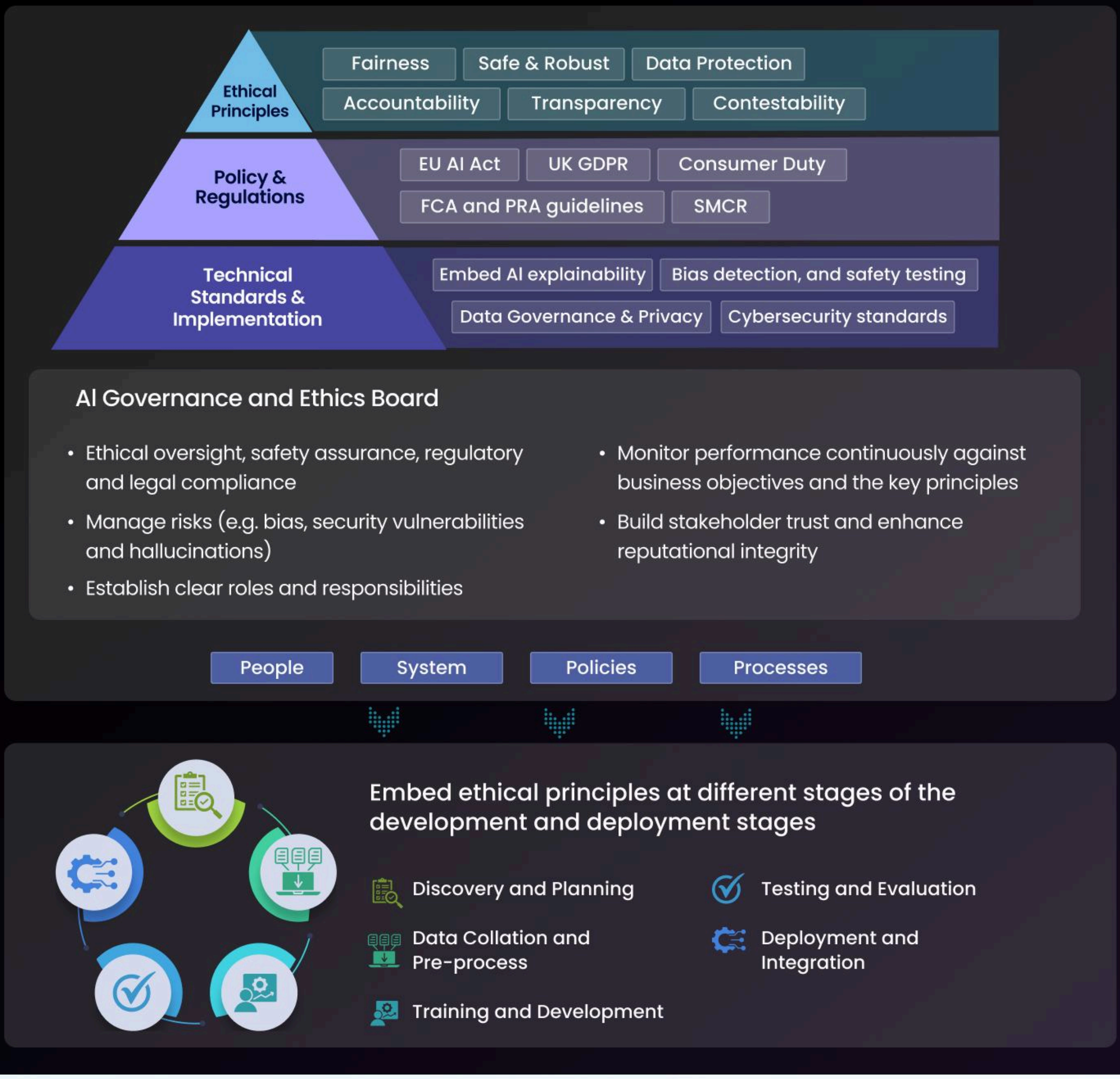
aveni Labs

FinLLM Safety Report Part I: **Governance**

Framework Development

AI Governance forms the foundation of Aveni's safety mitigation strategy. It covers all stages of the FinLM development lifecycle. Through comprehensive research into AI regulatory frameworks in the UK and other key global jurisdictions, we have developed a robust governance framework that primarily draws on the EU AI Act and existing FCA and PRA guidelines, interpreted through an AI-specific lens. Our approach is further strengthened by incorporating cutting-edge ethical AI research from our University of Edinburgh partners (as detailed in the 'Ethical Considerations in LLM' whitepaper which was co-authored and overseen by our Chief Data Scientist, Professor Alexandra Birch), alongside established industry standards such as Microsoft's Responsible AI standards and the NIST framework.

