AIGLE: AI Governance & Lifecycle Explorer

Comprehensive Documentation & User Guide

Arinze Okosieme, Trusted AI Safety Expert (TAISE), CISSP, CCSP, CCZT

December 30, 2025

Contents

AIGLE

AI Governance & Lifecycle Explorer

Comprehensive Documentation & User Guide

Created by

Arinze Okosieme

Trusted AI Safety Expert (TAISE), CISSP, CCSP, CCZT

Website: aigle.datadid.io

Contact: hello@datadid.io

Version 1.0

December 30, 2025

Contents

Executive Summary

Overview of AIGLE

AIGLE (AI Governance & Lifecycle Explorer) is a comprehensive, interactive web-based platform designed to help organisations navigate the complex landscape of AI governance, risk management, and responsible AI development. Built on established frameworks including the NIST AI Risk Management Framework (AI RMF), AIGLE provides a structured approach to understanding, implementing, and maturing AI governance practices.

The platform combines three nested governance frameworks into a single, cohesive visualisation that maps the entire AI lifecycle from strategic governance through operational deployment and continuous monitoring. With over 28 interactive governance elements, 50+ curated open-source tools, and a sophisticated maturity assessment system, AIGLE serves as both an educational resource and a practical implementation toolkit.

Purpose and Objectives

The primary objectives of AIGLE are to:

- **Demystify AI Governance**: Provide clear, accessible explanations of complex governance concepts and their practical applications
- **Enable Self-Assessment**: Offer organisations a structured way to evaluate their current AI governance maturity across multiple dimensions
- **Bridge Theory and Practice**: Connect governance principles directly to actionable tools and techniques through curated open-source recommendations
- **Promote Trustworthy AI**: Align organisational practices with the eight NIST trustworthiness characteristics through visual mapping and targeted guidance
- **Facilitate Continuous Improvement**: Support organisations in identifying gaps, prioritising improvements, and tracking progress over time

Target Audience

AIGLE is designed for a diverse range of stakeholders involved in AI development, deployment, and governance:

- **AI Governance Leaders**: Chief AI Officers, AI Ethics Officers, and governance committee members seeking to establish or mature governance frameworks
- **Risk and Compliance Professionals**: Risk managers, compliance officers, and auditors responsible for AI risk management and regulatory compliance
- **Technical Teams**: Data scientists, ML engineers, and AI developers looking to understand governance requirements and implement best practices
- **Executive Leadership**: C-suite executives and board members requiring strategic oversight of AI initiatives and associated risks
- **Consultants and Advisors**: External advisors supporting organisations in AI governance implementation

- **Educators and Researchers**: Academic institutions and research organisations teaching or studying AI governance

Key Benefits

Organizations using AIGLE can expect to realise several significant benefits:

Strategic Clarity: Gain a comprehensive understanding of how different governance activities interconnect across the AI lifecycle, enabling more coherent strategy development and resource allocation.

Risk Reduction: Identify and address governance gaps before they manifest as operational failures, compliance violations, or reputational damage through systematic assessment and targeted recommendations.

Accelerated Implementation: Reduce the time and effort required to establish governance practices by leveraging curated tool recommendations and proven frameworks rather than starting from scratch.

Stakeholder Alignment: Create shared understanding across technical and non-technical stakeholders through visual representations and accessible explanations of governance concepts.

Measurable Progress: Track governance maturity over time using standardised assessment criteria, enabling data-driven decisions about governance investments and demonstrable improvement to stakeholders.

Cost Efficiency: Leverage open-source tools and established frameworks to build robust governance capabilities without significant licensing costs or vendor lock-in.

Regulatory Readiness: Align practices with emerging AI regulations and standards by grounding governance in widely recognisedd frameworks like NIST AI RMF.

Introduction

What is AIGLE?

AIGLE (AI Governance & Lifecycle Explorer) is an innovative, interactive web application that transforms abstract AI governance concepts into tangible, actionable guidance. At its core, AIGLE presents a unique three-layer visualisation that integrates:

1. **The NIST AI Risk Management Framework (AI RMF)** as the outer strategic layer
2. **The Responsible AI Lifecycle** as the middle operational layer
3. **The Model Development Lifecycle** as the inner technical layer

This nested architecture reflects the reality that effective AI governance requires alignment across strategic, operational, and technical dimensions. Each layer contains multiple interactive elements that users can explore to understand specific governance activities, their outputs, associated risks, and recommended tools.

Beyond the core visualisation, AIGLE includes a comprehensive maturity assessment system with over 60 targeted questions, a trust lens overlay that maps governance elements to NIST's eight trustworthiness characteristics, and a curated library of 50+ open-source tools for implementing governance practices.

The Need for AI Governance

The rapid advancement and deployment of AI systems across industries has created unprecedented opportunities alongside significant risks. Organizations face mounting challenges:

Regulatory Pressure: Jurisdictions worldwide are implementing AI-specific regulations (EU AI Act, US Executive Orders, sector-specific rules) requiring demonstrable governance and risk management.

Reputational Risk: High-profile AI failures involving bias, privacy violations, or safety issues can cause lasting damage to organisational reputation and stakeholder trust.

Operational Complexity: AI systems introduce unique risks related to data quality, model behaviour, deployment contexts, and ongoing performance that traditional IT governance doesn't adequately address.

Stakeholder Expectations: Customers, employees, investors, and civil society increasingly demand transparency, fairness, and accountability in AI systems.

Technical Uncertainty: The probabilistic nature of AI, potential for unexpected behaviours, and rapid evolution of capabilities create inherent uncertainty requiring structured risk management.

Resource Constraints: Organizations struggle to identify which governance activities matter most and how to implement them efficiently with limited resources and expertise.

How AIGLE Addresses This Need

AIGLE directly addresses these challenges through several key mechanisms:

Framework Integration: Rather than presenting governance as a collection of disconnected activities, AIGLE shows how strategic governance (NIST AI RMF), operational lifecycle management (Responsible AI Lifecycle), and technical development (Model Development Lifecycle) interconnect and reinforce each other.

Practical Guidance: Each governance element includes not just theoretical descriptions but practical information about activities to perform, artifacts to produce, risks addressed, and common pitfalls to avoid.

Tool Recommendations: By mapping 50+ actively maintained open-source tools to specific governance elements, AIGLE bridges the gap between "what to do" and "how to do it," enabling organisations to move quickly from planning to implementation.

Maturity Assessment: The structured assessment system helps organisations understand their current state objectively, identify priority gaps, and track improvement over time using a standardised five-level maturity model.

Trust Alignment: The trust lens overlay makes explicit how different governance activities contribute to specific trustworthiness characteristics, helping organisations ensure comprehensive coverage of trust dimensions.

Accessibility: By presenting complex governance concepts through interactive visualisation and plain-language explanations, AIGLE makes governance accessible to both technical and non-technical stakeholders.

Core Philosophy and Approach

AIGLE is built on several foundational principles:

Framework-Based: AIGLE grounds its approach in established, widely recognisedd frameworks (particularly NIST AI RMF) rather than proprietary methodologies, ensuring alignment with emerging standards and regulations.

Lifecycle-Oriented: Governance is presented as an integrated set of activities spanning the entire AI lifecycle, from initial problem framing through deployment and ongoing monitoring, rather than as a one-time compliance exercise.

Practical Over Theoretical: While grounded in sound governance principles, AIGLE emphasizes practical implementation through tool recommendations, artifact templates, and actionable guidance.

Open and Transparent: By focussing on open-source tools and publicly available frameworks, AIGLE promotes transparency and avoids vendor lock-in, enabling organisations to build sustainable governance capabilities.

Maturity-Based: AIGLE recognises that governance maturity develops over time and provides a structured path for organisations to progress from initial awareness through optimised, continuously improving practices.

Trust-Centered: The eight NIST trustworthiness characteristics serve as the north star, ensuring that governance activities ultimately serve the goal of building AI systems worthy of stakeholder trust.

Inclusive: AIGLE is designed to serve diverse stakeholders across roles, technical backgrounds, and organisational contexts, promoting shared understanding and collaboration.

Framework Architecture

Three-Layer Governance Model

AIGLE's distinctive three-layer architecture reflects the multi-dimensional nature of effective AI governance. Each layer addresses governance at a different level of abstraction and organisational scope, while maintaining clear connections between layers.

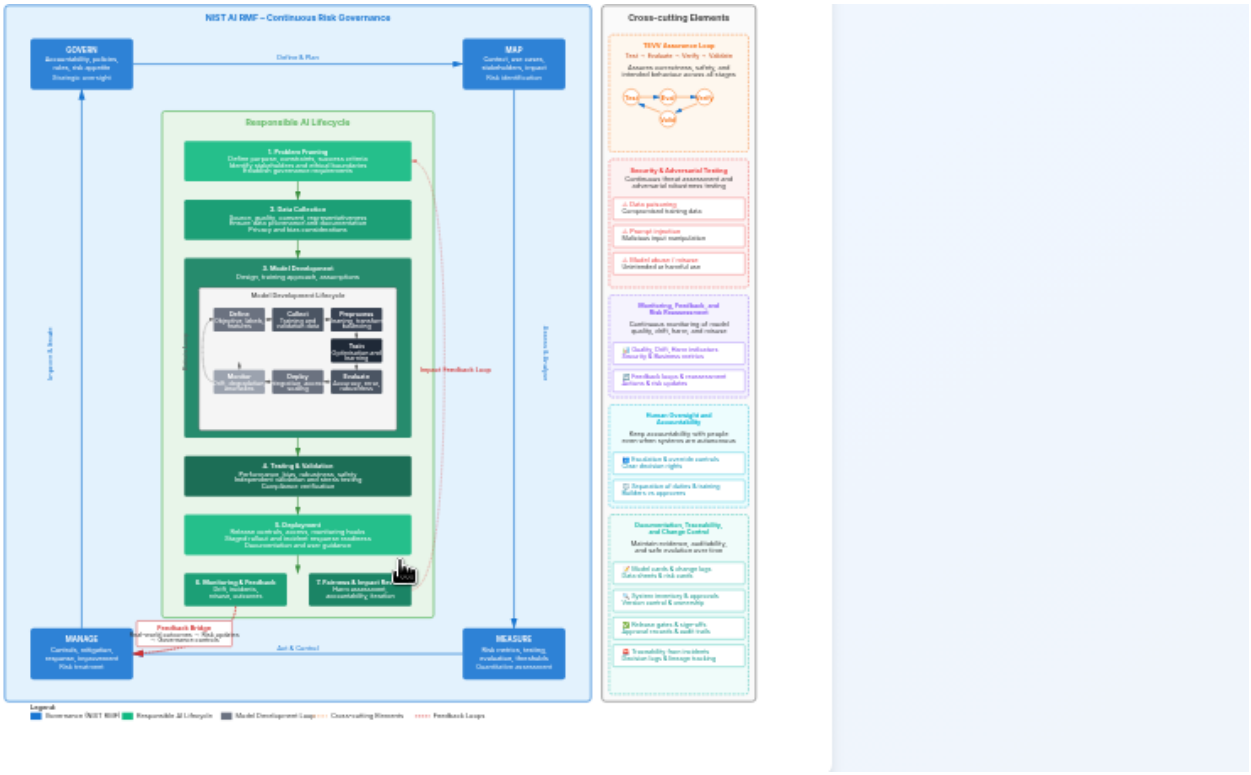

Figure 1: AIGLE Three-Layer Governance Diagram

Layer 1: NIST AI RMF (Outer Loop)

The outermost layer represents the NIST AI Risk Management Framework, which provides the strategic governance structure. This framework, developed by the U.S. National Institute of Standards and Technology, has become a global reference for AI risk management. The four core functions operate as a continuous cycle:

GOVERN: Organizational Governance Structure **Purpose**: Establish the organisational culture, structures, policies, and processes necessary for effective AI risk management across the enterprise.

Key Activities: - Define AI governance roles, responsibilities, and accountability structures - Establish AI risk appetite and tolerance levels aligned with organisational values - Create policies and procedures for AI development, deployment, and use - Implement governance mechanisms (committees, review boards, escalation paths) - Allocate re-

sources for AI risk management activities - Foster a culture of responsible AI through training and awareness

Artefacts Produced: - AI governance charter and policy framework - Roles and responsibilities matrix (RACI) - Risk appetite statement - Governance committee charters - Training and awareness materials - Resource allocation plans

Risks Addressed: - Lack of accountability for AI outcomes - Inconsistent risk management across AI initiatives - Insufficient resources for governance activities - Cultural resistance to responsible AI practices - Unclear decision-making authority

Common Pitfalls: - Creating governance structures that are too bureaucratic and slow innovation - Failing to secure executive sponsorship and adequate resources - Treating governance as a compliance checkbox rather than strategic enabler - Not adapting governance structures as AI capabilities and risks evolve - Insufficient integration with existing enterprise risk management

Trust Characteristics: Accountable & Transparent, Secure & Resilient

MAP: Context Identification and Risk Mapping **Purpose**: Understand the context in which AI systems will operate, identify relevant risks, and map them to potential impacts on individuals, organisations, and society.

Key Activities: - Identify and document AI system context (purpose, users, environment) - Catalog relevant legal, regulatory, and ethical requirements - Map stakeholders and their interests/concerns - Identify potential positive and negative impacts - Assess risk categories (bias, privacy, safety, security, etc.) - Document assumptions and limitations - Establish risk categorisation and prioritisation criteria

Artefacts Produced: - Context documentation and use case descriptions - Stakeholder analysis and engagement plans - Requirements traceability matrix - Risk register with categorised risks - Impact assessments (privacy, equity, safety) - Assumptions and limitations log

Risks Addressed: - Deploying AI in inappropriate contexts - Missing critical stakeholder perspectives - Overlooking relevant regulatory requirements - Failing to anticipate negative impacts - Inadequate understanding of system limitations

Common Pitfalls: - Conducting mapping as a one-time activity rather than ongoing process - Focusing only on technical risks while ignoring social/ethical dimensions - Insufficient stakeholder engagement, particularly with affected communities - Treating all risks as equal rather than prioritising based on severity and likelihood - Documenting risks without connecting them to mitigation strategies

Trust Characteristics: Accountable & Transparent, Fair with Harmful Bias Managed, Privacy-Enhanced

MEASURE: Assessment and Metrics **Purpose**: Quantify and qualify AI system performance, trustworthiness characteristics, and risk levels through systematic measurement and testing.

