

EU AI Act Compliance Checklist

10 Data Documentation Requirements

You Must Meet by August 2026

By David Cooper, CCIE #14019

Author of "The EU AI Act Survival Guide"

CRITICAL DEADLINE: AUGUST 2026

The EU AI Act becomes fully enforceable in August 2026. If you're building, deploying, or selling AI systems in the EU, you must demonstrate compliance with data documentation requirements.

Non-compliance penalties:

- Up to €35 million or 7% of global annual turnover
- Product bans in EU market
- Reputational damage

Use this checklist to:

- Audit your current compliance status
- Identify gaps in your documentation
- Build a roadmap to August 2026 compliance
- Demonstrate due diligence to regulators

The 10 Critical Requirements

1. Training Data Identification

Requirement: Document all training, validation, and testing datasets used.

What you need to document:

- Complete list of all datasets used
- Dataset names and versions
- Dataset sources
- Date acquired
- Purpose of each dataset

Why it matters: Regulators need to trace your AI system's behavior back to its training data.

2. Data Provenance Documentation

Requirement: Document origin, ownership, and chain of custody for all training data.

What you need to document:

- Where data came from
- Who owns the data
- How data was acquired
- Transformations applied
- Chain of custody

Why it matters: Proves you performed due diligence on data sources.

3. Temporal Documentation

Requirement: Document when each dataset was used in training, validation, and testing.

What you need to document:

- Date/time of each training run
- Datasets used in each run
- Model versions linked to datasets
- Timeline of updates
- Audit trail of access

Why it matters: Critical for identifying which models used problematic data.

4. Data Quality Assessment

Requirement: Assess and document quality, relevance, and representativeness.

What you need to document:

- Quality metrics
- Relevance to use case
- Representativeness analysis
- Data limitations
- Quality improvement steps

Why it matters: Poor quality data = unreliable AI systems.

5. Bias Detection and Mitigation

Requirement: Identify and mitigate bias, especially for protected characteristics.

What you need to document:

- Bias analysis performed
- Protected characteristics
- Identified biases
- Mitigation strategies
- Testing results

Why it matters: Discriminatory AI systems violate EU law.

6. Copyright and Licensing

Requirement: Respect IP rights and document compliance.

What you need to document:

- Copyright status
- Licenses for each dataset
- Terms of use compliance
- Rights to use for AI
- Attribution requirements

Why it matters: Training on copyrighted data without permission is illegal.

7. Personal Data Protection

Requirement: Comply with GDPR for personal data in training sets.

What you need to document:

- Legal basis for processing
- Data minimization
- Anonymization techniques
- Data subject rights
- DPIA if required

Why it matters: GDPR violations carry severe penalties.

8. Data Retention and Deletion

Requirement: Document retention periods and deletion procedures.

What you need to document:

- Retention periods
- Justification
- Deletion procedures
- Backup policies
- Deletion records

Why it matters: GDPR requires data minimization and limited retention.

9. Data Lineage and Versioning

Requirement: Track data lineage and maintain version control.

What you need to document:

- Dataset history
- Version numbers
- Relationships between datasets
- Model-data version links
- Reproducibility info

Why it matters: Version control enables reproducibility and debugging.

10. Audit Trail and Record-Keeping

Requirement: Maintain comprehensive records for 10+ years.

What you need to document:

- All above requirements
- Access logs
- Compliance reviews
- Incident reports
- Regular assessments

Why it matters: Regulators can audit at any time—missing records = non-compliance.

Your Compliance Roadmap

Phase 1: Audit Current State (Months 1-2)

- Inventory all datasets
- Assess documentation vs requirements
- Identify gaps and risks
- Prioritize compliance work

Phase 2: Implement Documentation (Months 3-8)

- Set up provenance tracking
- Document existing datasets
- Automate documentation for new data
- Train team

Phase 3: Build Processes (Months 9-14)

- Create acquisition procedures
- Implement quality workflows
- Build bias detection
- Establish retention policies

Phase 4: Validate and Test (Months 15-18)

- Internal compliance audit
- Third-party assessment
- Address gaps
- Document status

Phase 5: Maintain and Update (Ongoing)

- Quarterly reviews
- Update as data changes
- Monitor guidance
- Prepare for audits

How ProvenanceAI Helps

ProvenanceAI addresses 7 of the 10 requirements automatically:

- ✓ **Requirement 1:** Automatic dataset identification
- ✓ **Requirement 2:** Cryptographic proof of provenance
- ✓ **Requirement 3:** RFC 3339 timestamps
- ✓ **Requirement 6:** IP compliance tracking
- ✓ **Requirement 9:** SHA-256 version control
- ✓ **Requirement 10:** Legal-grade certificates
- ✓ **Zero-Knowledge:** GDPR-friendly architecture

Start your free trial today: ProvenanceAI.network
10 free certificates per month—no credit card required.

Next Steps

Immediate (This Week):

1. Print this checklist
2. Audit your current status
3. Identify biggest gaps
4. Assign compliance ownership

This Month:

1. Implement basic provenance tracking
2. Start ProvenanceAI free trial
3. Document critical datasets
4. Build compliance processes

By August 2026:

1. Complete documentation for all datasets
2. Establish automated workflows
3. Conduct internal audit
4. Be ready for regulatory inspection

Resources

Further Reading:

- "The EU AI Act Survival Guide" by David Cooper
Amazon: <https://www.amazon.com/dp/B0G7W2LF1D>
- EU AI Act Official Text
<https://eur-lex.europa.eu>
- ProvenanceAI Documentation
ProvenanceAI.network

Get Help:

- ProvenanceAI: ProvenanceAI.network
- Compliance Consulting: CyberShieldAustin.com
- Email: support@provenceai.network

About the Author

David Cooper, CCIE #14019, is a cybersecurity professional with 25 years of experience and author of 15+ technical books. He created ProvenanceAI to solve the data documentation challenge every AI company faces under the EU AI Act.

Connect:

LinkedIn: <https://www.linkedin.com/in/david-c-192601240/>
Twitter: <https://x.com/defiauditccie>
Website: CyberShieldAustin.com

Downloaded: December 19, 2025

© 2025 ProvenanceAI. Created by David Cooper, CCIE #14019.