The Aveni governance framework



Ethical Principles: Fairness, Safe & Robust, Data Protection, Accountability, Transparency, Contestability

Policy & Regulations: EU AI Act, UK GDPR, Consumer Duty, FCA and PRA guidelines, SMCR

Technical Standards & Implementation: Embed AI explainability, Bias detection, and safety testing, Data Governance & Privacy, Cybersecurity standards

AI Governance and Ethics Board

- Ethical oversight, safety assurance, regulatory and legal compliance
- Manage risks (e.g. bias, security vulnerabilities and hallucinations)
- Establish clear roles and responsibilities
- Monitor performance continuously against business objectives and the key principles
- Build stakeholder trust and enhance reputational integrity

People, System, Policies, Processes

Embed ethical principles at different stages of the development and deployment stages

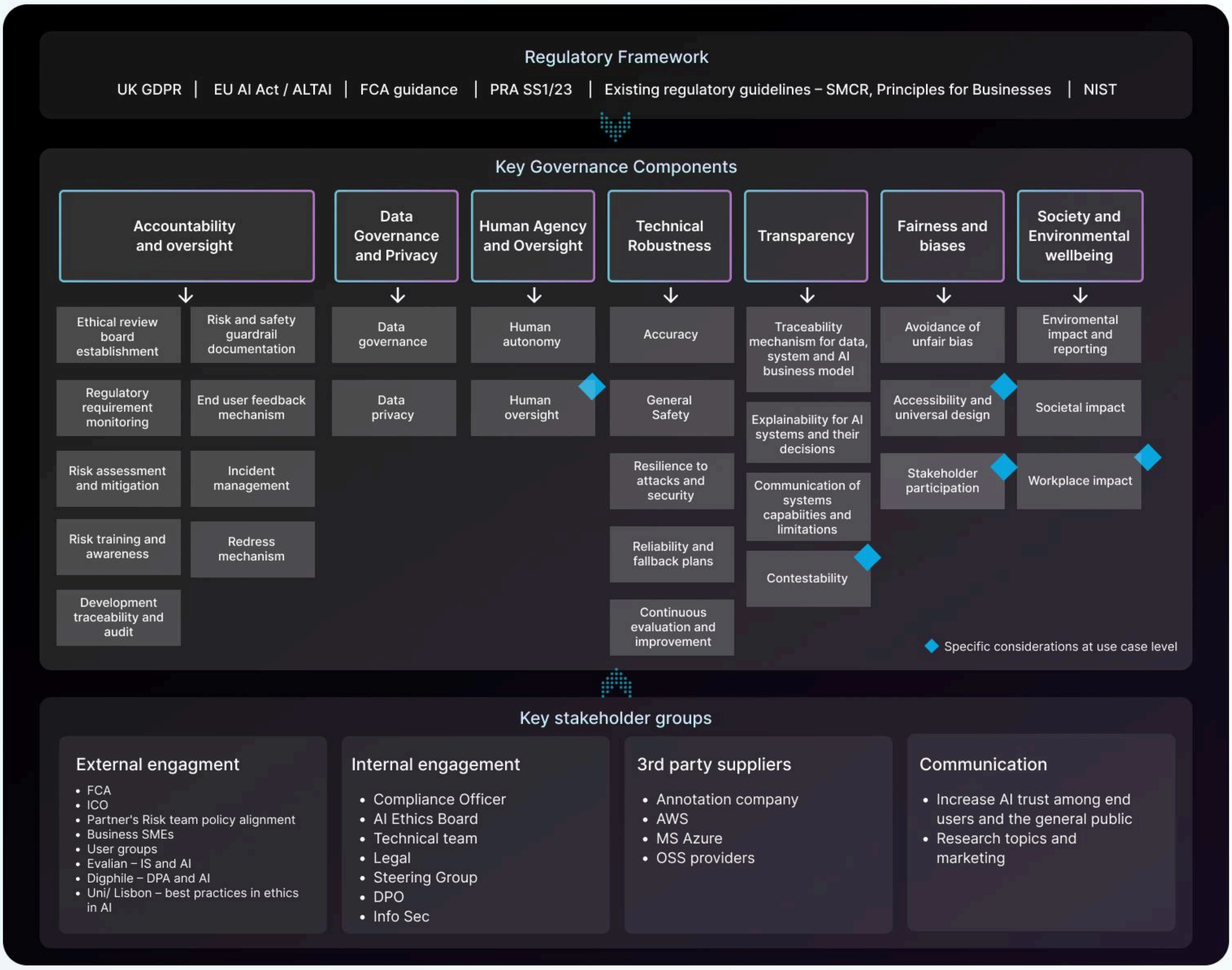
- Discovery and Planning
- Data Collation and Pre-process
- Training and Development
- Testing and Evaluation
- Deployment and Integration

