

Data Quality Fundamentals

- Data quality underpins AI readiness and analytics accuracy in 2026
- Poor data degrades models by 15-20% and drives flawed business decisions
- Learn to systematically check completeness, consistency, timeliness, and validity
- Roadmap: from core dimensions to a practical daily checklist



What 'Fit for Purpose' Really Means

Technically Valid



VS.

Fit for Purpose



Data quality = fitness for purpose, not abstract perfection



ISO 8000 levels: syntactic, semantic, and pragmatic quality



Technically correct data is not always useful data



Example: valid phone format, but belongs to wrong person



Example: perfect historical data, but too stale for today's decisions

The Core Dimensions of Data Quality

✓ **Completeness:** Are all required fields populated?

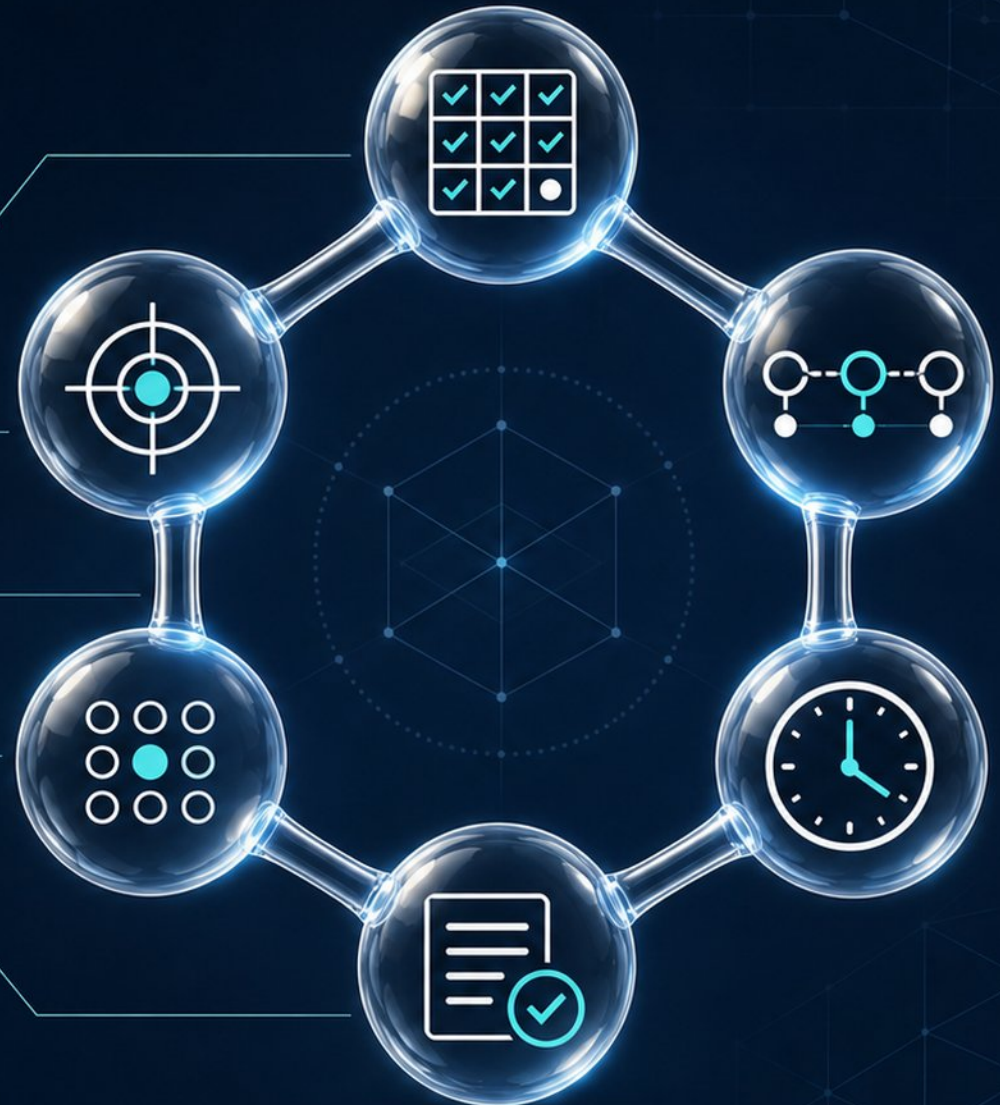
✓ **Consistency:** Do the same facts match across systems?