Key Activities: - Define metrics for trustworthiness characteristics (fairness, robustness, etc.) - Establish measurement methodologies and testing protocols - Implement continuous monitoring and measurement systems - Conduct regular assessments against defined criteria - Benchmark performance against industry standards - Track metrics over time to identify trends - Validate measurement approaches for reliability and validity

Artefacts Produced: - Metrics framework and definitions - Testing and evaluation protocols - Measurement results and scorecards - Benchmark comparisons - Trend analysis reports - Validation studies for measurement approaches

Risks Addressed: - Inability to detect performance degradation or emerging issues - Lack of objective evidence for trustworthiness claims - Insufficient visibility into system behaviour - Failure to identify bias or fairness issues - Inadequate basis for risk-informed decisions

Common Pitfalls: - Measuring only technical performance while ignoring trustworthiness dimensions - Using metrics that are easy to measure rather than meaningful - Failing to establish baselines and thresholds for action - Not validating that metrics actually measure what they claim to measure - Collecting data without analysing it or acting on insights

Trust Characteristics: Valid & Reliable, Fair with Harmful Bias Managed, Explainable & Interpretable

MANAGE: Risk Management and Response **Purpose**: Implement risk treatment strategies, respond to identified issues, and continuously improve risk management practices based on measurement insights.

Key Activities: - Develop and implement risk treatment plans (mitigate, transfer, accept, avoid) - Establish incident response and escalation procedures - Implement controls and safeguards - Monitor control effectiveness - Respond to identified issues and incidents - Document decisions and rationale - Continuously improve risk management based on lessons learned

Artefacts Produced: - Risk treatment plans and control documentation - Incident response procedures and playbooks - Control effectiveness reports - Incident logs and post-incident reviews - Decision logs with rationale - Lessons learned and improvement plans

Risks Addressed: - Inadequate response to identified risks - Ineffective risk controls - Slow or inappropriate incident response - Failure to learn from issues and improve - Lack of documentation for accountability

Common Pitfalls: - Implementing controls without verifying effectiveness - Treating risk management as separate from development rather than integrated - Failing to update risk treatments as context changes - Not establishing clear thresholds for escalation and response - Inadequate documentation of risk decisions and rationale

Trust Characteristics: Safe, Secure & Resilient, Accountable & Transparent

Layer 2: Responsible AI Lifecycle (Middle Flow)

The middle layer represents the operational AI lifecycle, showing the sequential flow of activities from initial problem definition through ongoing monitoring. This layer bridges strategic governance (outer layer) with technical development (inner layer).

Problem Framing **Purpose**: Define the problem to be solved, determine whether AI is appropriate, and establish success criteria before committing resources to development.

Key Activities: - Articulate the problem clearly and specifically - Assess whether AI is necessary and appropriate for the problem - Identify alternative approaches (including non-AI solutions) - Define success criteria and metrics - Conduct initial ethical and risk screening - Engage stakeholders in problem definition - Document assumptions and constraints

Artefacts Produced: - Problem statement and justification - AI appropriateness assessment - Success criteria and metrics definition - Initial risk screening results - Stakeholder input documentation - Assumptions and constraints log

Risks Addressed: - Applying AI to problems better solved by other means - Poorly defined problems leading to inappropriate solutions - Misalignment between technical solution and actual need - Overlooking ethical concerns early when changes are easier - Insufficient stakeholder input leading to rejected solutions

Common Pitfalls: - Rushing to AI solutions without considering alternatives - Defining problems too narrowly or too broadly - Focusing on technical feasibility without considering social implications - Not involving affected stakeholders in problem framing - Failing to establish clear, measurable success criteria

Trust Characteristics: Accountable & Transparent, Valid & Reliable

Data Collection & Preparation **Purpose**: Acquire, clean, and prepare data for model development while ensuring quality, representativeness, privacy, and appropriate documentation.

Key Activities: - Identify data sources and assess availability - Evaluate data quality, completeness, and representativeness - Implement privacy-preserving data collection methods - Clean and preprocess data - Address missing data and outliers - Document data provenance and lineage - Assess and mitigate data bias - Implement data governance and access controls

Artefacts Produced: - Data collection plan and protocols - Data quality assessment reports - Data documentation (datasheets, data cards) - Privacy impact assessments - Bias analysis reports - Data lineage documentation - Data governance policies

Risks Addressed: - Poor data quality leading to unreliable models - Biased data perpetuating or amplifying discrimination - Privacy violations through inappropriate data collection or use - Lack of data representativeness limiting generalization - Insufficient documentation hindering reproducibility and accountability

Common Pitfalls: - Assuming available data is appropriate without critical evaluation - Failing to document data limitations and biases - Inadequate privacy protections, especially for sensitive data - Not considering data representativeness across relevant subgroups - Insufficient data governance leading to inappropriate access or use

Trust Characteristics: Privacy-Enhanced, Fair with Harmful Bias Managed, Valid & Reliable

Model Development Purpose: Design, train, evaluate, and validate AI models through an iterative technical process (detailed in Layer 3).

This phase encompasses the inner cycle of feature engineering, model training, model evaluation, and model validation. See Layer 3 for detailed breakdown.

Key Activities (High-Level): - Engineer features from prepared data - Select and train appropriate model architectures - Evaluate model performance across multiple dimensions - Validate models against real-world conditions - Document model design decisions and trade-offs - Assess model trustworthiness characteristics

Artefacts Produced: - Trained model artifacts - Model cards and documentation - Performance evaluation reports - Validation test results - Design decision logs - Trustworthiness assessments

Risks Addressed: - Poor model performance on intended tasks - Models that perform well in training but fail in deployment - Lack of transparency in model behaviour - Undetected bias or fairness issues - Insufficient robustness to adversarial inputs or distribution shift

Common Pitfalls: - Optimizing for single metrics without considering trade-offs - Insufficient testing across diverse scenarios and subgroups - Inadequate documentation of model limitations - Not validating models in realistic deployment conditions - Treating model development as purely technical without governance integration

Trust Characteristics: Valid & Reliable, Explainable & Interpretable, Fair with Harmful Bias Managed

Deployment & Integration Purpose: Transition validated models into production environments, integrate with existing systems, and establish operational procedures.

Key Activities: - Plan deployment architecture and infrastructure - Implement model serving and inference systems - Integrate with upstream and downstream systems - Establish operational procedures and runbooks - Implement monitoring and alerting systems - Conduct deployment testing and validation - Train operational staff - Implement access controls and security measures - Plan rollback and contingency procedures

Artefacts Produced: - Deployment architecture documentation - Integration specifications - Operational procedures and runbooks - Deployment test results - Training materials for operational staff - Security and access control documentation - Rollback and contingency plans

Risks Addressed: - System failures during deployment - Integration issues with existing systems - Inadequate operational support leading to poor performance - Security

vulnerabilities in production environment - Lack of preparedness for issues requiring rollback

Common Pitfalls: - Insufficient testing in production-like environments before deployment - Inadequate operational documentation and training - Not implementing proper monitoring from day one - Failing to plan for rollback and incident response - Treating deployment as the end rather than beginning of operational lifecycle

Trust Characteristics: Secure & Resilient, Safe, Accountable & Transparent

Monitoring & Maintenance **Purpose**: Continuously monitor deployed AI systems, detect issues, maintain performance, and ensure ongoing trustworthiness.

Key Activities: - Monitor system performance and behaviour continuously - Track trustworthiness metrics (fairness, robustness, etc.) - Detect data drift and model degradation - Respond to alerts and incidents - Perform regular maintenance and updates - Collect feedback from users and stakeholders - Assess ongoing compliance with requirements - Document operational history and issues

Artefacts Produced: - Monitoring dashboards and reports - Incident logs and response documentation - Performance trend analysis - Drift detection reports - Maintenance logs - User feedback summaries - Compliance audit trails

Risks Addressed: - Undetected performance degradation - Model behaviour drift due to changing data distributions - Emerging fairness or bias issues in production - Security incidents and adversarial attacks - Non-compliance with evolving requirements

Common Pitfalls: - Monitoring only technical metrics without trustworthiness dimensions - Slow response to detected issues - Insufficient resources allocated to ongoing monitoring - Not collecting and acting on user feedback - Treating monitoring as automated rather than requiring human judgment

Trust Characteristics: Valid & Reliable, Safe, Secure & Resilient, Fair with Harmful Bias Managed

Impact Review & Iteration **Purpose**: Periodically assess actual impacts of deployed AI systems, compare to intended outcomes, and determine whether to continue, modify, or retire systems.

Key Activities: - Conduct comprehensive impact assessments - Compare actual outcomes to intended goals and success criteria - Assess unintended consequences and emergent behaviours - Gather stakeholder feedback on system impacts - Evaluate continued appropriateness and value - Make decisions about system continuation, modification, or retirement - Document lessons learned - Feed insights back into governance and future development

Artefacts Produced: - Impact assessment reports - Stakeholder feedback summaries - Outcome vs. goal comparison analysis - Continuation/modification/retirement decisions - Lessons learned documentation - Recommendations for governance improvements

Risks Addressed: - Continued operation of systems that no longer serve their purpose - Unaddressed negative impacts on stakeholders - Failure to learn from experience and improve practices - Misalignment between system operation and organisational values - Missed opportunities to optimise or enhance systems

Common Pitfalls: - Conducting reviews too infrequently or superficially - Focusing only on technical performance without assessing broader impacts - Insufficient stakeholder engagement in impact assessment - Not acting on review findings (continuing problematic systems) - Failing to feed lessons learned back into governance and development practices

Trust Characteristics: Accountable & Transparent, Fair with Harmful Bias Managed, Valid & Reliable

Layer 3: Model Development Lifecycle (Inner Cycle)

The innermost layer details the iterative technical process of developing AI models. This cycle typically repeats multiple times during the Model Development phase of Layer 2.

Feature Engineering Purpose: Transform raw data into features (input variables) that effectively represent the problem and enable model learning.

Key Activities: - Analyze data characteristics and relationships - Create new features through transformation and combination - Select relevant features and remove redundant ones - Encode categorical variables appropriately - Normalize and scale features - Handle temporal and spatial aspects - Document feature definitions and rationale - Assess feature fairness implications

Artefacts Produced: - Feature definitions and documentation - Feature engineering code and pipelines - Feature importance analysis - Feature fairness assessments - Transformation and encoding specifications

Risks Addressed: - Poor model performance due to inadequate feature representation - Introduction of bias through feature selection or engineering - Lack of interpretability due to opaque feature transformations - Data leakage through inappropriate feature construction - Inability to reproduce results due to undocumented features

Common Pitfalls: - Creating features that leak information from the target variable - Not considering fairness implications of feature choices - Over-engineering features leading to overfitting - Insufficient documentation of feature definitions and rationale - Not validating that engineered features make domain sense

Trust Characteristics: Valid & Reliable, Fair with Harmful Bias Managed, Explainable & Interpretable

Model Training Purpose: Use prepared data and engineered features to train AI models by optimising model parameters to minimise prediction errors.

Key Activities: - Select appropriate model architectures and algorithms - Configure hyperparameters - Split data into training, validation, and test sets - Train models using

appropriate optimisation techniques - Implement regularization to prevent overfitting - Use cross-validation for robust performance estimation - Document training procedures and configurations - Track experiments and model versions

Artefacts Produced: - Trained model artifacts and weights - Training configuration documentation - Experiment tracking logs - Hyperparameter tuning results - Cross-validation performance metrics - Model versioning records

Risks Addressed: - Overfitting to training data, poor generalization - Suboptimal model performance due to poor hyperparameter choices - Inability to reproduce training results - Lack of transparency in model development process - Inefficient use of computational resources

Common Pitfalls: - Not using proper train/validation/test splits, leading to overoptimistic performance estimates - Insufficient hyperparameter tuning - Training on biased or unrepresentative data samples - Not tracking experiments systematically - Optimizing for single metrics without considering trade-offs

Trust Characteristics: Valid & Reliable, Explainable & Interpretable

Model Evaluation Purpose: Assess trained model performance across multiple dimensions including accuracy, fairness, robustness, and other trustworthiness characteristics.

Key Activities: - Evaluate performance on held-out test data - Assess performance across demographic subgroups - Test fairness using multiple fairness metrics - Evaluate robustness to input perturbations - Assess calibration and uncertainty quantification - Analyze errors and failure modes - Compare to baseline and alternative models - Document evaluation results and limitations

Artefacts Produced: - Performance evaluation reports - Fairness assessment results - Robustness testing results - Error analysis documentation - Model comparison reports - Limitation and failure mode documentation

Risks Addressed: - Deploying models with inadequate performance - Undetected bias and fairness issues - Lack of robustness to real-world variations - Poor calibration leading to overconfident predictions - Insufficient understanding of model limitations

Common Pitfalls: - Evaluating only on overall metrics without subgroup analysis - Using single fairness metrics without considering trade-offs - Insufficient robustness testing - Not analysing errors to understand failure modes - Treating evaluation as one-time activity rather than iterative process

Trust Characteristics: Valid & Reliable, Fair with Harmful Bias Managed, Explainable & Interpretable

Model Validation Purpose: Verify that models meet requirements and perform acceptably in realistic deployment conditions before production release.