Principles → Regulations → Standards → Requirements → Safety Report → Artefacts

1. Consider the AI Principles
2. Examine the regulatory frameworks
3. Assign the standards and requirements for documentation to show we are adhering to the standards
4. Summarise our approach in a safety report (this document)
5. Create individual artefacts and documentation (e.g. policies, impact assessments) to demonstrate compliance

This allows clear traceability of documentation and technical practices directly to the core AI principles to ensure auditability.

The Governance Scope on a Page (SOAP), outlined below, serves as the foundation for our compliance requirements and best practices. These are carefully mapped to key AI principles and then across each stage of the model development lifecycle, ensuring consistent and comprehensive integration of ethical and regulatory considerations throughout our AI development process.



Regulatory Framework

UK GDPR | EU AI Act / ALTAI | FCA guidance | PRA SS1/23 | Existing regulatory guidelines – SMCR, Principles for Businesses | NIST

Key Governance Components

Accountability and oversight
- Ethical review board establishment
- Risk and safety guardrail documentation
- Regulatory requirement monitoring
- End user feedback mechanism
- Risk assessment and mitigation
- Incident management
- Risk training and awareness
- Redress mechanism
- Development traceability and audit

Data Governance and Privacy
- Data governance
- Data privacy

Human Agency and Oversight
- Human autonomy
- Human oversight ◆

Technical Robustness
- Accuracy
- General Safety
- Resilience to attacks and security
- Reliability and fallback plans
- Continuous evaluation and improvement

Transparency
- Traceability mechanism for data, system and AI business model
- Explainability for AI systems and their decisions
- Communication of systems capabiities and limitations
- Contestability ◆

Fairness and biases
- Avoidance of unfair bias
- Accessibility and universal design ◆
- Stakeholder participation ◆

Society and Environmental wellbeing
- Enviromental impact and reporting
- Societal impact
- Workplace impact ◆

◆ Specific considerations at use case level

Key stakeholder groups

External engagment
- FCA
- ICO
- Partner's Risk team policy alignment
- Business SMEs
- User groups
- Evalian – IS and AI
- Digphile – DPA and AI
- Uni/ Lisbon – best practices in ethics in AI

Internal engagement
- Compliance Officer
- AI Ethics Board
- Technical team
- Legal
- Steering Group
- DPO
- Info Sec

3rd party suppliers
- Annotation company
- AWS
- MS Azure
- OSS providers

Communication
- Increase AI trust among end users and the general public
- Research topics and marketing

1. Governance

We've established a robust accountability structure through our Governance and Ethics Board, which oversees all aspects of legal, regulatory, and ethical compliance throughout our AI development pipeline—from data collection to deployment.

Supporting this Board are specialised working groups focused on: Safety, Information security and Data governance. These groups conduct detailed risk analyses to inform decision-making and ensure compliance with key standards including ISO standards and the UK GDPR requirements.

This tiered governance approach enables thorough oversight while maintaining operational efficiency in our AI development processes.

This process is monitored and actioned by a range of owners across the business, from the direct implementation of the FinLLM Technical team to the general oversight of the Aveni Governance and Ethics Board.

Owner	Key Responsibilities
Governance & Ethics Board	• Ethical oversight, safety assurance, regulatory and legal compliance • Manage risks e.g. bias, security vulnerabilities and hallucinations. • Establish clear roles and responsibilities • Monitor performance continuously against business objectives and the key principles. • Build stakeholder trust and enhance reputational integrity
Risk & Compliance Team	• Outline risk assessments and frameworks • Liaise with legal representatives and data protection expertise • Maintain view of regulatory landscape
Technical Team	• Execute against deliverables • Provide feedback on practicalities and performance of strategic steer
Senior Leadership Team	• Guidance against commercial and technical requirements

1.1 Key documentation

- **FinLLM Model Cards** hosted on HuggingFace
- **Data Protection Impact Assessment** to ensure personal data used in FinLLM training is processed with respect for individuals' data protection rights and freedoms.
- **Data Collation and Copyright Policy** to document the principles and practices that Aveni will follow when carrying out data collation activities from various sources and to comply with EU AI Act Article 53(1)(c).
- **Model Documentation Sheet** to provide transparency of our model and comply with EU AI Act Article 53(1)(a & b).
- **Model Risk Management Policy** to document the principles and standards for identifying, managing, and mitigating model risk associated with the development and deployment of FinLLM.
- **RAID Log** to document project risks, assumptions, issues, and dependencies.
- **Privacy Notice** to document collection, use, and protection of Personal Data, and clients' rights in respect to their Personal Data.
- **Model Use Guidance** to provide guidance on the proper use of FinLLM to ensure responsible, effective, and secure interactions.

