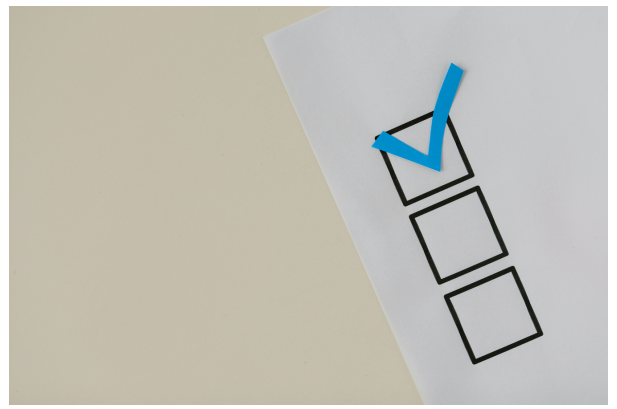


APPENDIX:

AI-Ready Data Checklist



This AI-Ready Data Checklist is designed to help construction and manufacturing SMBs assess the quality of their data before launching AI initiatives. Poor data quality is the leading cause of the 85% AI project failure rate (Chapter 1). By evaluating your data against the six dimensions of quality—accuracy, completeness, consistency, timeliness, relevance, and accessibility—you can identify gaps, prioritize improvements, and build a solid foundation for AI success. Use this checklist to audit your critical datasets, such as project records for construction or machine logs for manufacturing, and ensure they're ready to deliver ROI, like \$100,000–\$300,000 in annual savings (Chapter 6). For complex assessments, consider partnering with experienced consultants like Emphasis Tech, whose skilled professionals can streamline the process with their deep expertise in your industry (Chapter 8).



Select a critical dataset tied to your planned AI use case (e.g., predictive maintenance, quality control, resource optimization).



Answer each question below, checking "Yes" or "No." For "No" responses, note specific issues to address.



Calculate your readiness score by dividing the number of "Yes" answers by the total questions (20). A score of 80% or higher indicates AI readiness; below 80% suggests targeted improvements are needed.



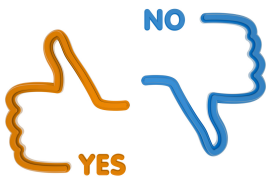
Refer to Chapters 4, 5, and 8 for strategies to address gaps, such as standardizing processes or cleaning data.

CHECKLIST: ACCURACY

DOES THE DATA REFLECT REALITY?

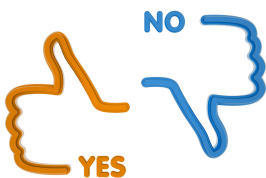
Are measurements and records free from transcription errors or manual entry mistakes?

EXAMPLE: CONSTRUCTION SITE REPORTS AVOID TYPOS IN MATERIAL QUANTITIES; MANUFACTURING SENSOR DATA MATCHES ACTUAL PRODUCTION OUTPUTS.



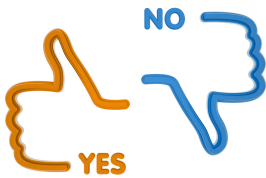
Is historical data validated against reliable sources (e.g., invoices, quality control logs)?

EXAMPLE: PAST PROJECT COSTS ALIGN WITH FINANCIAL RECORDS; MACHINE MAINTENANCE LOGS MATCH SERVICE REPORTS.



Are data collection methods (e.g., sensors, forms) calibrated or standardized to ensure precision?

EXAMPLE: SENSORS ON PRODUCTION LINES ARE REGULARLY CALIBRATED; FIELD DATA FORMS USE CONSISTENT UNITS.

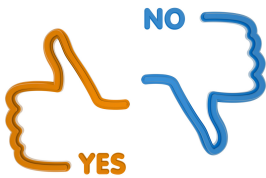


CHECKLIST: COMPLETENESS

ARE THERE GAPS IN THE DATA?

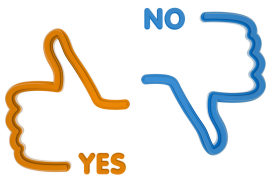
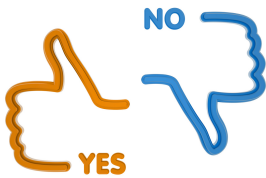
Do critical fields (e.g., project timelines, defect rates) have values for at least 95% of records?

