

GITG-APF™ (Grumpy Old IT Guy AI Profiling Framework™)

Version 1.0

Copyright © 2026 Grumpy Old IT Guy Publishers / Angelo R. Bobak

Executive Summary

The **Grumpy Old IT Guy AI Profiling Framework™ (GITG-APF™)** is an original methodology for applying Artificial Intelligence to enterprise data profiling. Traditional profiling focuses on measuring the statistical condition of data by counting null values, duplicates, invalid formats, and broken relationships. GITG-APF™ extends this approach by using AI to discover hidden patterns, predict future quality issues, identify emerging risks, recommend remediation actions, and continuously improve organizational data assets through machine-assisted analysis.

Unlike conventional profiling tools that describe what currently exists, GITG-APF™ is designed to explain why quality problems occur, estimate where they are likely to appear next, and recommend the most effective corrective actions. The framework combines traditional profiling techniques with machine learning, metadata intelligence,

business rule evaluation, master data analysis, and governance policies to create a continuously improving data quality ecosystem.

Introduction

As organizations accumulate increasingly large volumes of structured and semi-structured data, manual profiling techniques become less effective. Modern enterprises require intelligent systems capable of analyzing millions—or even billions—of records while automatically recognizing patterns that would be impossible to detect manually.

The **Grumpy Old IT Guy AI Profiling Framework™ (GITG-APF™)** was introduced in 2026 as part of the Grumpy Old IT Guy Data Quality Framework to address these challenges. It provides a repeatable methodology for integrating AI-assisted analysis into traditional data profiling activities while maintaining strong governance and human oversight.

The framework is technology-neutral and can be implemented using SQL Server, Microsoft Fabric, Microsoft Purview, Azure AI services, Python, machine learning platforms, or other modern data management technologies.

Purpose of the Framework

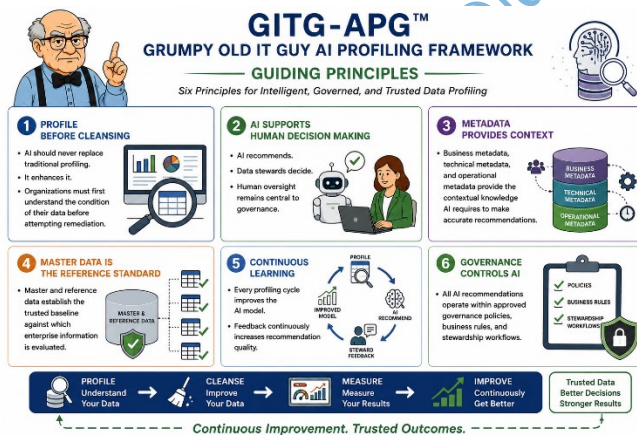
The primary objectives of GITG-APF™ are to:

- Accelerate enterprise data profiling
- Detect hidden data quality issues

- Predict future quality problems
- Recommend remediation activities
- Prioritize data quality risks
- Improve governance decisions
- Reduce manual profiling effort
- Support continuous improvement

Guiding Principles

The framework is built upon six guiding principles.



AI Profiling Framework Guiding Principles

1. Profile Before Cleansing

AI should never replace traditional profiling.

It enhances it.

Organizations must first understand the condition of their data before attempting remediation.

2. AI Supports Human Decision Making

AI recommends.

Data stewards decide.

Human oversight remains central to governance.

3. Metadata Provides Context

Business metadata, technical metadata, and operational metadata provide the contextual knowledge AI requires to make accurate recommendations.

4. Master Data Is the Reference Standard

Master and reference data establish the trusted baseline against which enterprise information is evaluated.

5. Continuous Learning

Every profiling cycle improves the AI model.

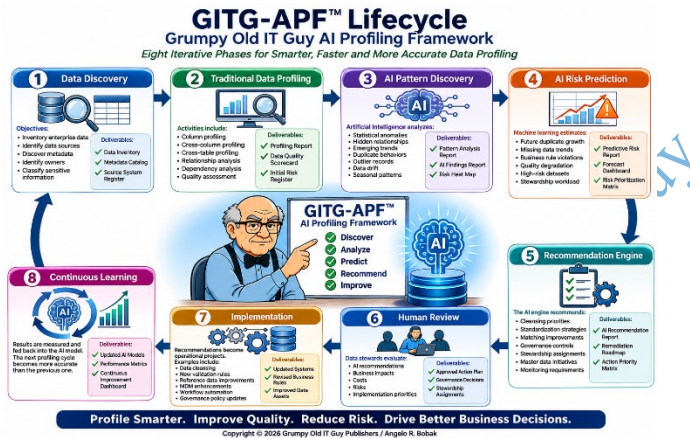
Feedback continuously increases recommendation quality.

6. Governance Controls AI

All AI recommendations operate within approved governance policies, business rules, and stewardship workflows.

The GITG-APF™ Lifecycle

The framework consists of eight iterative phases.