1.2 Base Model Selection

Our commitment to safe model design begins with our selection of base models. FinLLM is a suite of LLMs ranging in size and transparency to account for differing risk-appetites of applications. To decide on the base models on which we will perform pre- and post-training, we created a scoring rubric against the most popular model families, assessing them **against 10 key transparency criteria**:

1. Architecture Disclosure
2. Training Data Transparency
3. Training Process Disclosure
4. Weights Availability
5. Code Availability
6. Licensing Openness
7. Evaluation Transparency
8. Responsible Use Guidance
9. Safety Benchmarking
10. Location

We take inspiration for our categories from the Foundation Model Transparency Index[1] but have adapted these for our requirements and considerations. In particular, whether the model has a commercially-permissive license, and the geographic region in which the model was developed (we place higher scores for models developed in Europe due to adherence to the EU AI Act and GDPR regulations for data privacy).

[1] https://crfm.stanford.edu/fmti/paper.pdf

We further measure the models against usability and logistic criteria such as how well the model is adopted (through HuggingFace downloads), the training framework supported, maximum context length, and availability of model sizes.

This approach allows us to systematically evaluate the transparency and usability of these models based on criteria important to Aveni and our partners, enabling us to make an informed decision about our choice of base models[2].

Model Family	Total[3]	Summary
LLaMA 3 (Meta)	13	partially open
Mistral / Mixtral	**15**	**partially open**
Qwen2.5 (Alibaba)	**13**	**partially open**
Olmo (AllenAI)	**18**	**open**
LLM360	19	open
Phi-4 (Microsoft)	15	partially open
Granite 3 (IBM)	**16**	**partially open**
Gemini (Google)	9	closed
GPT (OpenAI)	5	closed

[3] Sum of scores of the 10 categories. Out of 20. Model Transparency

1.3 Wider industry engagement

We engage with the wider regulatory community, specifically the Financial Conduct Authority (FCA), the University of Edinburgh (UoE), legal and regulatory compliance subject matter experts, and our project partners Lloyds Banking Group (LBG) and Nationwide Building Society (NBS).

[2] The final decision on base model selection is made through a combination of transparency criteria, usability, and performance against key evaluation benchmarks.

We also seek expert legal advice from specialists in data protection, artificial intelligence, and digital regulation. These partnerships allow us to remain abreast of the latest developments and decisions within the regulatory environment while remaining guided by the practical requirements of the industry.

Collaboration:

We've established strategic partnerships to strengthen our AI governance approach:

1. **FCA Collaboration:**

- Accepted into the FCA Digital Sandbox and participated in their 2 days in-person AI sprint. This served to validate our approach to governance with industry peers as well acknowledging the common challenges in the application. In particular, we gained a better appreciation of:

    a. The importance of AI literacy and our role in enhancing public trust in the use of AI by incorporating end users' concerns and needs into the design of the product.

    b. Potential future use cases for our use case analysis e.g. investment decision making, agentic AI, intelligent monitoring, generating training materials for compliance and financial literacy, scenario simulations and so forth.

    c. The AI Opportunity Plan which will form the roadmap for the UK government's strategy for enhancing growth and productivity.

2. **Regulatory Compliance specialists:**

- Engaged external specialists to navigate complex regulatory changes and data protection requirements
- Implemented company-wide AI regulation training to establish baseline risk awareness. This ensures that every employee understands how AI regulations specifically apply to their individual role and responsibilities.
- Submitted response to the AI Copyright Consultation Paper to support our case for an UK exception to data mining, akin to the EU.

3. Research Partnerships:

- Continue with ethical AI development through collaboration with the University of Edinburgh to translate research into real life application.
- Leverage learnings from the EuroLLM. Key members of our engineering team have worked on the project.
- Currently developing hallucination training datasets and guardrails with research student
- Working with European ethics committees to enhance our approach

In summary, our governance process follows a detailed structure to provide a solid foundation on which to develop FinLLM. We refer to these principles throughout the development process and continue to maintain strong relationships with the wider industry and regulatory landscape.

The next installment will focus on our safety approach to data collection, training, and evaluation.

Get in touch with our team to explore how FinLLM can be deployed in your organisation

Reach us at hello@aveni.ai to start
the conversation.

aveni Labs