Key Activities: - Test models in production-like environments - Validate against real-world data and scenarios - Conduct user acceptance testing - Verify compliance with requirements and constraints - Assess operational feasibility and resource requirements

- Test integration with surrounding systems - Validate monitoring and alerting systems - Obtain stakeholder sign-off for deployment

Artefacts Produced: - Validation test plans and results - User acceptance testing documentation - Requirements compliance verification - Integration test results - Operational readiness assessment - Stakeholder approval documentation

Risks Addressed: - Models that perform well in development but fail in production - Unmet requirements discovered after deployment - Inadequate operational support infrastructure - Stakeholder rejection of deployed systems - Compliance violations

Common Pitfalls: - Validating only on historical data without realistic deployment scenarios - Insufficient user acceptance testing - Not validating operational aspects (latency, resource usage, etc.) - Treating validation as rubber-stamp rather than critical gate - Inadequate documentation of validation results and decisions

Trust Characteristics: Valid & Reliable, Safe, Secure & Resilient

Cross-Cutting Elements

Four critical elements span all layers and phases of the AI lifecycle, requiring continuous attention throughout:

Documentation & Transparency

Purpose: Maintain comprehensive, accessible documentation of AI systems, decisions, and processes to enable transparency, accountability, and reproducibility.

Key Activities Across Lifecycle: - Document system purpose, context, and intended use - Record design decisions and rationale - Maintain data and model documentation (datasheets, model cards) - Document testing and evaluation results - Record operational procedures and incidents - Create user-facing documentation and disclosures - Maintain audit trails of key decisions

Artefacts: - System documentation (purpose, context, use cases) - Data documentation (datasheets, data cards) - Model documentation (model cards, technical specifications) - Decision logs and rationale - User documentation and disclosures - Audit trails and compliance records

Trust Characteristics: Accountable & Transparent, Explainable & Interpretable

Human Oversight & Control

Purpose: Ensure appropriate human involvement in AI system development, deployment, and operation to maintain accountability and enable intervention.

Key Activities Across Lifecycle: - Define human roles and responsibilities - Implement human-in-the-loop mechanisms where appropriate - Establish human review and approval gates - Enable human override and intervention capabilities - Train humans to effectively oversee AI systems - Monitor and support human decision-makers - Maintain human accountability for AI outcomes

Artefacts: - Human oversight procedures and protocols - Roles and responsibilities documentation - Training materials for human overseers - Intervention and override procedures - Human review and approval records

Trust Characteristics: Accountable & Transparent, Safe, Secure & Resilient

Stakeholder Engagement

Purpose: Involve relevant stakeholders throughout the AI lifecycle to incorporate diverse perspectives, build trust, and ensure systems serve stakeholder needs.

Key Activities Across Lifecycle: - Identify relevant stakeholders (users, affected parties, domain experts, etc.) - Engage stakeholders in problem framing and requirements definition - Gather stakeholder input on design decisions - Conduct user testing and gather feedback - Communicate with stakeholders about system capabilities and limitations - Address stakeholder concerns and incorporate feedback - Maintain ongoing stakeholder relationships

Artefacts: - Stakeholder analysis and engagement plans - Stakeholder input documentation - User testing and feedback summaries - Communication materials and disclosures - Stakeholder concern tracking and resolution

Trust Characteristics: Accountable & Transparent, Fair with Harmful Bias Managed

Continuous Learning & Adaptation

Purpose: Systematically learn from experience and adapt governance and development practices based on new knowledge, changing contexts, and lessons learned.

Key Activities Across Lifecycle: - Collect and analyse operational data and feedback - Conduct regular reviews and retrospectives - Identify lessons learned and improvement opportunities - Update practices, policies, and procedures based on insights - Stay current with evolving standards, regulations, and best practices - Share knowledge across teams and projects - Foster culture of continuous improvement

Artefacts: - Lessons learned documentation - Practice and policy updates - Retrospective and review reports - Knowledge sharing materials - Improvement tracking and implementation plans

Trust Characteristics: Valid & Reliable, Secure & Resilient, Accountable & Transparent

Core Features

Interactive Master Diagram

The centrepiece of AIGLE is its interactive master diagram, which visualises all three governance layers in a single, cohesive view. This diagram serves as both an educational tool and a navigation interface for exploring governance elements in depth.

Visual Representation

The diagram uses a nested circular design that intuitively represents the relationship between layers:

- **Outer Ring**: The four NIST AI RMF functions (GOVERN, MAP, MEASURE, MANAGE) form the strategic perimeter
- **Middle Flow**: The six Responsible AI Lifecycle phases flow sequentially around the middle layer
- **Inner Cycle**: The four Model Development Lifecycle steps form an iterative cycle at the centre
- **Cross-Cutting Elements**: Four elements are positioned to indicate their relevance across all layers

Color coding provides visual distinction between layers and element types: - **Blue tones** (#3B82F6): Strategic governance elements - **Green tones** (#10B981): Operational lifecycle elements - **Grey tones** (#6B7280): Technical development elements - **Accent colours**: Cross-cutting elements

Interactive Elements

Each of the 28+ governance elements in the diagram is clickable and interactive:

Hover Effects: Elements highlight on hover, providing immediate visual feedback and indicating interactivity.

Click Actions: Clicking an element opens a detailed side panel with comprehensive information about that element, including: - Purpose and objectives - Key activities to perform - Artefacts produced - Risks addressed - Common pitfalls to avoid - Trust characteristics alignment - Recommended open-source tools

Visual Feedback: Selected elements remain highlighted while their detail panel is open, helping users maintain context.

Container Interactions: The three large background containers (outer ring, middle flow, inner cycle) are themselves interactive, providing high-level strategic overviews of each layer when clicked.

Responsive Design

The diagram adapts to different screen sizes and devices:

- **Desktop**: Full-size diagram with all elements clearly visible and labelled

- **Tablet**: Scaled diagram maintaining readability and interactivity
- **Mobile**: Optimized layout with touch-friendly element sizing

Accessibility Features

The diagram includes several accessibility considerations:

- **Keyboard Navigation**: All interactive elements are keyboard-accessible
- **Screen Reader Support**: Semantic HTML and ARIA labels provide context for screen readers
- **High Contrast**: Color choices meet WCAG contrast requirements
- **Focus Indicators**: Clear visual indicators for keyboard focus

Trust Lens Toggle

The Trust Lens is an innovative overlay feature that maps governance elements to the eight NIST trustworthiness characteristics, helping users understand how different activities contribute to building trustworthy AI systems.

Eight NIST Trustworthiness Dimensions

The NIST AI RMF identifies eight characteristics of trustworthy AI systems:

1. **Valid & Reliable**: Systems perform consistently and accurately for their intended purpose
2. **Safe**: Systems do not pose unreasonable risks to safety or health
3. **Secure & Resilient**: Systems are protected against threats and can recover from disruptions
4. **Accountable & Transparent**: Systems enable appropriate accountability and transparency
5. **Explainable & Interpretable**: Systems provide appropriate explanations of their behaviour
6. **Privacy-Enhanced**: Systems protect privacy and data confidentiality
7. **Fair with Harmful Bias Managed**: Systems avoid harmful bias and promote fairness
8. **Environmentally Sustainable**: Systems minimise environmental impact (future consideration)

Visual Overlay Mapping

When activated, the Trust Lens overlay:

Highlights Elements: Governance elements glow with colours corresponding to the trust dimensions they primarily support.

Shows Connections: Visual indicators (glow effects) highlight relationships between elements and trust characteristics.

Displays Legend: A collapsible legend panel appears in the bottom-right corner showing all eight trust characteristics with their colour-coded indicators.

Enhances Understanding: Click any element to see detailed information about which specific trust dimensions it impacts and why.

Color-Coded Indicators

Each trust dimension is assigned a distinct colour: - Valid & Reliable: Blue - Safe: Green - Secure & Resilient: Purple - Accountable & Transparent: Orange - Explainable & Interpretable: Yellow - Privacy-Enhanced: Pink - Fair with Harmful Bias Managed: Teal - Environmentally Sustainable: Lime (future)

Elements that contribute to multiple trust dimensions show blended or multiple colour indicators.

Use Cases

The Trust Lens serves several important purposes:

Gap Analysis: Organizations can identify which trust dimensions are well-covered by their current governance activities and which need more attention.

Prioritization: When resources are limited, organisations can prioritise governance elements that address their most critical trust concerns.

Communication: The visual mapping helps communicate to stakeholders how governance activities translate into trustworthy outcomes.

Education: The Trust Lens helps users understand the multidimensional nature of AI trustworthiness and how different activities contribute.

Maturity Assessment

AIGLE includes a comprehensive maturity assessment system that enables organisations to evaluate their current AI governance capabilities and identify improvement priorities.

Assessment Structure

The assessment consists of 60+ targeted questions organised by governance element:

Element-Specific Questions: Each of the 28 governance elements has 2-4 questions specifically designed to assess maturity in that area.

Balanced Coverage: Questions span strategic, operational, and technical dimensions to provide comprehensive assessment.

Clear Language: Questions are written in accessible language suitable for both technical and non-technical respondents.

Context-Appropriate: Questions are designed to be relevant across different organisational contexts and AI use cases.

Five-Level Maturity Model

Responses are evaluated against a five-level maturity model:

Level 0 - Initial/Ad Hoc: Governance activities are absent or performed inconsistently without formal processes. AI development proceeds without structured governance.

Level 1 - Aware: Organization recognises the need for AI governance and has begun initial activities. Some documentation and processes exist but are incomplete and inconsistently applied.

Level 2 - Defined: Formal governance processes and policies are documented and communicated. Governance activities are performed consistently for most AI projects, though integration may be incomplete.

Level 3 - Managed: Governance is well-integrated into AI development and operations. Metrics are tracked, and governance effectiveness is monitored. Continuous improvement processes are in place.

Level 4 - Optimized: Governance is continuously optimised based on data and feedback. Organization is a leader in responsible AI practices. Governance enables innovation while effectively managing risks.

Scoring and Results

After completing the assessment, users receive:

Overall Maturity Score: Aggregate score across all governance elements, indicating overall governance maturity level.

Element-Level Scores: Individual scores for each governance element, showing strengths and gaps.

Layer Scores: Aggregate scores for each of the three layers (NIST AI RMF, Responsible AI Lifecycle, Model Development Lifecycle).

Trust Dimension Scores: Scores indicating maturity in supporting each of the eight trustworthiness characteristics.

Visual Dashboard: Graphical representation of scores using charts and heat maps for easy interpretation.

Personalized Recommendations

Based on assessment results, AIGLE provides:

Priority Improvements: Identification of governance elements with the largest gaps that should be prioritised for improvement.

Specific Actions: Concrete recommendations for activities to perform to advance maturity in each area.

Tool Recommendations: Suggestions for open-source tools that can help implement recommended improvements.

Resource Estimates: Guidance on the level of effort and resources typically required for recommended improvements.

Sequencing Guidance: Recommendations on the order in which to tackle improvements based on dependencies and typical maturity progression.

PDF Report Generation

Assessment results can be exported as a comprehensive PDF report including:

- Executive summary of overall maturity
- Detailed scores by element, layer, and trust dimension
- Visual charts and graphs
- Prioritized recommendations
- Tool suggestions
- Comparison to typical maturity progression

Email Delivery

Users can request their assessment report be delivered via email:

Privacy-Preserving: Email addresses are used only for report delivery and not stored or used for other purposes.

GDPR Compliant: Email handling follows GDPR requirements for data minimization and purpose limitation.

Professional Format: Emailed reports are professionally formatted and suitable for sharing with stakeholders.

Open Source Tools Library

One of AIGLE's most valuable features is its curated library of 50+ open-source tools for implementing AI governance practices. This library bridges the gap between governance concepts and practical implementation.

Important Note: Not all governance elements have software tools available. Some elements, particularly those focused on methodology-based activities like stakeholder mapping and context discovery (MAP), rely on human expertise and facilitation processes rather than software automation. For these elements, AIGLE provides guidance on manual approaches and best practices.

Tool Assignment Philosophy

Tools are mapped to governance elements based on their primary function and alignment with element objectives:

Organizational vs. Technical Governance: - **GOVERN element** focuses on organizational governance platforms (policy frameworks, compliance management, governance processes) - **MANAGE/DEPLOYMENT elements** focus on technical implementation (policy enforcement tools, runtime controls, deployment automation) - Example: Policy authoring platforms belong in GOVERN, while policy enforcement engines (like OPA, Kyverno) belong in MANAGE/DEPLOYMENT

Measurement vs. Context Discovery: - **MEASURE element** includes quantitative tools (metrics, testing, bias measurement with specific scores) - **MAP element** has no software tools—it requires methodology-based approaches (stakeholder workshops, interviews, impact assessments) that demand human expertise and judgment

Element-Specific Mapping: Tools may appear in multiple elements when they serve different purposes across the lifecycle (e.g., MLflow for experiment tracking in development and audit trails in documentation).

Curation Criteria

Tools included in the library meet strict criteria:

Actively Maintained: Tools must be under active development with recent commits and releases.

Open Source: All tools are open source with permissive licences (MIT, Apache 2.0, BSD, etc.).

Production-Ready: Tools are mature enough for production use, not just research prototypes.

Well-Documented: Tools have comprehensive documentation enabling adoption.

Community Support: Tools have active user communities and support channels.

Proven Value: Tools have demonstrated value through adoption by organisations and positive community feedback.