The GITG-APF™ Lifecycle

Phase 1 — Data Discovery

Objectives:

- Inventory enterprise data
- Identify data sources
- Discover metadata
- Identify owners
- Classify sensitive information

Deliverables:

- Data Inventory
- Metadata Catalog
- Source System Register

Phase 2 — Traditional Data Profiling

Activities include:

- Column profiling
- Cross-column profiling
- Cross-table profiling
- Relationship analysis
- Dependency analysis
- Quality assessment

Deliverables:

- Profiling Report
- Data Quality Scorecard
- Initial Risk Register

Phase 3 — AI Pattern Discovery

Artificial Intelligence analyzes:

- Statistical anomalies
- Hidden relationships
- Emerging trends
- Duplicate behaviors
- Outlier records
- Data drift
- Seasonal patterns

Deliverables:

- Pattern Analysis Report
- AI Findings Report
- Risk Heat Map

Phase 4 — AI Risk Prediction

Machine learning estimates:

- Future duplicate growth
- Missing data trends
- Business rule violations
- Quality degradation
- High-risk datasets
- Stewardship workload

Deliverables:

- Predictive Risk Report
- Forecast Dashboard
- Risk Prioritization Matrix

Phase 5 — Recommendation Engine

The AI engine recommends:

- Cleansing priorities
- Standardization strategies
- Matching improvements
- Governance controls
- Stewardship assignments
- Master data initiatives
- Monitoring requirements

Deliverables:

- AI Recommendation Report
- Remediation Roadmap
- Action Priority Matrix

Phase 6 — Human Review

Data stewards evaluate:

- AI recommendations
- Business impacts
- Costs
- Risks
- Implementation priorities

Deliverables:

- Approved Action Plan
- Governance Decisions
- Stewardship Assignments

Phase 7 — Implementation

Recommendations become operational projects.

Examples include:

- Data cleansing
- New validation rules
- Reference data improvements
- MDM enhancements
- Workflow automation
- Governance policy updates

Deliverables:

- Updated Systems
- Revised Business Rules
- Improved Data Assets

Phase 8 — Continuous Learning

Results are measured and fed back into the AI model.

The next profiling cycle becomes more accurate than the previous one.

Deliverables:

- Updated AI Models
- Performance Metrics
- Continuous Improvement Dashboard

Framework Deliverables

A complete GITG-APF™ implementation typically produces:

- Enterprise Data Inventory
- Metadata Catalog
- Data Profiling Report
- Data Quality Scorecard
- AI Pattern Discovery Report
- Predictive Risk Assessment
- Risk Heat Map
- AI Recommendation Report
- Remediation Roadmap
- Stewardship Action Plan
- Governance Decision Log
- Continuous Monitoring Dashboard
- Executive Summary Report

I leave this up to you to implement.

Recommended Templates

The framework is supported by a collection of reusable templates, including:

GITG-APF-101

Enterprise Data Inventory Template

GITG-APF-102

Metadata Assessment Worksheet

GITG-APF-103

Data Profiling Checklist

GITG-APF-104

AI Pattern Discovery Report

GITG-APF-105

Data Quality Risk Register

GITG-APF-106

Predictive Risk Assessment Matrix

GITG-APF-107

AI Recommendation Report

GITG-APF-108

Remediation Roadmap

GITG-APF-109

Data Steward Review Form

GITG-APF-110

Governance Approval Checklist

GITG-APF-111

Continuous Monitoring Scorecard

GITG-APF-112

Lessons Learned Register

These also I leave up to you to create!

Benefits

Organizations implementing GITG-APF™ can expect:

- Earlier identification of data quality issues
- Reduced manual profiling effort
- Faster root-cause analysis
- Improved stewardship productivity
- Better prioritization of remediation projects
- More accurate master data
- Stronger governance decisions
- Enhanced regulatory compliance
- Lower operational risk
- Continuous improvement through AI-assisted learning

Relationship with Other GITG Frameworks

GITG-APF™ is intended to complement—not replace—the broader Grumpy Old IT Guy framework family.

- **GITG-DQF™** defines the overall Data Quality Framework.
- **GITG-DPFF™** provides the traditional Data Profiling Framework and foundational profiling activities.
- **GITG-DQLF™** defines the continuous Data Quality Lifecycle.
- **GITG-DQMM™** measures organizational Data Quality Maturity.
- **GITG-DQMF™** defines the core dimensions of enterprise data quality.
- **GITG-APF™** extends these frameworks by introducing AI-assisted profiling, predictive analytics, intelligent recommendations, and continuous learning.

Together, these frameworks form an integrated methodology for assessing, governing, improving, and sustaining enterprise data quality in modern, AI-enabled organizations.

Conclusion

The **Grumpy Old IT Guy AI Profiling Framework™ (GITG-APF™)** represents the evolution of traditional data profiling into an intelligent, predictive discipline. By combining statistical profiling, metadata analysis,

governance principles, machine learning, and human stewardship, the framework enables organizations to move beyond simply identifying data quality problems toward anticipating future issues and implementing targeted, measurable improvements. Rather than replacing experienced data professionals, GITG-APF™ amplifies their capabilities by delivering actionable insights and prioritizing remediation efforts to foster a culture of continuous improvement in data quality across the enterprise.

© 2026 Angelo R. Bobak
All rights reserved.

First Edition: February 2026

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopying, recording, or otherwise—without the publisher's prior written permission.

Grumpy Old IT Guy Publishers
Westchester County, NY, USA
www.grumpyolditguy.com