/gdpr-compliance/the-intersection-of-gdpr-and-ai-and-6-compliance-best-practices/>
7. GDPR-Compliant AI in Healthcare: A Guide to Data Privacy - Ailoitte Technologies, accessed on July 24, 2025, <https://www.ailoitte.com/insights/gdpr-compliant-healthcare-application/>
8. AI and the GDPR: Understanding the Foundations of Compliance - TechGDPR, accessed on July 24, 2025, <https://techgdpr.com/blog/ai-and-the-gdpr-understanding-the-foundations-of-compliance/>
9. muhami.ae, accessed on July 24, 2025, <https://muhami.ae/articles/how-is-health-data-protected-in-the-uae/#:~:text=The%20law%20places%20strict%20limits,data%20remains%20de%2Didentified%20where>
10. Data protection laws | The Official Portal of the UAE Government, accessed on July 24, 2025, <https://u.ae/en/about-the-uae/digital-uae/data/data-protection-laws>
11. UAE law protecting personal health data and information, accessed on July 24, 2025, <https://www.worldhealthexpo.com/insights/healthcare-regulation/uae-law-protecting-personal-health-data-and-information>
12. AI: right or wrong? 4 ethical considerations of AI in therapy - Upheal, accessed on July 24, 2025, <https://www.upheal.io/blog/ai-right-or-wrong-4-ethical-considerations-of-ai-in-therapy>
13. Ethical Considerations in Artificial Intelligence Interventions for Mental Health and Well-Being: Ensuring Responsible Implementation and Impact - MDPI, accessed on July 24, 2025, <https://www.mdpi.com/2076-0760/13/7/381>
14. Navigating the Promise and Risks of Artificial Intelligence in Mental Health Care, accessed on July 24, 2025, <https://www.learntolive.com/insights/navigating-the-promise-and-risks-of-artificial-intelligence-in-mental-health-care>
15. Artificial Intelligence in Behavioral Health: Challenging Ethical Issues, accessed on July 24, 2025, <https://www.addictioncounselorce.com/blog/artificial-intelligence-in-behavioral-health-challenging-ethical-issues>
16. Best Practices for the Use of Artificial Intelligence by Mental Health Therapists, accessed on July 24, 2025, <https://ai.utah.gov/wp-content/uploads/Best-Practices-Mental-Health-Therapists.pdf>
17. Build HIPAA/GDPR-compliant custom mental health software - Andersen, accessed on July 24, 2025, <https://andersenlab.com/industries/healthcare/mental-health-digital>
18. AI can help detect suicide risk earlier, FIU Business research finds., accessed on July 24, 2025, <https://business.fiu.edu/news/2025/ai-can-help-detect-suicide-risk-earlier-fiu-business-research-finds.html>
19. Balancing risks and benefits: clinicians' perspectives on the use of ..., accessed on July 24, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC12158938/>
20. Leveraging AI for Suicide Risk Prediction - American Psychoanalytic Association, accessed on July 24, 2025, <https://apsa.org/leveraging-ai-for-suicide-risk-prediction/>
21. Artificial intelligence in psychiatric education: Enhancing clinical ..., accessed on July 24, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC12077637/>
22. Responsible AI for Mental Health | Coursera, accessed on July 24, 2025, <https://www.coursera.org/learn/responsible-ai-for-mental-health>
23. Introduction to Artificial Intelligence for Mental Health Professionals - Psyched to Practice, accessed on July 24, 2025, <https://www.psychedtopractice.com/course/introduction-to-artificial-intelligence-for-mental-health-professionals>
24. Artificial intelligence in positive mental health: a narrative review - PMC, accessed on July 24, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC10982476/>
25. Responsible AI Integration in Mental Health Research: Issues, Guidelines, and Best Practices - PMC - PubMed Central, accessed on July 24, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC11624515/>