Tool Categories

Tools are organised into categories aligned with governance needs:

Fairness & Bias: - AI Fairness 360 (IBM): Comprehensive fairness metrics and mitigation algorithms - Fairlearn (Microsoft): Fairness assessment and unfairness mitigation - Aequitas: Bias and fairness audit toolkit - What-If Tool: Visual interface for fairness analysis

Explainability & Interpretability: - SHAP: Unified approach to explaining model predictions - LIME: Local interpretable model-agnostic explanations - InterpretML: Interpretable machine learning algorithms - Alibi: Algorithms for explaining ML models - ELI5: Debugging and explaining ML classifiers

Model Testing & Validation: - Garak: LLM vulnerability scanner - DeepChecks: Testing and validation for ML models - Evidently: ML model monitoring and testing - Great Expectations: Data validation and testing

Data Quality & Validation: - Great Expectations: Data quality testing - Pandera: Statistical data validation - Deequ: Data quality validation at scale - TensorFlow Data Validation: Data validation for ML pipelines

Model Monitoring: - Evidently: ML monitoring and observability - WhyLabs: ML monitoring and observability - Fiddler: ML monitoring and explainability - Arize: ML observability platform

Privacy & Security: - Opacus: Training models with differential privacy - PySyft: Privacy-preserving ML - TensorFlow Privacy: Privacy-preserving ML - Adversarial Robustness Toolbox: Defending against adversarial attacks

MLOps & Governance: - MLflow: ML lifecycle management - DVC: Data and model versioning - Weights & Biases: Experiment tracking - Kubeflow: ML workflows on Kubernetes

Documentation & Transparency: - Model Card Toolkit: Generating model cards - Datasheets for Datasets: Dataset documentation - VerifyML: Model documentation and validation

Tool Information

For each tool, AIGLE provides:

Description: Clear explanation of what the tool does and its primary use cases.

Key Features: Highlight of the tool's most important capabilities.

Language/Framework: Programming languages and ML frameworks supported.

License: Open source licence type.

Links: Direct links to: - GitHub repository - Official documentation - Project website - Tutorials and examples

Governance Element Mapping: Clear indication of which governance elements the tool supports.

Trust Dimension Mapping: Indication of which trustworthiness characteristics the tool helps achieve.

Integration with Governance Elements

Tools are mapped to specific governance elements, so when users explore an element in the diagram, they immediately see relevant tool recommendations. This contextual presentation helps users quickly identify tools that can help them implement specific governance activities.

Filtering and Search

Users can discover tools through multiple pathways:

By Governance Element: See tools relevant to a specific governance activity.

By Trust Dimension: Find tools that support specific trustworthiness characteristics.

By Category: Browse tools by functional category (fairness, explainability, etc.).

By Technology: Filter by programming language or ML framework.

Search: Free-text search across tool names, descriptions, and features.

Guided Interactive Tour

To help new users understand AIGLE's features and navigate the complex governance landscape, the platform includes a comprehensive guided tour system.

Tour Structure

The tour consists of 15 carefully sequenced steps that progressively introduce users to AIGLE's features:

1. **Welcome**: Introduction to AIGLE and tour overview
2. **Three-Layer Model**: Explanation of the nested governance architecture
3. **NIST AI RMF Layer**: Overview of the outer strategic layer
4. **Responsible AI Lifecycle**: Overview of the middle operational layer
5. **Model Development Lifecycle**: Overview of the inner technical layer
6. **Cross-Cutting Elements**: Introduction to elements that span all layers
7. **Interactive Elements**: How to click and explore governance elements
8. **Element Details**: Understanding the information in detail panels
9. **Trust Lens**: Introduction to the trustworthiness overlay
10. **Trust Dimensions**: Explanation of the eight NIST trust characteristics
11. **Tool Recommendations**: How to discover and use open-source tools
12. **Maturity Assessment**: Introduction to the assessment system
13. **Assessment Process**: How to complete and interpret assessments
14. **Reports**: Generating and using assessment reports
15. **Next Steps**: Guidance on how to begin using AIGLE for your organisation

Interactive Features

The tour includes several features to enhance learning:

Spotlight Effects: Visual highlighting draws attention to the specific UI element being discussed in each step.

Automated Positioning: The tour automatically scrolls and positions the view to ensure highlighted elements are visible.

Contextual Information: Each step provides clear, concise information about the feature being introduced.

Visual Indicators: Progress indicators show users where they are in the tour sequence.

Examples: Where appropriate, steps include concrete examples to illustrate concepts.

Navigation Options

Users have multiple ways to navigate the tour:

Next/Back Buttons: Step forward or backward through the tour sequence.

Keyboard Shortcuts: Arrow keys for next/back, Escape to exit tour.

Skip Option: Users can skip the tour if they prefer to explore independently.

Resume Capability: If users exit the tour, they can resume from where they left off.

Restart Option: Users can restart the tour at any time from the beginning.

Educational Context

Each tour step provides educational context that helps users understand not just how to use AIGLE, but why AI governance matters:

Governance Concepts: Brief explanations of key governance concepts as they're introduced.

Real-World Relevance: Examples of why specific governance activities matter in practise.

Best Practices: Tips on how to apply governance concepts in organisational contexts.

Common Challenges: Acknowledgment of typical challenges organisations face in implementing governance.

Accessibility

The tour system is designed to be accessible:

Keyboard Navigation: Full keyboard support for all tour controls.

Screen Reader Compatible: Tour content is accessible to screen readers.

Adjustable Pace: Users control the pace, with no time limits on steps.

Dismissible: Users can exit the tour at any time without losing access to features.

Detailed Element Descriptions

This section provides comprehensive descriptions of each governance element in AIGLE. Elements are organised by layer and presented in the order users typically encounter them in the AI lifecycle.

NIST AI RMF Layer Elements

GOVERN

Purpose: Establish the organisational culture, structures, policies, and processes necessary for effective AI risk management across the enterprise.

Key Activities: - Define AI governance roles, responsibilities, and accountability structures - Establish AI risk appetite and tolerance levels aligned with organisational values - Create policies and procedures for AI development, deployment, and use - Implement governance mechanisms (committees, review boards, escalation paths) - Allocate resources for AI risk management activities - Foster a culture of responsible AI through training and awareness - Integrate AI governance with enterprise risk management - Establish metrics for governance effectiveness

Artefacts Produced: - AI governance charter and policy framework - Roles and responsibilities matrix (RACI) - Risk appetite statement - Governance committee charters - Training and awareness materials - Resource allocation plans - Integration plans with enterprise risk management - Governance metrics and dashboards

Risks Addressed: - Lack of accountability for AI outcomes - Inconsistent risk management across AI initiatives - Insufficient resources for governance activities - Cultural resistance to responsible AI practices - Unclear decision-making authority - Misalignment between AI initiatives and organisational values - Inadequate executive oversight of AI risks

Common Pitfalls: - Creating governance structures that are too bureaucratic and slow innovation - Failing to secure executive sponsorship and adequate resources - Treating governance as a compliance checkbox rather than strategic enabler - Not adapting governance structures as AI capabilities and risks evolve - Insufficient integration with existing enterprise risk management - Focusing on policies without ensuring implementation and enforcement - Not measuring governance effectiveness

Trust Characteristics: Accountable & Transparent, Secure & Resilient

Recommended Tools: - **VerifyWise**: Comprehensive AI governance platform for documenting and auditing AI governance processes, integrating policy, tools, roles, and reporting to ensure alignment with ethical standards and regulations - **OpenRMF**: Web-based compliance automation tool for cyber and AI risk management, supporting NIST AI RMF compliance reporting, STIG checklists, and generation of governance artifacts like POAMs and risk dashboards

Note: The GOVERN element focuses on organizational governance platforms rather than technical implementation tools. Policy enforcement tools (like OPA, Kyverno) are mapped to MANAGE/DEPLOYMENT elements where they implement technical controls.

MAP

Purpose: Understand the context in which AI systems will operate, identify relevant risks, and map them to potential impacts on individuals, organisations, and society.

Key Activities: - Identify and document AI system context (purpose, users, environment) - Catalog relevant legal, regulatory, and ethical requirements - Map stakeholders and their interests/concerns - Identify potential positive and negative impacts - Assess risk categories (bias, privacy, safety, security, etc.) - Document assumptions and limitations - Establish risk categorisation and prioritisation criteria - Conduct initial impact assessments

Artefacts Produced: - Context documentation and use case descriptions - Stakeholder analysis and engagement plans - Requirements traceability matrix - Risk register with categorised risks - Impact assessments (privacy, equity, safety) - Assumptions and limitations log - Risk categorisation framework - Initial risk heat maps

Risks Addressed: - Deploying AI in inappropriate contexts - Missing critical stakeholder perspectives - Overlooking relevant regulatory requirements - Failing to anticipate negative impacts - Inadequate understanding of system limitations - Misalignment between AI system and organisational values - Insufficient consideration of societal impacts

Common Pitfalls: - Conducting mapping as a one-time activity rather than ongoing process - Focusing only on technical risks while ignoring social/ethical dimensions - Insufficient stakeholder engagement, particularly with affected communities - Treating all risks as equal rather than prioritising based on severity and likelihood - Documenting risks without connecting them to mitigation strategies - Not updating risk maps as context changes - Inadequate consideration of cumulative and systemic risks

Trust Characteristics: Accountable & Transparent, Fair with Harmful Bias Managed, Privacy-Enhanced

Recommended Approach (No Software Tools):

The MAP element has **no software tools** because its activities are fundamentally **methodology-based** and require human expertise:

Why No Tools: - **Stakeholder mapping** requires facilitation, interviews, and workshops - **Context understanding** demands cultural awareness and judgment - **Impact assessment** needs ethical reasoning and domain expertise - **Requirements gathering** involves negotiation and consensus-building

How to Approach MAP: 1. **Conduct stakeholder mapping workshops** with diverse participants using matrices and influence maps 2. **Use interviews and surveys** to engage directly with affected stakeholders and gather qualitative data 3. **Apply ethical impact assessment frameworks** manually (IEEE 7000, ALTAI, DPIA) with ethics review panels 4. **Document findings** in organizational governance platforms (like VerifyWise) or standard documentation tools

Note: Tools like AI Fairness 360 and Fairlearn measure bias quantitatively (MEASURE element), but they don't help identify stakeholders or understand context (MAP element).

The distinction between discovery (MAP) and measurement (MEASURE) is intentional and important.

MEASURE

Purpose: Quantify and qualify AI system performance, trustworthiness characteristics, and risk levels through systematic measurement and testing.

Key Activities: - Define metrics for trustworthiness characteristics (fairness, robustness, etc.) - Establish measurement methodologies and testing protocols - Implement continuous monitoring and measurement systems - Conduct regular assessments against defined criteria - Benchmark performance against industry standards - Track metrics over time to identify trends - Validate measurement approaches for reliability and validity - Report measurement results to stakeholders

Artefacts Produced: - Metrics framework and definitions - Testing and evaluation protocols - Measurement results and scorecards - Benchmark comparisons - Trend analysis reports - Validation studies for measurement approaches - Stakeholder reports and dashboards

Risks Addressed: - Inability to detect performance degradation or emerging issues - Lack of objective evidence for trustworthiness claims - Insufficient visibility into system behaviour - Failure to identify bias or fairness issues - Inadequate basis for risk-informed decisions - Inability to demonstrate compliance with requirements - Missing early warning signs of problems

Common Pitfalls: - Measuring only technical performance while ignoring trustworthiness dimensions - Using metrics that are easy to measure rather than meaningful - Failing to establish baselines and thresholds for action - Not validating that metrics actually measure what they claim to measure - Collecting data without analysing it or acting on insights - Insufficient frequency of measurement - Not considering measurement limitations and potential gaming

Trust Characteristics: Valid & Reliable, Fair with Harmful Bias Managed, Explainable & Interpretable

Recommended Tools: - Evidently: ML model monitoring and testing - DeepChecks: Comprehensive model testing - AI Fairness 360: Fairness metrics and assessment - WhyLabs: ML monitoring and observability - Great Expectations: Data quality metrics

MANAGE

Purpose: Implement risk treatment strategies, respond to identified issues, and continuously improve risk management practices based on measurement insights.

Key Activities: - Develop and implement risk treatment plans (mitigate, transfer, accept, avoid) - Establish incident response and escalation procedures - Implement controls and safeguards - Monitor control effectiveness - Respond to identified issues and incidents - Document decisions and rationale - Continuously improve risk management based on lessons learned - Communicate risk management activities to stakeholders

Artefacts Produced: - Risk treatment plans and control documentation - Incident response procedures and playbooks - Control effectiveness reports - Incident logs and post-incident reviews - Decision logs with rationale - Lessons learned and improvement plans - Stakeholder communications

Risks Addressed: - Inadequate response to identified risks - Ineffective risk controls - Slow or inappropriate incident response - Failure to learn from issues and improve - Lack of documentation for accountability - Insufficient stakeholder communication about risk management - Inability to demonstrate due diligence

Common Pitfalls: - Implementing controls without verifying effectiveness - Treating risk management as separate from development rather than integrated - Failing to update risk treatments as context changes - Not establishing clear thresholds for escalation and response - Inadequate documentation of risk decisions and rationale - Slow incident response due to unclear procedures - Not learning from incidents and near-misses

Trust Characteristics: Safe, Secure & Resilient, Accountable & Transparent

Recommended Tools: - **Open Policy Agent (OPA)**: Policy-based control for microservices, Kubernetes, CI/CD pipelines, and API gateways, enabling declarative policy enforcement across the AI system lifecycle - **Kyverno**: Kubernetes-native policy management for validating, mutating, and generating configurations, ensuring compliance and security in containerized AI deployments - **Prometheus**: Open-source monitoring and alerting toolkit for tracking control effectiveness and system health metrics in real-time - **OpenRMF**: Risk management and compliance tracking (also in GOVERN for organizational governance) - **Evidently**: Monitor control effectiveness through ML model monitoring - **Garak**: Test LLM vulnerabilities and validate security controls

Responsible AI Lifecycle Elements

Problem Framing

Purpose: Define the problem to be solved, determine whether AI is appropriate, and establish success criteria before committing resources to development.