✓ **Timeliness:** Is the data fresh enough to use?

✓ **Validity:** Does data match the required format and rules?

✓ **Uniqueness:** Are there unwanted duplicates?

✓ **Accuracy:** Does the data reflect real-world truth?



Completeness: Is Everything Here?



- Measure presence of required data: missing records and blank critical fields



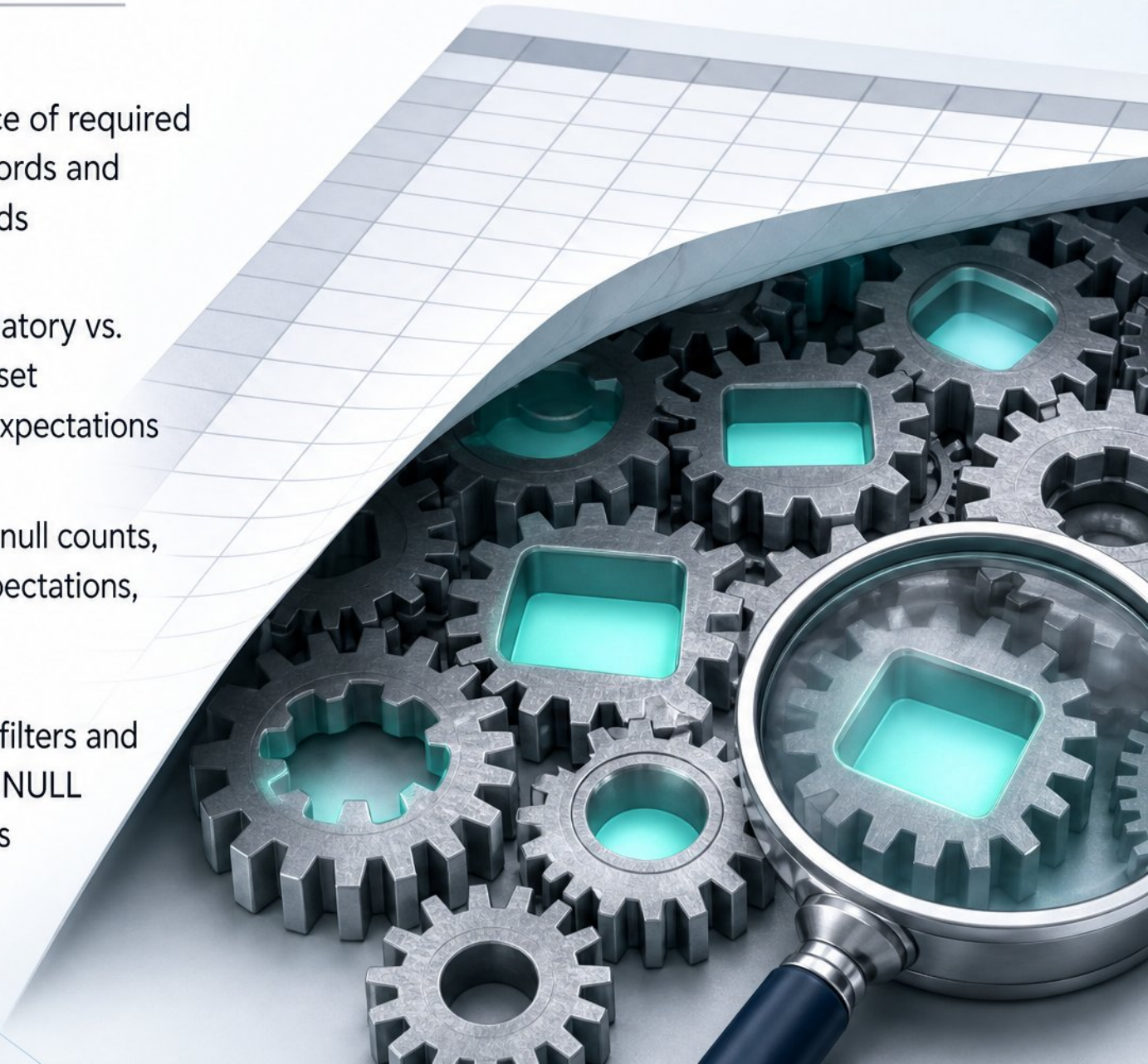
- Distinguish mandatory vs. optional fields to set column-specific expectations



