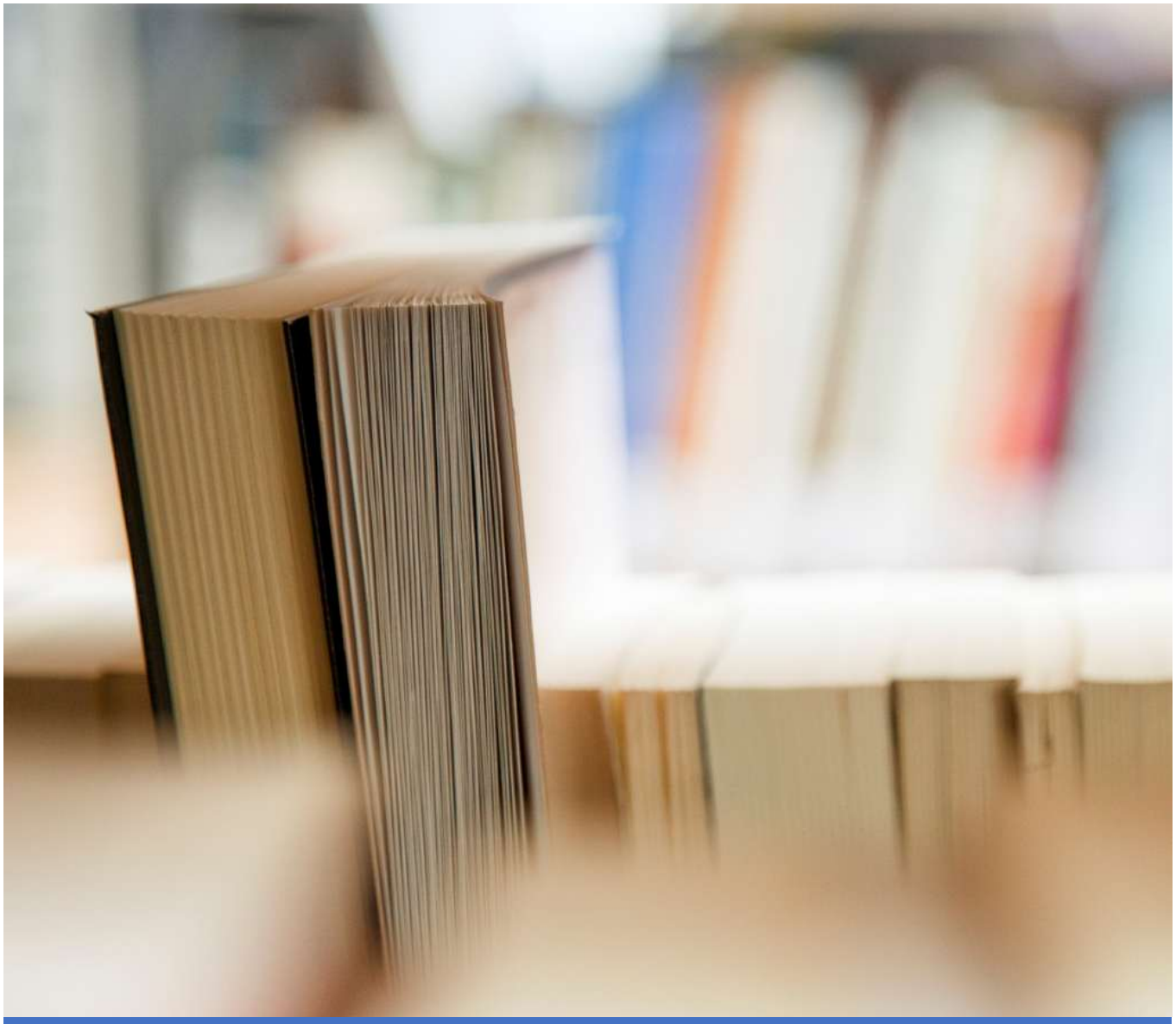

Due Diligence Checklist for an AI/ML Company

1. Corporate & Strategic Overview

- Company history, incorporation documents, cap table (fully diluted)
- Founder background (technical depth vs business execution)
- Vision: product vs platform vs infrastructure play
- Strategic positioning:
    - Vertical AI (healthcare, fintech, manufacturing)
    - Horizontal AI (LLMs, analytics, automation)
- Market category:
    - SaaS AI / AI Infra / DeepTech / Applied AI
- Key partnerships (cloud, data providers, enterprise clients)
- Exit readiness (IPO vs acquisition positioning)