Key Activities: - Articulate the problem clearly and specifically - Assess whether AI is necessary and appropriate for the problem - Identify alternative approaches (including non-AI solutions) - Define success criteria and metrics - Conduct initial ethical and risk screening - Engage stakeholders in problem definition - Document assumptions and constraints - Establish project scope and boundaries

Artefacts Produced: - Problem statement and justification - AI appropriateness assessment - Success criteria and metrics definition - Initial risk screening results - Stakeholder input documentation - Assumptions and constraints log - Project charter and scope document

Risks Addressed: - Applying AI to problems better solved by other means - Poorly defined problems leading to inappropriate solutions - Misalignment between technical solution and actual need - Overlooking ethical concerns early when changes are easier

- Insufficient stakeholder input leading to rejected solutions - Scope creep and mission drift - Unrealistic expectations about AI capabilities

Common Pitfalls: - Rushing to AI solutions without considering alternatives - Defining problems too narrowly or too broadly - Focusing on technical feasibility without considering social implications - Not involving affected stakeholders in problem framing - Failing to establish clear, measurable success criteria - Treating problem framing as one-time activity rather than iterative - Not documenting assumptions that may need revisiting

Trust Characteristics: Accountable & Transparent, Valid & Reliable

Recommended Tools: - Model Card Toolkit: Document problem framing and intended use - Great Expectations: Validate assumptions about data availability

Data Collection & Preparation

Purpose: Acquire, clean, and prepare data for model development while ensuring quality, representativeness, privacy, and appropriate documentation.

Key Activities: - Identify data sources and assess availability - Evaluate data quality, completeness, and representativeness - Implement privacy-preserving data collection methods - Clean and preprocess data - Address missing data and outliers - Document data provenance and lineage - Assess and mitigate data bias - Implement data governance and access controls - Create data documentation (datasheets)

Artefacts Produced: - Data collection plan and protocols - Data quality assessment reports - Data documentation (datasheets, data cards) - Privacy impact assessments - Bias analysis reports - Data lineage documentation - Data governance policies - Cleaned and prepared datasets

Risks Addressed: - Poor data quality leading to unreliable models - Biased data perpetuating or amplifying discrimination - Privacy violations through inappropriate data collection or use - Lack of data representativeness limiting generalization - Insufficient documentation hindering reproducibility and accountability - Data leakage and security breaches - Inability to trace data provenance

Common Pitfalls: - Assuming available data is appropriate without critical evaluation - Failing to document data limitations and biases - Inadequate privacy protections, especially for sensitive data - Not considering data representativeness across relevant subgroups - Insufficient data governance leading to inappropriate access or use - Removing outliers without understanding their meaning - Not maintaining data lineage and provenance

Trust Characteristics: Privacy-Enhanced, Fair with Harmful Bias Managed, Valid & Reliable

Recommended Tools: - Great Expectations: Data quality validation - TensorFlow Data Validation: Data validation for ML pipelines - Pandera: Statistical data validation - Deequ: Data quality validation at scale - Aequitas: Bias analysis in datasets - DVC: Data versioning and lineage - Opacus: Differential privacy for data

Feature Engineering

Purpose: Transform raw data into features (input variables) that effectively represent the problem and enable model learning.

Key Activities: - Analyze data characteristics and relationships - Create new features through transformation and combination - Select relevant features and remove redundant ones - Encode categorical variables appropriately - Normalize and scale features - Handle temporal and spatial aspects - Document feature definitions and rationale - Assess feature fairness implications - Validate features make domain sense

Artefacts Produced: - Feature definitions and documentation - Feature engineering code and pipelines - Feature importance analysis - Feature fairness assessments - Transformation and encoding specifications - Feature validation results

Risks Addressed: - Poor model performance due to inadequate feature representation - Introduction of bias through feature selection or engineering - Lack of interpretability due to opaque feature transformations - Data leakage through inappropriate feature construction - Inability to reproduce results due to undocumented features - Features that don't generalise to deployment contexts

Common Pitfalls: - Creating features that leak information from the target variable - Not considering fairness implications of feature choices - Over-engineering features leading to overfitting - Insufficient documentation of feature definitions and rationale - Not validating that engineered features make domain sense - Using features that won't be available at inference time - Not assessing feature stability over time

Trust Characteristics: Valid & Reliable, Fair with Harmful Bias Managed, Explainable & Interpretable

Recommended Tools: - SHAP: Analyze feature importance and interactions - AI Fairness 360: Assess fairness implications of features - Great Expectations: Validate feature distributions - DVC: Version control for feature engineering code

Model Training

Purpose: Use prepared data and engineered features to train AI models by optimising model parameters to minimise prediction errors.

Key Activities: - Select appropriate model architectures and algorithms - Configure hyperparameters - Split data into training, validation, and test sets - Train models using appropriate optimisation techniques - Implement regularization to prevent overfitting - Use cross-validation for robust performance estimation - Document training procedures and configurations - Track experiments and model versions - Assess training data representativeness

Artefacts Produced: - Trained model artifacts and weights - Training configuration documentation - Experiment tracking logs - Hyperparameter tuning results - Cross-validation performance metrics - Model versioning records - Training data documentation

Risks Addressed: - Overfitting to training data, poor generalization - Suboptimal

model performance due to poor hyperparameter choices - Inability to reproduce training results - Lack of transparency in model development process - Inefficient use of computational resources - Training on biased or unrepresentative data samples

Common Pitfalls: - Not using proper train/validation/test splits, leading to overoptimistic performance estimates - Insufficient hyperparameter tuning - Training on biased or unrepresentative data samples - Not tracking experiments systematically - Optimizing for single metrics without considering trade-offs - Not documenting random seeds and other factors affecting reproducibility - Insufficient regularization leading to overfitting

Trust Characteristics: Valid & Reliable, Explainable & Interpretable

Recommended Tools: - MLflow: Experiment tracking and model versioning - Weights & Biases: Experiment tracking and hyperparameter tuning - DVC: Data and model versioning - TensorBoard: Training visualisation - Optuna: Hyperparameter optimisation

Model Evaluation

Purpose: Assess trained model performance across multiple dimensions including accuracy, fairness, robustness, and other trustworthiness characteristics.

Key Activities: - Evaluate performance on held-out test data - Assess performance across demographic subgroups - Test fairness using multiple fairness metrics - Evaluate robustness to input perturbations - Assess calibration and uncertainty quantification - Analyze errors and failure modes - Compare to baseline and alternative models - Document evaluation results and limitations - Test for spurious correlations

Artefacts Produced: - Performance evaluation reports - Fairness assessment results - Robustness testing results - Error analysis documentation - Model comparison reports - Limitation and failure mode documentation - Calibration analysis

Risks Addressed: - Deploying models with inadequate performance - Undetected bias and fairness issues - Lack of robustness to real-world variations - Poor calibration leading to overconfident predictions - Insufficient understanding of model limitations - Models that exploit spurious correlations - Disparate performance across subgroups

Common Pitfalls: - Evaluating only on overall metrics without subgroup analysis - Using single fairness metrics without considering trade-offs - Insufficient robustness testing - Not analysing errors to understand failure modes - Treating evaluation as one-time activity rather than iterative process - Not testing calibration and uncertainty quantification - Insufficient comparison to baselines and alternatives

Trust Characteristics: Valid & Reliable, Fair with Harmful Bias Managed, Explainable & Interpretable

Recommended Tools: - AI Fairness 360: Comprehensive fairness metrics - Fairlearn: Fairness assessment and mitigation - Aequitas: Bias and fairness audit - DeepChecks: Model testing and validation - SHAP: Model explanation and analysis - LIME: Local model explanations - Evidently: Model evaluation and testing

Model Validation

Purpose: Verify that models meet requirements and perform acceptably in realistic deployment conditions before production release.

Key Activities: - Test models in production-like environments - Validate against real-world data and scenarios - Conduct user acceptance testing - Verify compliance with requirements and constraints - Assess operational feasibility and resource requirements - Test integration with surrounding systems - Validate monitoring and alerting systems - Obtain stakeholder sign-off for deployment - Conduct red-teaming and adversarial testing

Artefacts Produced: - Validation test plans and results - User acceptance testing documentation - Requirements compliance verification - Integration test results - Operational readiness assessment - Stakeholder approval documentation - Red-team testing results

Risks Addressed: - Models that perform well in development but fail in production - Unmet requirements discovered after deployment - Inadequate operational support infrastructure - Stakeholder rejection of deployed systems - Compliance violations - Security vulnerabilities - Inadequate monitoring capabilities

Common Pitfalls: - Validating only on historical data without realistic deployment scenarios - Insufficient user acceptance testing - Not validating operational aspects (latency, resource usage, etc.) - Treating validation as rubber-stamp rather than critical gate - Inadequate documentation of validation results and decisions - Not testing edge cases and failure modes - Insufficient adversarial and security testing

Trust Characteristics: Valid & Reliable, Safe, Secure & Resilient

Recommended Tools: - Garak: LLM vulnerability scanning - Adversarial Robustness Toolbox: Adversarial testing - DeepChecks: Validation testing - Evidently: Production readiness testing - Great Expectations: Data validation in production

Deployment & Integration

Purpose: Transition validated models into production environments, integrate with existing systems, and establish operational procedures.

Key Activities: - Plan deployment architecture and infrastructure - Implement model serving and inference systems - Integrate with upstream and downstream systems - Establish operational procedures and runbooks - Implement monitoring and alerting systems - Conduct deployment testing and validation - Train operational staff - Implement access controls and security measures - Plan rollback and contingency procedures - Conduct phased rollout if appropriate

Artefacts Produced: - Deployment architecture documentation - Integration specifications - Operational procedures and runbooks - Deployment test results - Training materials for operational staff - Security and access control documentation - Rollback and contingency plans - Deployment approval documentation

Risks Addressed: - System failures during deployment - Integration issues with existing systems - Inadequate operational support leading to poor performance - Security vulnerabilities in production environment - Lack of preparedness for issues requiring rollback - Insufficient monitoring leading to undetected issues - Inadequate access controls

Common Pitfalls: - Insufficient testing in production-like environments before deployment - Inadequate operational documentation and training - Not implementing proper monitoring from day one - Failing to plan for rollback and incident response - Treating deployment as the end rather than beginning of operational lifecycle - Not conducting phased rollout to limit risk - Insufficient security hardening for production

Trust Characteristics: Secure & Resilient, Safe, Accountable & Transparent

Recommended Tools: - MLflow: Model deployment and serving - Kubeflow: ML workflows and deployment on Kubernetes - TensorFlow Serving: Model serving infrastructure - Evidently: Deployment monitoring - Prometheus: Infrastructure monitoring

Monitoring & Maintenance

Purpose: Continuously monitor deployed AI systems, detect issues, maintain performance, and ensure ongoing trustworthiness.

Key Activities: - Monitor system performance and behaviour continuously - Track trustworthiness metrics (fairness, robustness, etc.) - Detect data drift and model degradation - Respond to alerts and incidents - Perform regular maintenance and updates - Collect feedback from users and stakeholders - Assess ongoing compliance with requirements - Document operational history and issues - Conduct periodic reviews and audits

Artefacts Produced: - Monitoring dashboards and reports - Incident logs and response documentation - Performance trend analysis - Drift detection reports - Maintenance logs - User feedback summaries - Compliance audit trails - Periodic review reports

Risks Addressed: - Undetected performance degradation - Model behaviour drift due to changing data distributions - Emerging fairness or bias issues in production - Security incidents and adversarial attacks - Non-compliance with evolving requirements - User dissatisfaction due to poor performance - Inability to diagnose and resolve issues

Common Pitfalls: - Monitoring only technical metrics without trustworthiness dimensions - Slow response to detected issues - Insufficient resources allocated to ongoing monitoring - Not collecting and acting on user feedback - Treating monitoring as automated rather than requiring human judgment - Not monitoring for data drift and distribution shift - Inadequate incident response procedures

Trust Characteristics: Valid & Reliable, Safe, Secure & Resilient, Fair with Harmful Bias Managed

Recommended Tools: - Evidently: ML monitoring and drift detection - WhyLabs: ML observability - Fiddler: ML monitoring and explainability - Arize: ML observability platform - Prometheus: Infrastructure monitoring - Grafana: Monitoring dashboards

Impact Review & Iteration

Purpose: Periodically assess actual impacts of deployed AI systems, compare to intended outcomes, and determine whether to continue, modify, or retire systems.

Key Activities: - Conduct comprehensive impact assessments - Compare actual outcomes to intended goals and success criteria - Assess unintended consequences and emergent behaviours - Gather stakeholder feedback on system impacts - Evaluate continued appropriateness and value - Make decisions about system continuation, modification, or retirement - Document lessons learned - Feed insights back into governance and future development - Update risk assessments based on operational experience

Artefacts Produced: - Impact assessment reports - Stakeholder feedback summaries - Outcome vs. goal comparison analysis - Continuation/modification/retirement decisions - Lessons learned documentation - Recommendations for governance improvements - Updated risk assessments

Risks Addressed: - Continued operation of systems that no longer serve their purpose - Unaddressed negative impacts on stakeholders - Failure to learn from experience and improve practices - Misalignment between system operation and organisational values - Missed opportunities to optimise or enhance systems - Accumulation of technical debt - Erosion of stakeholder trust

Common Pitfalls: - Conducting reviews too infrequently or superficially - Focusing only on technical performance without assessing broader impacts - Insufficient stakeholder engagement in impact assessment - Not acting on review findings (continuing problematic systems) - Failing to feed lessons learned back into governance and development practices - Not considering cumulative and systemic impacts - Treating reviews as compliance exercise rather than genuine learning opportunity

Trust Characteristics: Accountable & Transparent, Fair with Harmful Bias Managed, Valid & Reliable

Recommended Tools: - Evidently: Long-term performance analysis - Model Card Toolkit: Update model documentation based on operational experience - MLflow: Track system evolution and decisions

Cross-Cutting Elements

Documentation & Transparency

Purpose: Maintain comprehensive, accessible documentation of AI systems, decisions, and processes to enable transparency, accountability, and reproducibility.