- Practical checks: null counts, row counts vs expectations, and fill rates



- Use spreadsheet filters and SQL COUNT / IS NULL patterns to assess completeness



Consistency: Does It All Match?

- Ensure uniform formats, units, and codes across all data sources
- Cross-field rules: e.g., order totals must equal line-item sums
- Cross-table checks maintain referential integrity between systems
- Use lookup tables and schema enforcement to catch value drift
- Detect conflicts like mismatched customer names in CRM vs. billing



Timeliness: How Fresh Must the Data Be?

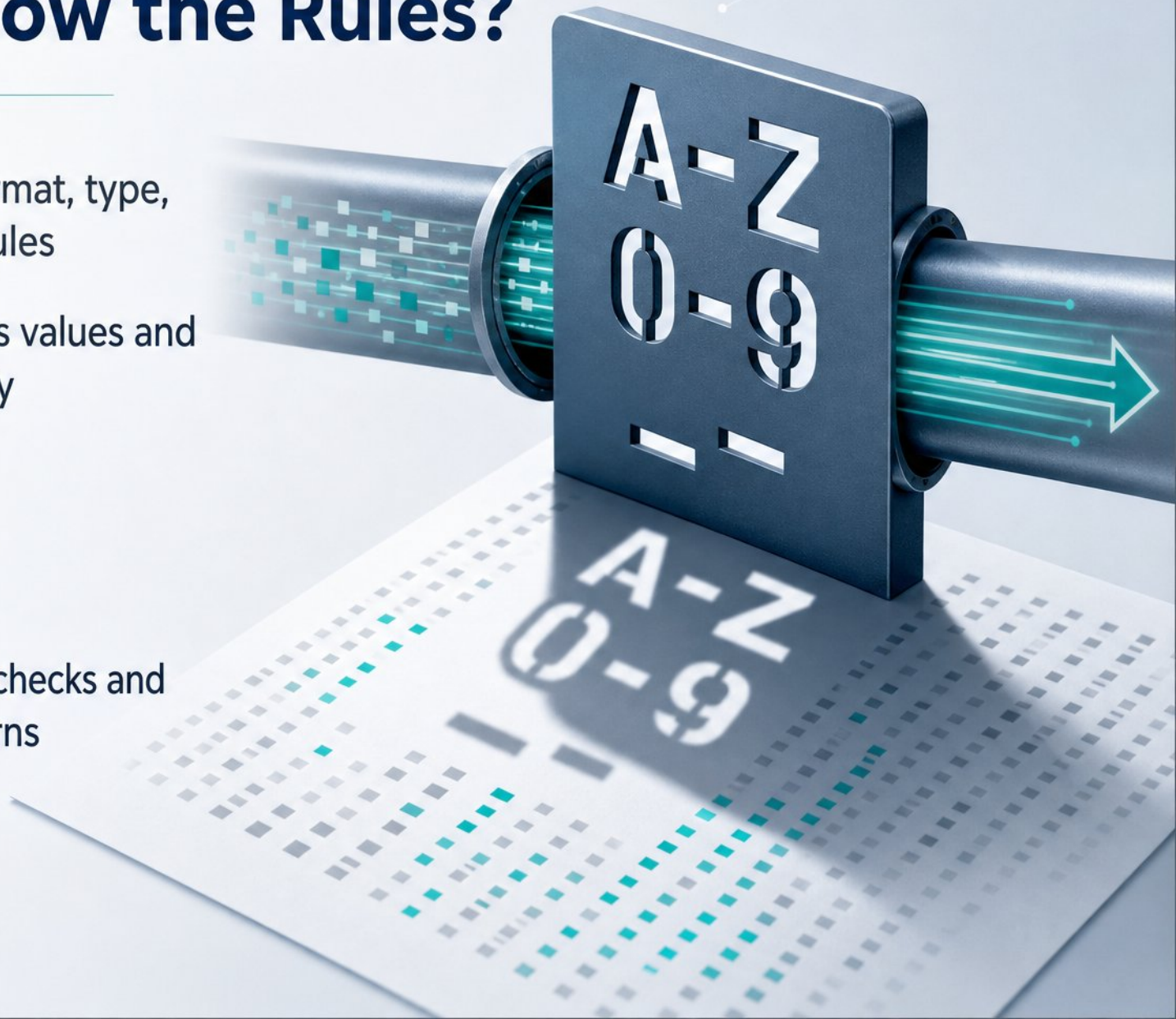
- Age, latency, and update frequency define whether data is current enough
- Set freshness thresholds per analysis or report type, not one-size-fits-all
- Check last-modified timestamps, load times, and staleness flags
- Stale data breaks decisions: yesterday's inventory for today's orders