EXAMPLE: NO MISSING DELIVERY DATES IN CONSTRUCTION LOGS; NO BLANK QUALITY CONTROL ENTRIES IN PRODUCTION DATA.



Are all relevant data points captured for your AI use case?

EXAMPLE: PREDICTIVE MAINTENANCE INCLUDES RUNTIME, TEMPERATURE, AND VIBRATION DATA; RESOURCE OPTIMIZATION INCLUDES LABOR AND EQUIPMENT USAGE.

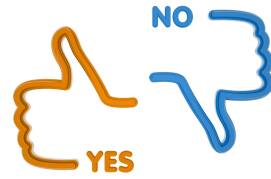


CHECKLIST: CONSISTENCY

IS DATA REPRESENTED UNIFORMLY
ACROSS SYSTEMS?

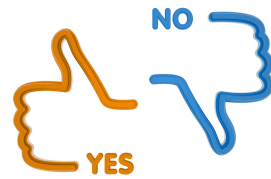
Are naming conventions, units, and formats consistent across departments and systems?

EXAMPLE: ALL CONSTRUCTION PROJECTS USE “MM” FOR MEASUREMENTS; MANUFACTURING DATA USES THE SAME PRODUCT CODES ACROSS ERP AND QUALITY SYSTEMS.



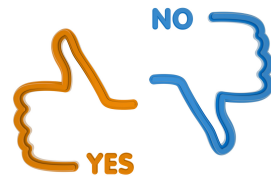
Do integrated systems (e.g., project management and accounting) reconcile without discrepancies?

EXAMPLE: MATERIAL COSTS MATCH BETWEEN FIELD REPORTS AND FINANCIAL SOFTWARE; PRODUCTION OUTPUTS ALIGN ACROSS MACHINE LOGS AND INVENTORY SYSTEMS.



Are duplicate records (e.g., customers, equipment) identified and resolved?

EXAMPLE: NO MULTIPLE ENTRIES FOR THE SAME SUBCONTRACTOR; NO REDUNDANT MACHINE IDS IN MAINTENANCE LOGS.

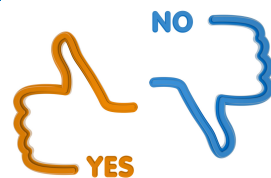


CHECKLIST: TIMELINESS

IS THE DATA CURRENT ENOUGH FOR AI?

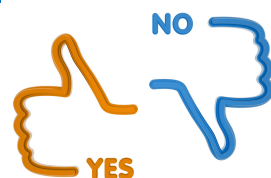
Is data updated frequently enough to support real-time or near-real-time AI applications?

EXAMPLE: CONSTRUCTION PROGRESS REPORTS ARE LOGGED DAILY; MANUFACTURING QUALITY DATA IS CAPTURED PER SHIFT.



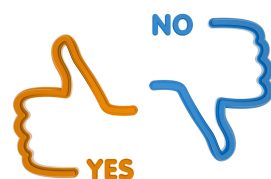
Are outdated records (e.g., old project data, obsolete product specs) archived or removed?

EXAMPLE: ONLY THE LAST 5 YEARS OF PROJECT DATA ARE USED FOR AI TRAINING; DISCONTINUED PRODUCTS ARE EXCLUDED FROM QUALITY ANALYSIS.



Do data collection processes minimize delays between event and entry?

EXAMPLE: MOBILE APPS ENABLE INSTANT FIELD DATA ENTRY; IOT SENSORS STREAM MACHINE DATA IN REAL TIME.

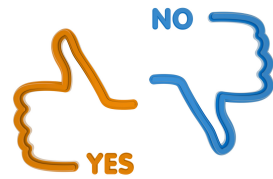


CHECKLIST: RELEVANCE

IS THE DATA ALIGNED WITH YOUR AI GOALS?

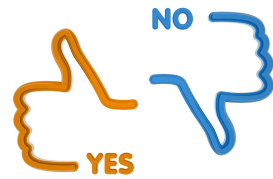
Does the dataset include only data relevant to the AI use case?

EXAMPLE: PREDICTIVE MAINTENANCE EXCLUDES UNRELATED METRICS LIKE EMPLOYEE SCHEDULES; QUALITY PREDICTION FOCUSES ON DEFECT-RELATED VARIABLES.