Spans All Phases: Documentation is required throughout the entire AI lifecycle, from initial problem framing through ongoing operation.

Key Activities: - Document system purpose, context, and intended use - Record design decisions and rationale - Maintain data and model documentation (datasheets, model cards) - Document testing and evaluation results - Record operational procedures and incidents - Create user-facing documentation and disclosures - Maintain audit trails of key decisions - Update documentation as systems evolve

Artefacts: - System documentation (purpose, context, use cases) - Data documentation (datasheets, data cards) - Model documentation (model cards, technical specifications) - Decision logs and rationale - User documentation and disclosures - Audit trails and compliance records - Operational documentation and runbooks

Risks Addressed: - Lack of transparency hindering trust and accountability - Inability to reproduce results or understand system behaviour - Insufficient information for stakeholders to make informed decisions - Difficulty diagnosing and resolving issues - Compliance violations due to inadequate documentation - Knowledge loss when team members change - Inability to demonstrate due diligence

Common Pitfalls: - Treating documentation as afterthought rather than ongoing activity - Creating documentation that is too technical for non-technical stakeholders - Not updating documentation as systems evolve - Documenting what was done without explaining why - Insufficient detail to enable reproducibility - Not making documentation accessible to those who need it - Focusing on compliance documentation while neglecting operational documentation

Trust Characteristics: Accountable & Transparent, Explainable & Interpretable

Recommended Tools: - Model Card Toolkit: Generate model cards - Datasheets for Datasets: Create dataset documentation - DVC: Version control for documentation - MLflow: Track and document experiments - VerifyML: Model documentation and validation

Human Oversight & Control

Purpose: Ensure appropriate human involvement in AI system development, deployment, and operation to maintain accountability and enable intervention.

Spans All Phases: Human oversight is required at all stages, from governance through ongoing monitoring.

Key Activities: - Define human roles and responsibilities - Implement human-in-the-loop mechanisms where appropriate - Establish human review and approval gates - Enable human override and intervention capabilities - Train humans to effectively oversee AI systems - Monitor and support human decision-makers - Maintain human accountability for AI outcomes - Design appropriate levels of automation

Artefacts: - Human oversight procedures and protocols - Roles and responsibilities documentation - Training materials for human overseers - Intervention and override procedures - Human review and approval records - Automation level assessments - Human factors analysis

Risks Addressed: - Lack of accountability for AI outcomes - Inability to intervene when AI systems behave inappropriately - Over-reliance on AI leading to automation bias - Inadequate human understanding of AI system behaviour - Deskilling of human decision-makers - Inappropriate levels of automation - Unclear responsibility for decisions

Common Pitfalls: - Treating humans as mere rubber-stamps for AI decisions - Not providing humans with sufficient information and tools to effectively oversee AI - Im-

plementing human oversight that is too burdensome and gets bypassed - Not training humans to understand AI capabilities and limitations - Failing to design for appropriate levels of automation - Not monitoring for automation bias and over-reliance - Unclear accountability when humans and AI collaborate

Trust Characteristics: Accountable & Transparent, Safe, Secure & Resilient

Recommended Tools: - What-If Tool: Enable human exploration of model behaviour - InterpretML: Provide explanations to support human oversight - SHAP: Help humans understand model decisions - LIME: Local explanations for human review

Stakeholder Engagement

Purpose: Involve relevant stakeholders throughout the AI lifecycle to incorporate diverse perspectives, build trust, and ensure systems serve stakeholder needs.

Spans All Phases: Stakeholder engagement is required from problem framing through impact review.

Key Activities: - Identify relevant stakeholders (users, affected parties, domain experts, etc.) - Engage stakeholders in problem framing and requirements definition - Gather stakeholder input on design decisions - Conduct user testing and gather feedback - Communicate with stakeholders about system capabilities and limitations - Address stakeholder concerns and incorporate feedback - Maintain ongoing stakeholder relationships - Ensure diverse stakeholder representation

Artefacts: - Stakeholder analysis and engagement plans - Stakeholder input documentation - User testing and feedback summaries - Communication materials and disclosures - Stakeholder concern tracking and resolution - Engagement activity logs - Stakeholder satisfaction assessments

Risks Addressed: - Systems that don't meet stakeholder needs - Missing important perspectives leading to blind spots - Stakeholder rejection of deployed systems - Unintended negative impacts on affected communities - Erosion of trust due to lack of engagement - Failure to identify and address concerns early - Lack of diverse perspectives leading to bias

Common Pitfalls: - Engaging only convenient stakeholders, missing affected communities - Treating engagement as one-time activity rather than ongoing - Not acting on stakeholder feedback - Insufficient engagement with diverse stakeholders - Engaging stakeholders too late when changes are difficult - Not communicating clearly about AI capabilities and limitations - Treating engagement as public relations rather than genuine collaboration

Trust Characteristics: Accountable & Transparent, Fair with Harmful Bias Managed

Recommended Tools: - Model Card Toolkit: Create stakeholder-facing documentation - What-If Tool: Enable stakeholder exploration of model behaviour - Aequitas: Facilitate stakeholder discussions about fairness

Continuous Learning & Adaptation

Purpose: Systematically learn from experience and adapt governance and development practices based on new knowledge, changing contexts, and lessons learned.

Spans All Phases: Learning and adaptation should occur throughout the lifecycle and feed back into future iterations.

Key Activities: - Collect and analyse operational data and feedback - Conduct regular reviews and retrospectives - Identify lessons learned and improvement opportunities - Update practices, policies, and procedures based on insights - Stay current with evolving standards, regulations, and best practices - Share knowledge across teams and projects - Foster culture of continuous improvement - Experiment with new approaches and tools

Artefacts: - Lessons learned documentation - Practice and policy updates - Retrospective and review reports - Knowledge sharing materials - Improvement tracking and implementation plans - Experiment results and recommendations - Best practise documentation

Risks Addressed: - Repeating mistakes across projects - Failure to improve practices over time - Falling behind evolving standards and best practices - Inability to adapt to changing contexts and requirements - Knowledge silos limiting organisational learning - Stagnation and complacency - Missed opportunities for innovation

Common Pitfalls: - Not allocating time for learning and improvement activities - Conducting retrospectives without acting on findings - Failing to share lessons learned across teams - Not staying current with evolving standards and regulations - Treating learning as individual rather than organisational activity - Not experimenting with new approaches - Focusing only on technical learning while ignoring governance lessons

Trust Characteristics: Valid & Reliable, Secure & Resilient, Accountable & Transparent

Recommended Tools: - MLflow: Track experiments and learnings - Weights & Biases: Experiment tracking and comparison - DVC: Version control for evolving practices - Evidently: Analyze trends and patterns for learning

User Guide

Getting Started

Accessing AIGLE

AIGLE is a web-based application accessible through any modern web browser. To access AIGLE:

1. Navigate to **aigle.datadid.io** in your web browser
2. The application loads directly without requiring login or registration
3. AIGLE works best on desktop and tablet devices with screen widths of 768px or greater
4. Mobile access is supported but some features may be optimised for larger screens

Browser Compatibility: AIGLE is compatible with: - Chrome/Edge (version 90+) - Firefox (version 88+) - Safari (version 14+) - Other modern browsers supporting ES6 and CSS Grid

No Installation Required: AIGLE is a progressive web application that runs entirely in your browser. No software installation or downloads are required.

Navigation Basics

AIGLE's interface is designed for intuitive navigation:

Main Diagram: The central interactive diagram occupies the majority of the screen and serves as the primary navigation interface.

Top Bar: Contains the AIGLE logo, title, and action buttons: - **How to Use**: Toggle button to show/hide the instructions panel - **View Report**: View your maturity assessment report (appears after completing assessments) - **Take a Tour**: Begin the guided tour of the platform

Floating Controls (Bottom-Right): - **Trust Lens Toggle**: Floating button to activate/deactivate the trustworthiness overlay - **Trust Characteristics Legend**: Collapsible panel showing the eight NIST trust dimensions (appears when Trust Lens is active)

Side Panel: Opens on the right side when you click a governance element, displaying detailed information.

Footer: Contains links to: - Book a Consultation (datadid.io) - Visit Website (okosieme.org) - Privacy Policy - Contact Information

Understanding the Interface

Color Coding: Different colours represent different layers and element types: - Blue tones: Strategic governance (NIST AI RMF) - Green tones: Operational lifecycle (Responsible AI Lifecycle) - Grey tones: Technical development (Model Development Lifecycle) - Accent colours: Cross-cutting elements

Visual Hierarchy: The nested circular design shows relationships: - Outer elements provide strategic context - Middle elements show operational flow - Inner elements detail technical iteration - Cross-cutting elements span all layers

Interactive Indicators: Visual cues show interactivity: - Hover effects: Elements highlight when you move your cursor over them - Cursor changes: Pointer cursor indicates clickable elements - Selected state: Active elements remain highlighted

Exploring Elements

Clicking and Selecting

To explore a governance element:

1. **Hover** over any element in the diagram to see it highlight
2. **Click** the element to open its detail panel
3. The element remains highlighted while its panel is open
4. **Click another element** to switch to that element's details
5. **Click the X button** or **click outside the panel** to close it

Container Elements: The three large background areas (outer ring, middle flow, inner cycle) are also clickable and provide high-level overviews of each layer.

Cross-Cutting Elements: The four cross-cutting elements (Documentation, Human Oversight, Stakeholder Engagement, Continuous Learning) can be clicked to see how they apply across all phases.

Reading Element Details

When you click an element, the side panel displays comprehensive information:

Element Name and Layer: Header shows which element you're viewing and which layer it belongs to.

Purpose: Clear statement of why this governance activity matters and what it aims to achieve.

Key Activities: Bulleted list of specific activities to perform as part of this governance element.

Artefacts Produced: List of documents, records, and outputs that should result from these activities.

Risks Addressed: Explanation of what risks this governance element helps mitigate.

Common Pitfalls: Warning about typical mistakes organisations make with this element and how to avoid them.

Trust Characteristics: Indication of which NIST trustworthiness dimensions this element primarily supports.

Recommended Tools: Curated list of open-source tools that can help implement this governance element, with: - Tool name and brief description - Links to

GitHub repository, documentation, and website - License information - Programming language/framework

Understanding Relationships

AIGLE helps you understand how governance elements relate to each other:

Sequential Flow: Elements in the Responsible AI Lifecycle (middle layer) flow sequentially, showing the typical progression of AI projects.

Iterative Cycle: Elements in the Model Development Lifecycle (inner layer) form an iterative cycle that repeats multiple times.

Strategic Oversight: Elements in the NIST AI RMF (outer layer) provide ongoing strategic governance throughout the lifecycle.

Cross-Cutting Connections: The four cross-cutting elements connect to all other elements, indicating activities that span the entire lifecycle.

Trust Lens Mapping: When activated, the Trust Lens shows which elements contribute to specific trustworthiness characteristics.

Using the Trust Lens

Activating the Overlay

To activate the Trust Lens:

1. Locate the **“Trust Lens”** floating button in the bottom-right corner of the screen
2. Click the button to activate the overlay
3. The button turns purple and displays “Trust Lens: ON”
4. The diagram transforms to show trust dimension mappings
5. Elements glow with colours corresponding to trust characteristics they support
6. A collapsible legend panel appears above the button showing all eight trust characteristics

To deactivate: 1. Click the **“Trust Lens”** button again 2. The button returns to grey and displays “Trust Lens: OFF” 3. The legend panel disappears and the diagram returns to its standard view

Interpreting Color Codes

Each of the eight NIST trustworthiness characteristics is represented by a distinct colour:

- **Blue**: Valid & Reliable - Systems perform consistently and accurately
- **Green**: Safe - Systems don’t pose unreasonable safety risks
- **Purple**: Secure & Resilient - Systems are protected and can recover from disruptions
- **Orange**: Accountable & Transparent - Systems enable accountability and transparency
- **Yellow**: Explainable & Interpretable - Systems provide appropriate explanations

- **Pink**: Privacy-Enhanced - Systems protect privacy and data confidentiality
- **Teal**: Fair with Harmful Bias Managed - Systems avoid harmful bias and promote fairness
- **Lime**: Environmentally Sustainable - Systems minimise environmental impact (future)

Multiple Colors: Some elements contribute to multiple trust dimensions and may show blended colours or multiple indicators.

Intensity: The intensity of the glow effect indicates the strength of the element's contribution to that trust dimension.

Understanding Dimension Mappings

The Trust Lens helps you understand:

Coverage: Which trust dimensions are well-covered by your governance activities and which may need more attention.

Connections: How specific governance activities contribute to trustworthiness outcomes.

Priorities: Which elements to focus on if you're particularly concerned about specific trust dimensions.

Gaps: Areas where additional governance activities may be needed to achieve comprehensive trustworthiness.

Use Cases: - **Gap Analysis**: Identify which trust dimensions lack sufficient governance coverage - **Prioritization**: Focus on elements that address your most critical trust concerns - **Communication**: Explain to stakeholders how governance translates to trustworthy outcomes - **Education**: Learn about the multidimensional nature of AI trustworthiness

Taking Maturity Assessments

Starting an Assessment

To begin a maturity assessment:

1. Click on any governance element in the diagram to open its detailed information panel
2. Scroll down in the panel and click the **"Begin Assessment"** button
3. The assessment interface opens with targeted questions specific to that element
4. Answer the questions to evaluate your organisation's maturity in that area

Alternatively, you can start a comprehensive assessment covering all elements by clicking the **"Begin Assessment"** button in the instructions panel (if available).

Time Commitment: The full assessment typically takes 30-45 minutes to complete thoughtfully.

Saving Progress: Your responses are saved automatically as you progress, so you can pause and resume later.