2. Market & Competitive Landscape

- TAM / SAM / SOM (bottom-up preferred)
- Market maturity (emerging vs crowded)
- Key competitors:
    - Direct (same model/application)
    - Indirect (manual process, SaaS alternatives)
- Competitive moat:
    - Data advantage
    - Model performance
    - Switching costs
- Benchmarking:
    - Accuracy vs competitors
    - Cost per inference
    - Latency

3. AI/ML Technology Assessment (Core DD Area)

A. Model Architecture

- Type:
    - LLM / CV / NLP / RL / Hybrid
- Proprietary vs open-source base (e.g., fine-tuned models)
- Model size, training approach (pretraining vs fine-tuning vs RAG)
- Explainability (XAI capability)

B. Model Performance

- Metrics:
    - Accuracy / Precision / Recall / F1
    - BLEU / ROUGE (for NLP)
- Benchmark vs industry standards
- Drift detection & retraining frequency
- Real-world vs lab performance gap

C. MLOps & Infrastructure

- CI/CD pipelines for ML
- Model deployment:
    - Cloud (AWS, Azure, GCP)
    - On-prem / edge
- Monitoring:
    - Latency

    - Model degradation
- Versioning & rollback mechanisms

D. Scalability

- Inference cost per transaction
- GPU/compute efficiency
- Horizontal scaling capability
- Dependency on third-party APIs (e.g., OpenAI risk)

4. Data Due Diligence (MOST CRITICAL)

- Data sources:
    - Proprietary / licensed / scraped
- Data ownership & rights
- Data quality:
    - Label accuracy
    - Bias / imbalance
- Volume & uniqueness (core moat)
- Data pipeline:
    - Ingestion → cleaning → labeling → storage
- Compliance:
    - GDPR, HIPAA, DPDP Act (India)
- Synthetic data usage (if any)

5. Product & Use Case Validation

- Product maturity:
    - MVP / Beta / Production
- Core use cases & ROI for customers
- Integration complexity (API-first vs heavy deployment)
- UI/UX quality
- Customization vs standardized product
- Customer stickiness:
    - Daily usage?
    - Embedded in workflow?

6. Revenue Model & Financials

A. Revenue Streams

- SaaS subscription / usage-based / licensing
- Enterprise contracts vs SMB
- % of AI-driven vs services revenue

B. Financial Metrics

- ARR / MRR growth
- Gross margin (AI infra heavy vs SaaS-like)
- CAC & LTV
- Burn rate & runway
- Contribution margin per customer

C. Unit Economics (AI-specific)

- Cost per inference
- Training cost vs revenue scaling
- GPU/cloud spend as % of revenue

7. Customers & Traction

- Key clients (logos + concentration risk)
- Contract types:
    - Pilot vs long-term
- Retention:
    - Net Revenue Retention (NRR)
- Pipeline visibility
- Case studies (quantified ROI)

8. Legal, IP & Compliance

A. Intellectual Property

- Patents (filed/granted)
- Ownership of models & code
- Open-source usage:
    - License compliance (MIT, GPL risk)

B. Regulatory

- AI regulations (EU AI Act, India DPDP)
- Industry-specific compliance (healthcare, BFSI)

C. Contracts

- Data usage agreements
- Customer SLAs
- Vendor/cloud contracts

9. Security & Risk

- Data security architecture
- Cybersecurity audits
- Model risks:
    - Hallucinations (LLMs)
    - Bias / fairness issues
- Dependency risk:
    - Reliance on APIs (e.g., OpenAI, Anthropic)

10. Team & Talent

- Founders:
    - AI depth (PhD vs applied engineers)
- Team composition:
    - ML engineers vs software engineers
- Attrition risk
- Hiring pipeline (AI talent scarcity)

11. R&D & Innovation Pipeline

- Roadmap:
    - Model upgrades
    - New use cases
- Research partnerships (universities, labs)
- % revenue invested in R&D

12. ESG & Ethical AI

- Responsible AI policies
- Bias mitigation strategies
- Environmental impact (compute-heavy models)
- Governance frameworks

13. Exit & Scalability Assessment

- Strategic buyers:
    - Big Tech, SaaS companies
- Platform potential vs feature risk
- Global scalability
- Defensibility over 5–10 years