Validity:

Does It Follow the Rules?

- Data conforms to format, type, range, and domain rules
- Detect out-of-bounds values and format violations early
- Validation: regex, allowable-value lists, range checks
- Spreadsheet-friendly checks and SQL constraint patterns



Uniqueness: No Unwanted Copies

- True duplicates vs legitimate repeated transaction records
- Exact-match and fuzzy-match detection approaches
- Define keys and unique identifiers before merging
- Duplicate records silently break joins, aggregations, and trust



Accuracy: The Hardest Dimension

- Accuracy is correctness against the real world, not just rule conformity
- A valid format does not guarantee the value describes the right entity
- Measurement requires comparison to a trusted reference source or primary research
- Use sampling, spot checks, and third-party verification for practical assessment
- Communicate accuracy limitations clearly when trusted sources are unavailable



Assembling Your Pre-Analysis Data Checklist



- 6-step workflow: Completeness → Consistency → Timeliness → Validity → Uniqueness → Accuracy



- Check missing records and stale data before digging into format rules



- Document findings in plain language for business stakeholders



- Use a template: what to check, how to check, how to report

Profiling and Baselining Before You Build



- You can't improve what you haven't measured in data quality



- Use modern platform tools for automated profiling (Databricks, Snowflake)



- Create a baseline with null ratios, freshness metrics, and duplicate counts



- Prioritize cleanup efforts based on baseline results for maximum impact



Common Pitfalls and How to Avoid Them

- Over-relying on summary stats without inspecting raw data
- Assuming imported data is clean without verification
- Confusing display formatting problems with actual data errors
- The "no nulls" trap: blanket rules create alert fatigue
- Real cautionary tales from analysis and AI projects