Answering Questions

The assessment presents 60+ questions organised by governance element:

Question Format: Each question presents a statement about a governance practise, and you select the response that best describes your organisation's current state.

Response Options: Five options corresponding to maturity levels: - **Level 0**: Practice is absent or ad hoc - **Level 1**: Aware of need, initial activities begun - **Level 2**: Formal processes defined and documented - **Level 3**: Processes well-integrated and monitored - **Level 4**: Processes continuously optimised

Honest Assessment: For accurate results, answer based on your organisation's actual current state, not aspirational goals.

Context Consideration: Consider your organisation's specific context when answering. What constitutes "mature" governance may vary by organisation size, industry, and AI use cases.

Don't Know: If you're unsure about a particular area, select the most conservative (lower maturity) option or skip the question.

Navigation: - **Next**: Move to the next question - **Previous**: Return to previous questions to review or change answers - **Progress Indicator**: Shows how many questions you've completed

Viewing Results

After completing the assessment, you'll see comprehensive results:

Overall Maturity Score: Your aggregate maturity level across all governance elements (0-4 scale).

Layer Scores: Separate scores for: - NIST AI RMF (Strategic Governance) - Responsible AI Lifecycle (Operational Governance) - Model Development Lifecycle (Technical Governance)

Element Scores: Individual scores for each of the 28 governance elements, showing specific strengths and gaps.

Trust Dimension Scores: Scores indicating your maturity in supporting each of the eight NIST trustworthiness characteristics.

Visual Dashboard: Charts and heat maps providing visual representation of results: - Radar chart showing scores across layers - Heat map showing element-level scores - Bar charts comparing trust dimension scores

Interpretation Guidance: Explanation of what your scores mean and typical maturity progression patterns.

Generating Reports

To generate a PDF report of your assessment results:

1. Review your results in the web interface
2. Click the **“Generate PDF Report”** button
3. The system creates a comprehensive PDF including:
    - Executive summary
    - Detailed scores and visualizations
    - Prioritized recommendations
    - Tool suggestions
    - Action planning guidance
4. The PDF downloads to your device

Report Contents: - **Executive Summary**: High-level overview suitable for leadership - **Detailed Results**: Comprehensive scores and analysis - **Visualizations**: Charts and graphs from the web interface - **Recommendations**: Prioritized improvement suggestions - **Tool Suggestions**: Relevant open-source tools for priority areas - **Action Planning**: Guidance on sequencing improvements

Customization: You can add notes or context to your report before generating it.

Email Delivery

To receive your assessment report via email:

1. After completing the assessment, click **“Email Report”**
2. Enter your email address
3. Optionally add recipient names and a message
4. Click **“Send Report”**
5. You’ll receive the PDF report via email within a few minutes

Privacy: Email addresses are used only for report delivery and are not stored or used for other purposes. See the Privacy Policy for details.

GDPR Compliance: Email handling follows GDPR requirements for data minimization and purpose limitation.

Professional Format: Emailed reports are professionally formatted and suitable for sharing with stakeholders.

Discovering Tools

Browsing Recommendations

AIGLE includes 50+ curated open-source tools for implementing AI governance. You can discover tools in several ways:

By Element: When viewing an element’s detail panel, scroll to the “Recommended Tools” section to see tools relevant to that specific governance activity. **Note**: Some elements (like MAP) have no software tools as they require methodology-based approaches—these elements show guidance on manual processes instead.

By Category: Tools are organised into functional categories: - Fairness & Bias - Explainability & Interpretability - Model Testing & Validation - Data Quality & Validation - Model Monitoring - Privacy & Security - MLOps & Governance - Documentation & Transparency

By Trust Dimension: When the Trust Lens is active, you can see which tools support specific trustworthiness characteristics.

Search: Use the search function to find tools by name, description, or capability.

Filtering by Element

To find tools for a specific governance activity:

1. Click the governance element in the diagram
2. Scroll to the "Recommended Tools" section in the detail panel
3. Review the curated list of tools relevant to that element
4. Each tool listing includes:
    - Tool name and brief description
    - Key capabilities
    - Programming language/framework
    - License type
    - Links to resources

Contextual Recommendations: Tools are specifically selected for their relevance to the governance element you're viewing.

Multiple Tools: Most elements have multiple tool recommendations, giving you options based on your technology stack and preferences.

Accessing External Resources

Each tool listing includes links to external resources:

GitHub Repository: Direct link to the tool's source code repository where you can: - Review the code - Check activity and maintenance status - Read issues and discussions - View stars and community engagement

Documentation: Link to official documentation where you can: - Learn how to install and use the tool - Review API references - Find tutorials and examples - Understand capabilities and limitations

Project Website: Link to the tool's official website (if available) for: - High-level overview - Use cases and examples - Community resources - News and updates

Opening Links: All external links open in new tabs/windows so you don't lose your place in AIGLE.

Evaluation: When evaluating tools, consider: - Active maintenance (recent commits and releases) - Community size and engagement - Documentation quality - Compatibility with your technology stack - License compatibility with your requirements - Maturity and production-readiness

Privacy and Compliance

GDPR Compliance

AIGLE is designed with privacy by default:

Data Minimization: Only essential data is collected: - Email addresses (only for report delivery, not stored) - Assessment responses (stored locally in browser) - No personal information required to use AIGLE

Purpose Limitation: Data is used only for stated purposes: - Email addresses used only for report delivery - Assessment data used only for generating results - No data used for marketing or other purposes

User Rights: Users can: - Access their assessment data (stored locally) - Delete their data (clear browser storage) - Export their data (PDF reports)

Transparency: Privacy policy clearly explains: - What data is collected - How data is used - How data is protected - User rights and choices

Data Storage

Local Storage: Assessment responses stored in browser local storage: - Data remains on user's device - Not transmitted to servers except for report generation - User can clear at any time

No User Accounts: AIGLE doesn't require user accounts: - No passwords to manage - No personal information collected - Reduced privacy risk

Report Generation: When generating reports: - Assessment data temporarily transmitted to server - Report generated and returned - Data not retained on server - Secure transmission (HTTPS)

Security

HTTPS: All traffic encrypted in transit

No Sensitive Data: AIGLE doesn't collect or store sensitive personal information

Third-Party Tools: Links to external tools are provided but AIGLE doesn't embed third-party tracking

Regular Updates: Dependencies regularly updated to address security vulnerabilities

Use Cases

AIGLE serves diverse use cases across different organisational contexts and stakeholder needs. This section illustrates how different users can leverage AIGLE to achieve their goals.

AI Governance Education

Scenario: A university professor teaching a course on AI ethics and governance wants to help students understand the practical implementation of governance frameworks.

How AIGLE Helps: - **Visual Learning**: The interactive diagram provides a visual representation of abstract governance concepts - **Comprehensive Coverage**: Students can explore all aspects of the AI lifecycle and governance - **Practical Connection**: Tool recommendations show students how governance concepts translate to real implementation - **Self-Paced Exploration**: Students can explore at their own pace, diving deep into areas of interest - **Guided Tour**: The tour provides structured introduction for students new to AI governance

Outcomes: - Students gain comprehensive understanding of AI governance frameworks - Students can connect theoretical concepts to practical tools and techniques - Students are prepared to implement governance in their future careers

Executive Training

Scenario: A Chief AI Officer needs to educate the executive team and board of directors about AI governance requirements and the organisation's current state.

How AIGLE Helps: - **High-Level Overview**: Container elements provide strategic overviews suitable for executive audiences - **Visual Communication**: The diagram helps executives understand complex governance relationships - **Maturity Assessment**: Assessment results provide objective data on current governance state - **Gap Identification**: Results clearly show where governance needs strengthening - **Professional Reports**: PDF reports are suitable for board presentations

Outcomes: - Executives understand AI governance requirements and organisational gaps - Board can make informed decisions about governance investments - Clear communication of governance strategy across leadership

Compliance Planning

Scenario: A compliance officer needs to prepare the organisation for upcoming AI regulations (e.g., EU AI Act) and demonstrate due diligence.

How AIGLE Helps: - **Framework Alignment**: NIST AI RMF aligns with many regulatory requirements - **Comprehensive Coverage**: AIGLE covers all aspects of AI governance required by regulations - **Gap Analysis**: Assessment identifies areas needing attention for compliance - **Documentation**: Tool recommendations include documentation tools for compliance evidence - **Audit Trail**: Assessment reports provide evidence of governance efforts

Outcomes: - Organization is prepared for regulatory requirements - Clear roadmap for achieving compliance - Documentation to demonstrate due diligence to regulators

Risk Assessment

Scenario: A risk manager needs to assess AI-related risks across the organisation's AI portfolio and prioritise risk mitigation efforts.

How AIGLE Helps: - **Risk Mapping**: MAP function helps identify and categorise AI risks - **Comprehensive Risk Coverage**: AIGLE covers technical, operational, and strategic risks - **Maturity Assessment**: Assessment reveals which risk management capabilities are weak - **Prioritization**: Results help prioritise which risks to address first - **Tool Recommendations**: Specific tools for risk assessment and mitigation

Outcomes: - Comprehensive understanding of AI risk landscape - Prioritized risk mitigation roadmap - Tools and techniques for ongoing risk management

Tool Selection

Scenario: A data science team lead needs to select tools for implementing fairness testing and model monitoring in their ML pipeline.

How AIGLE Helps: - **Curated Recommendations**: 50+ actively maintained, production-ready tools - **Contextual Suggestions**: Tools mapped to specific governance needs - **Comprehensive Information**: Links to documentation, repositories, and examples - **Multiple Options**: Several tools for each need, allowing selection based on tech stack - **Trust Dimension Mapping**: Understand which trustworthiness characteristics each tool supports

Outcomes: - Efficient tool selection without extensive research - Confidence in tool quality and maintenance - Clear understanding of tool capabilities and fit

Organizational Maturity Evaluation

Scenario: An AI governance leader needs to evaluate the organisation's current AI governance maturity and track improvement over time.

How AIGLE Helps: - **Structured Assessment**: 60+ questions provide comprehensive maturity evaluation - **Standardized Model**: Five-level maturity model enables consistent assessment - **Multi-Dimensional View**: Scores across layers and trust dimensions show strengths and gaps - **Baseline Establishment**: Initial assessment provides baseline for tracking progress - **Periodic Reassessment**: Regular assessments track maturity improvement over time

Outcomes: - Objective understanding of current governance maturity - Clear baseline for measuring improvement - Data-driven decisions about governance investments - Demonstrable progress to stakeholders

Project Planning

Scenario: A project manager is planning a new AI initiative and needs to ensure appropriate governance is built into the project plan.

How AIGLE Helps: - **Lifecycle Mapping**: Responsible AI Lifecycle shows all phases to plan for - **Activity Identification**: Each element details specific activities to include in project plan - **Artefact Planning**: Clear list of governance artefacts to produce at each phase - **Resource Planning**: Understanding of governance activities helps estimate resources - **Tool Selection**: Identify tools needed for governance activities

Outcomes: - Comprehensive project plan including governance activities - Realistic resource and timeline estimates - Governance integrated from project start, not bolted on later

Stakeholder Communication

Scenario: An AI product manager needs to communicate with diverse stakeholders (users, regulators, executives) about AI system trustworthiness.

How AIGLE Helps: - **Trust Lens**: Visual mapping of governance to trustworthiness characteristics - **Clear Explanations**: Accessible language for non-technical stakeholders - **Comprehensive Coverage**: Address all stakeholder concerns (fairness, safety, privacy, etc.) - **Professional Reports**: Shareable documentation of governance efforts - **Visual Communication**: Diagram helps explain complex concepts

Outcomes: - Effective communication with diverse stakeholders - Increased stakeholder trust and confidence - Clear demonstration of responsible AI practices

Vendor Evaluation

Scenario: A procurement officer needs to evaluate AI vendors and their governance practices as part of vendor selection.

How AIGLE Helps: - **Evaluation Framework**: AIGLE provides comprehensive framework for assessing vendor governance - **Question Bank**: Assessment questions can be adapted for vendor questionnaires - **Maturity Model**: Standardized model for comparing vendor maturity - **Comprehensive Coverage**: Ensure all governance dimensions are evaluated - **Tool Expectations**: Understand what tools and practices to expect from mature vendors

Outcomes: - Structured, comprehensive vendor evaluation - Consistent comparison across vendors - Confidence in vendor governance capabilities

Internal Audit

Scenario: An internal auditor needs to audit AI systems and governance practices to ensure compliance with organisational policies.

How AIGLE Helps: - **Audit Framework**: AIGLE provides comprehensive framework for AI governance audits - **Checklist Development**: Element activities and artefacts

inform audit checklists - **Maturity Assessment**: Assessment can be used as audit tool - **Gap Identification**: Clear identification of governance gaps and risks - **Recommendations**: Specific, actionable recommendations for addressing gaps

Outcomes: - Thorough, structured AI governance audits - Clear audit findings and recommendations - Roadmap for addressing audit findings

Best Practices

Integrating AIGLE into Governance Workflows

Start with Education

Initial Exploration: Before implementing governance changes, ensure stakeholders understand the governance landscape: - Have key stakeholders complete the guided tour - Explore the diagram together in team meetings - Discuss how AIGLE's framework relates to your organisation's context - Use AIGLE as a shared reference point for governance discussions

Build Common Language: Use AIGLE's terminology and structure to create shared understanding across technical and non-technical stakeholders.