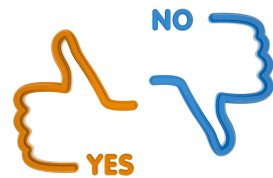
Are data sources prioritized based on their impact on business outcomes (e.g., cost savings, efficiency)?

EXAMPLE: RESOURCE OPTIMIZATION PRIORITIZES LABOR AND MATERIAL DATA OVER MARKETING METRICS; QUALITY CONTROL FOCUSES ON HIGH-DEFECT PROCESSES.



Is the dataset large enough to support the AI model's requirements?

EXAMPLE: MACHINE LEARNING MODELS HAVE AT LEAST 6–12 MONTHS OF HISTORICAL DATA; RULE-BASED AI SYSTEMS HAVE COMPREHENSIVE RULE SETS.

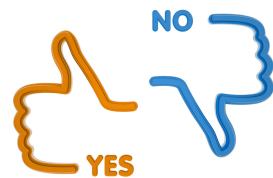


CHECKLIST: ACCESSIBILITY

CAN AI SYSTEMS ACCESS THE DATA?

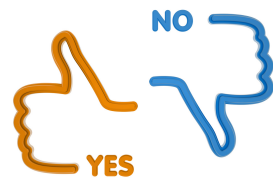
Is data stored in a format compatible with AI tools (e.g., CSV, JSON, or databases)?

EXAMPLE: PROJECT DATA IS EXPORTED FROM SPREADSHEETS TO A DATABASE; MACHINE LOGS ARE IN A STANDARDIZED DIGITAL FORMAT.



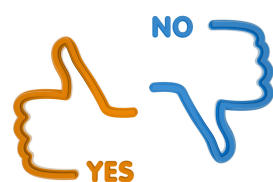
Are data silos eliminated or bridged through integration (e.g., APIs, data layers)?

EXAMPLE: CONSTRUCTION ERP AND FIELD APPS SHARE DATA VIA APIS; MANUFACTURING SYSTEMS FEED INTO A CENTRAL DATA WAREHOUSE.



Do team members and AI systems have appropriate access permissions without compromising security?

EXAMPLE: DATA STEWARDS CAN EDIT RECORDS, BUT SENSITIVE FINANCIAL DATA IS RESTRICTED; AI TOOLS ACCESS ONLY NECESSARY DATASETS.

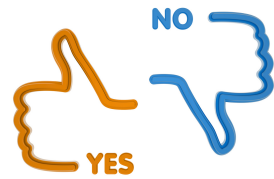


CHECKLIST: GOVERNANCE & MONITORING

IS THE DATA MANAGED EFFECTIVELY?

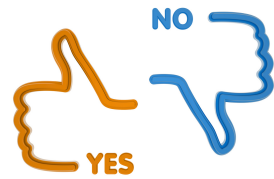
Are data quality standards (e.g., validation rules, governance policies) documented and enforced?

EXAMPLE: CONSTRUCTION TEMPLATES MANDATE REQUIRED FIELDS; MANUFACTURING SYSTEMS FLAG INCONSISTENT ENTRIES AUTOMATICALLY.



Are regular audits scheduled to monitor data quality and catch emerging issues?

EXAMPLE: QUARTERLY REVIEWS CHECK FOR COMPLETENESS IN PROJECT LOGS; MONTHLY AUDITS VERIFY MACHINE DATA ACCURACY.



SCORING & NEXT STEPS

CALCULATE READINESS

Divide the number of “Yes” answers by 20. A score of 80% (16/20) or higher indicates your data is AI-ready. Below 80% means you need targeted improvements.

ADDRESS GAPS

For each “No” answer, refer to Chapters 4 and 5 for solutions, such as standardizing processes, cleaning data, or integrating systems. Chapter 8’s roadmap can guide your action plan.

SEEK EXPERT SUPPORT

If gaps are complex or time is limited, engage experienced consultants. Emphasis Tech’s skilled professionals, with extensive expertise in data quality, can conduct audits, resolve issues, and ensure your data is AI-ready faster and more effectively.

By using this checklist, you’ll avoid the pitfalls of poor data quality, ensure your AI investments deliver results like 15–20% better project estimates or 20–30% less downtime (Chapter 6), and build a lasting data foundation.