Automation and AI for Data Quality Checks



- Rule-based validation at ingestion prevents errors from entering systems



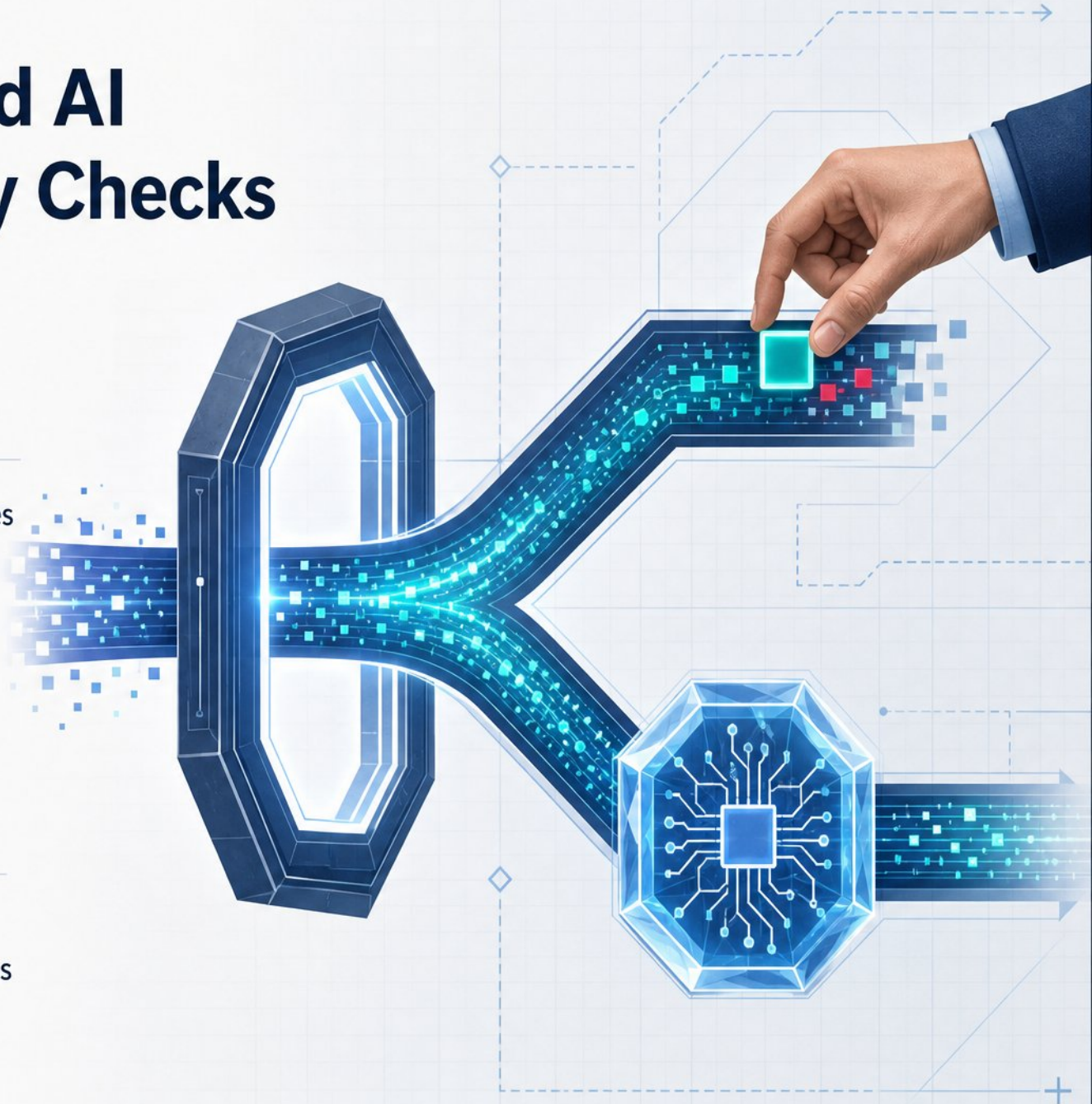
- AI-anomaly detection scales checks but still needs human oversight



- Data contracts enforce schemas and standards between producers and consumers



- Real-time monitoring dashboards route violations to stewards and engineers



Data Quality Governance and Culture



- Data quality requires org-wide culture change, not just tech fixes



- Data Owners: senior leaders accountable for domain value



- Data Stewards: daily oversight, quality enforcement, and alerts



- Shared metrics and cross-functional accountability prevent silos



- Embed quality champions in teams for peer-to-peer learning



Putting Data Quality into Daily Practice



- Embed checks in your first 5 minutes with data



- Match quality checks to your specific analysis goal



- Start today with spreadsheets and basic SQL queries



- Track ROI: time saved, errors prevented, trust built



- Explore provided templates and continued learning paths



PersonWise.

PersonWise • AI digital-human course platform



Scan the code or click anywhere to watch this course live, with voice Q&A

personwise.ai/t/d63ad29d/public