Conduct Baseline Assessment

Establish Starting Point: Early in your governance journey, conduct a comprehensive maturity assessment: - Involve diverse stakeholders in assessment for comprehensive view - Be honest about current state, not aspirational goals - Document context and assumptions - Share results with leadership to build support for governance investments

Identify Quick Wins: Use assessment results to identify: - High-impact, low-effort improvements to build momentum - Critical gaps that need immediate attention - Areas where you're already strong that can be leveraged

Develop Phased Roadmap

Prioritize Improvements: Don't try to implement everything at once: - Focus on highest-priority gaps identified in assessment - Consider dependencies (some elements build on others) - Balance quick wins with longer-term strategic improvements - Align with organisational priorities and risk appetite

Sequence Thoughtfully: Typical maturity progression: 1. **Governance Foundation** (GOVERN): Establish policies, roles, and resources 2. **Risk Understanding** (MAP): Identify and document risks 3. **Measurement Capability** (MEASURE): Implement metrics and monitoring 4. **Risk Management** (MANAGE): Implement controls and response procedures 5. **Lifecycle Integration**: Embed governance throughout AI lifecycle 6. **Continuous Improvement**: Optimize based on experience

Integrate with Existing Processes

Don't Create Silos: Integrate AI governance with existing processes: - Connect to enterprise risk management - Align with existing project management methodologies - Integrate with existing compliance and audit processes - Leverage existing tools and systems where possible

Adapt to Context: Customize AIGLE's framework to your organisation: - Scale governance rigor to AI system risk level - Adapt terminology to your organisational culture

- Integrate with your existing governance structures - Tailor tool recommendations to your technology stack

Implement Tools Strategically

Start with High-Impact Tools: Based on AIGLE's recommendations: - Prioritize tools that address your highest-priority governance gaps - Start with tools that integrate well with your existing tech stack - Consider tools that address multiple governance needs - Evaluate tools thoroughly before production deployment

Build Capability Gradually: Don't try to implement all 50+ tools: - Start with 3-5 tools addressing critical needs - Build team capability with initial tools before expanding - Share learnings across teams as tools are adopted - Continuously evaluate tool effectiveness

Recommended Assessment Frequency

Initial Assessment

Timing: Conduct initial assessment when: - Beginning AI governance journey - Launching new AI governance programme - Preparing for regulatory compliance - Responding to AI-related incident or concern

Purpose: Establish baseline understanding of current maturity

Regular Reassessments

Quarterly Reviews: For organisations actively building governance capabilities: - Track progress on improvement initiatives - Identify emerging gaps as AI use expands - Adjust priorities based on changing context - Maintain momentum and accountability

Annual Assessments: For organisations with mature governance: - Comprehensive evaluation of governance effectiveness - Strategic planning for next year's improvements - Benchmarking against evolving standards - Reporting to leadership and board

Trigger-Based Assessments

Conduct Assessment When: - Launching significant new AI initiatives - Experiencing AI-related incidents or issues - Facing new regulatory requirements - Undergoing organisational changes affecting AI governance - Receiving stakeholder concerns about AI systems

Project-Specific Assessments

For Individual AI Projects: Conduct focussed assessments: - At project initiation to identify governance requirements - At key milestones to verify governance implementation - Before deployment to ensure readiness - Post-deployment to evaluate effectiveness

Collaboration Strategies

Cross-Functional Teams

Governance Requires Diverse Perspectives: - **Technical Teams**: Data scientists, ML engineers, software developers - **Risk and Compliance**: Risk managers, compliance officers, legal counsel - **Business Stakeholders**: Product managers, business leaders, domain experts - **Ethics and Social Impact**: Ethicists, social scientists, affected community representatives

Collaboration Approaches: - Use AIGLE as shared reference point in cross-functional meetings - Assign different teams to lead different governance elements based on expertise - Conduct joint assessments with diverse stakeholder participation - Share AIGLE tool recommendations across teams

Governance Committees

Establish AI Governance Committee: - Cross-functional membership representing diverse perspectives - Clear charter and decision-making authority - Regular meetings to review AI initiatives and governance - Use AIGLE framework to structure committee activities

Committee Activities: - Review maturity assessment results and improvement plans - Approve high-risk AI initiatives - Review incidents and lessons learned - Update governance policies and procedures - Oversee governance tool implementation

Knowledge Sharing

Build Organizational Capability: - Conduct training sessions using AIGLE as curriculum - Share lessons learned from governance implementation - Create internal documentation building on AIGLE's framework - Establish communities of practise for AI governance

External Engagement: - Participate in industry forums and working groups - Share your governance experiences (while respecting confidentiality) - Learn from others' governance journeys - Contribute to open-source governance tools

Documentation Approaches

Leverage AIGLE's Structure

Organize Documentation by AIGLE Elements: - Create documentation templates for each governance element - Use AIGLE's artefact lists as documentation checklists - Structure internal governance portal around AIGLE's framework - Reference AIGLE in internal policies and procedures

Maintain Living Documentation

Documentation Should Evolve: - Update documentation as governance practices mature - Document lessons learned and incorporate into procedures - Keep tool docu-

mentation current as tools are adopted - Version control governance documentation

Create Stakeholder-Appropriate Documentation

Different Audiences Need Different Documentation: - **Executive**: High-level summaries, dashboards, strategic implications - **Technical**: Detailed procedures, tool documentation, technical specifications - **Operational**: Runbooks, checklists, escalation procedures - **External**: Public-facing transparency reports, model cards, user documentation

Use AIGLE's Tool Recommendations

Documentation Tools from AIGLE: - Model Card Toolkit for model documentation - Datasheets for Datasets for data documentation - VerifyML for comprehensive model documentation - DVC for versioning documentation alongside code and data

Assessment Reports as Documentation

Leverage Assessment Reports: - Include assessment reports in governance documentation - Use reports to communicate governance state to leadership - Track assessment reports over time to show progress - Share reports with auditors and regulators as evidence of governance

Continuous Improvement

Regular Reviews

Establish Review Cadence: - Quarterly governance reviews to assess effectiveness - Annual strategic reviews to update governance strategy - Post-incident reviews to learn from issues - Post-project reviews to capture lessons learned

Metrics and Monitoring

Track Governance Effectiveness: - Monitor governance metrics (coverage, compliance, incidents) - Track maturity scores over time - Measure time and resources spent on governance - Assess stakeholder satisfaction with governance

Stay Current

Governance Must Evolve: - Monitor evolving AI regulations and standards - Stay current with AI governance best practices - Evaluate new tools as they emerge - Update governance practices based on new knowledge

Foster Learning Culture

Encourage Continuous Learning: - Celebrate governance successes and learnings - Treat governance gaps as learning opportunities, not failures - Encourage experimentation with new governance approaches - Share knowledge across teams and projects

Contact Information

Website: aigle.datadid.io

Email: hello@datadid.io

Arinze Okosieme. Website: okosieme.org

Consultation Booking: datadid.io

Appendices

Appendix A: Glossary of Terms

AI Governance: The system of policies, processes, and practices that guide the responsible development, deployment, and use of AI systems within an organisation.

AI Risk Management Framework (AI RMF): Framework developed by NIST providing a structured approach to managing risks associated with AI systems.

Algorithmic Bias: Systematic and repeatable errors in AI systems that create unfair outcomes, often reflecting historical biases in training data.

Artefact: A document, record, or output produced as part of governance activities (e.g., model card, risk assessment, test results).

Calibration: The degree to which an AI system's predicted probabilities match actual outcomes.

Data Drift: Changes in the statistical properties of input data over time that can degrade model performance.

Differential Privacy: A mathematical framework for providing privacy guarantees when analysing datasets.

Explainability: The ability to explain how an AI system produces its outputs in terms understandable to humans.

Fairness: The absence of bias or discrimination in AI system outcomes across different demographic groups.

Feature Engineering: The process of transforming raw data into features (input variables) suitable for machine learning.

Governance Element: A specific component of the AI governance framework representing a set of related activities (e.g., Problem Framing, Model Evaluation).

Human-in-the-Loop: AI system design that includes human involvement in decision-making or oversight.

Interpretability: The degree to which a human can understand the cause of an AI system's decision.

Maturity Model: A framework for assessing and improving organisational capabilities, typically with multiple levels from initial to optimised.

Model Card: Standardized documentation providing information about a machine learning model's intended use, performance, and limitations.

Model Drift: Degradation in model performance over time due to changes in the relationship between inputs and outputs.

NIST: National Institute of Standards and Technology, U.S. agency that develops technology standards and guidelines.

Responsible AI: The practise of designing, developing, and deploying AI systems in ways that are ethical, fair, transparent, and accountable.

Robustness: The ability of an AI system to maintain performance under varying conditions, including adversarial attacks.

Stakeholder: Any individual or group with an interest in or affected by an AI system (users, developers, affected parties, regulators, etc.).

Trust Characteristics: The eight dimensions of trustworthy AI identified by NIST: Valid & Reliable, Safe, Secure & Resilient, Accountable & Transparent, Explainable & Interpretable, Privacy-Enhanced, Fair with Harmful Bias Managed, and Environmentally Sustainable.

Trustworthy AI: AI systems that exhibit the characteristics necessary to earn and maintain stakeholder trust.

Validation: The process of verifying that an AI system meets requirements and performs acceptably in realistic deployment conditions.

Appendix B: NIST AI RMF References

Primary Reference: - NIST AI Risk Management Framework (AI RMF 1.0), January 2023 - Available at: https://www.nist.gov/itl/ai-risk-management-framework

Related NIST Publications: - NIST AI RMF Playbook (provides implementation guidance) - NIST Special Publication 1270: Towards a Standard for Identifying and Managing Bias in AI - NIST Trustworthy and Responsible AI Resource Center

Key Concepts from NIST AI RMF:

Four Core Functions: 1. GOVERN: Cultivate organisational culture and structure for AI risk management 2. MAP: Establish context and identify risks 3. MEASURE: Assess and benchmark AI risks 4. MANAGE: Allocate resources and implement risk treatment

Eight Trustworthiness Characteristics: 1. Valid and Reliable 2. Safe 3. Secure and Resilient 4. Accountable and Transparent 5. Explainable and Interpretable 6. Privacy-Enhanced 7. Fair with Harmful Bias Managed 8. Environmentally Sustainable (emerging consideration)

Risk Management Approach: - Socio-technical: Considers both technical and social dimensions - Lifecycle-oriented: Applies throughout AI system lifecycle - Continuous: Ongoing process, not one-time activity - Stakeholder-inclusive: Involves diverse perspectives

Appendix C: Additional Resources

AI Governance Frameworks and Standards: - ISO/IEC 42001: AI Management System - EU AI Act: European Union AI regulation - OECD AI Principles: International AI governance principles - IEEE 7000 Series: Standards for ethical AI

Responsible AI Resources: - Partnership on AI: Multi-stakeholder organisation advancing responsible AI - AI Ethics Guidelines Global Inventory: Comprehensive collection of AI ethics guidelines - Responsible AI Institute: Resources and certification for responsible AI

Technical Resources: - Papers with Code: ML papers with implementation code - Hugging Face: ML models, datasets, and tools - ML Commons: Benchmarks and best practices for ML

Open Source Tool Ecosystems: - Linux Foundation AI & Data: Open source AI projects - LFAI Landscape: Comprehensive map of open source AI tools - MLOps Community: Resources for ML operations

Educational Resources: - Fast.ai: Practical deep learning courses - Coursera AI Ethics: Courses on AI ethics and governance - Elements of AI: Free AI fundamentals course

Industry Organizations: - AI Now Institute: Research on social implications of AI - Data & Society: Research on data and society - AlgorithmWatch: Monitoring algorithmic decision-making

Appendix D: Privacy Policy Summary

Data Collection: - AIGLE collects minimal data necessary for functionality - Email addresses collected only for report delivery (not stored) - Assessment responses stored locally in user's browser - No personal information required to use AIGLE

Data Use: - Email addresses used only for report delivery - Assessment data used only for generating results and recommendations - No data used for marketing or purposes other than stated

Data Storage: - Assessment responses stored in browser local storage (user's device) - No user data stored on AIGLE servers - Users can clear local storage at any time

Data Sharing: - AIGLE does not share user data with third parties - Links to external tools provided but no data shared with those tools - No tracking or analytics beyond basic website usage statistics

User Rights: - Access: Users can access their assessment data in browser storage - Deletion: Users can delete their data by clearing browser storage - Export: Users can export their data via PDF reports - Portability: Assessment data is portable via PDF export

Security: - All traffic encrypted via HTTPS - No sensitive personal information collected - Regular security updates to dependencies

GDPR Compliance: - Data minimization: Only essential data collected - Purpose limitation: Data used only for stated purposes - Transparency: Clear privacy policy - User rights: Access, deletion, export supported

Contact: For privacy questions or concerns: hello@datadid.io

Full Privacy Policy: Available at: aigle.datadid.io/privacy

Appendix E: Version History

Version 1.0 - December 30, 2025: - Initial release of AIGLE platform - Three-layer interactive governance diagram - 28 governance elements with detailed information - 50+ open-source tool recommendations - Comprehensive maturity assessment system - Trust lens overlay with 8 NIST trust dimensions - 15-step guided tour - PDF report generation - Email report delivery - Dark mode support - Responsive design for desktop, tablet, and mobile - Comprehensive documentation

Planned Future Enhancements: - Additional tool recommendations as new tools emerge - Enhanced assessment analytics and benchmarking - Industry-specific guidance and customization - Integration with governance tool APIs - Collaborative assessment features for teams - Assessment comparison over time - Additional language support - Enhanced accessibility features

Appendix F: Acknowledgments

Framework Sources: - NIST AI Risk Management Framework team - OECD AI Policy Observatory - Partnership on AI - IEEE Standards Association

Open Source Community: - Developers and maintainers of the 50+ tools featured in AIGLE - Open source AI governance tool ecosystem - Contributors to AI fairness, explainability, and safety research

Technology Stack: - Next.js and React teams - Framer Motion developers - Tailwind CSS community - TypeScript team

Inspiration and Guidance: - Organizations pioneering responsible AI practices - Researchers advancing AI governance and trustworthy AI - Practitioners sharing lessons learned from AI governance implementation

---

End of Document

For more information, visit aigle.datadid.io or contact hello@datadid.io