Journal of Digital Assets and Monetary Research

Volume 1, Issue 1

Research Article

Date of Submission: 13 February 2026
Date of Acceptance: 16 March 2026
Date of Publication: 23 March 2026

AEGIS: An Autonomous Equity Grading and Intelligence System

A Unified Framework for Continuous, Probabilistic, and AI-Augmented Global Equity Valuation

Divyanshu Verma*
Verma Research Capital LLC Wisconsin, United States

***Corresponding Author:** Divyanshu Verma, Verma Research Capital LLC Wisconsin, United States.

Citation: Divyanshu, V. (2026). AEGIS: An Autonomous Equity Grading and Intelligence System. *J Digi Assets Monetary Res*. 1(1), 01-28.

Abstract

This paper introduces AEGIS, the Autonomous Equity Grading and Intelli- gence System, a proprietary valuation infrastructure developed at Verma Research Capital (VRC). AEGIS integrates large language model processing, stochastic discounted cash flow mechanics, Bayesian belief updating, multi- factor quantitative scoring, natural language analysis of earnings transcripts, macroeconomic regime detection, and continuous monitoring pipelines into a unified, automated platform for global equity valuation and opportunity identification. The system is motivated by the philosophical conviction, articulated across the VRC research canon, that intrinsic value is not a static calculation but a dynamic probability distribution that must be continuously refined as new information arrives. AEGIS operationalizes this conviction at scale, processing hundreds of annual reports, quarterly earnings calls, regulatory filings, and macroeconomic data releases to produce a composite score for each covered company, a probability-weighted intrinsic value range, and a real-time flag whenever the market price diverges materially from that range. The paper provides the complete theoretical foundation, mathematical formalism, system architecture, LLM integration design, quant scoring methodology, management quality assessment model, competitive moat quantification framework, and valuation synthesis engine that together constitute AEGIS. Case illustrations from the Indian banking sector, consistent with prior VRC research, are used throughout to ground the abstract machinery in concrete analytical substance.

Introduction

The central problem of active equity investing is deceptively simple to state and extraordinarily difficult to solve: determine what a business is worth, compare that estimate to what the market is willing to pay, and act on any meaningful gap. In practice, this problem is compounded by the sheer volume of information that must be processed to form a credible view, the speed with which that information changes, the cognitive biases that distort human judgment in the presence of uncertainty, and the structural incentive conflicts that corrupt the institutional research machinery through which most investors access fundamental analysis.

Verma Research Capital (VRC) was founded on the conviction that these obstacles, while formidable, are not insuperable. The firm believes that the disciplined integration of quantitative rigor, large language model intelligence, probabilistic reasoning, and deep fundamental analysis into a unified computational platform can produce a durable analytical edge. AEGIS is the expression of that conviction in software and mathematics.

The acronym AEGIS stands for Autonomous Equity Grading and Intelligence System. The name reflects the system's dual function: it grades companies along a rich set of quantitative and qualitative dimensions, and it applies genuine analytical intelligence rather than mere mechanical screening to derive investment signals. The word autonomous is deliberate. AEGIS is designed to operate continuously, without human intervention in the data ingestion and initial analysis pipeline, surfacing to the human analyst only those situations where the margin of opportunity is sufficiently large or the analytical signal sufficiently novel to warrant direct attention.

This paper is organized as follows. Section 2 reviews the intellectual foundations on which AEGIS rests, drawing together contributions from fundamental valuation theory, machine learning, natural language processing, Bayesian statistics, and behavioral finance. Section 3 presents the system architecture of AEGIS, describing the data ingestion layer, the LLM processing pipeline, the quantitative scoring engine, and the synthesis and output layer. Section 4 develops the mathematical formalism of the stochastic DCF engine at the heart of AEGIS. Section 5 presents the multi-factor composite scoring model. Section 6 describes the LLM pipeline in detail, including prompting design, retrieval-augmented generation, and extraction of structured financial signals from unstructured text. Section 7 covers management quality assessment. Section 8 covers competitive moat quantification. Section 9 presents the macroeconomic regime integration layer. Section 10 describes the continuous monitoring and Bayesian update mechanism. Section 11 discusses portfolio integration and position sizing. Section 12 presents illustrative case applications. Section 13 discusses limitations and areas for further development. Section 14 concludes.

Theoretical Foundations

Intrinsic Value as a Probability Distribution

The intellectual lineage of AEGIS begins with the most basic principle in finance: the value of any asset is the present value of the cash flows it is expected to generate. Williams formalized this intuition in the dividend discount model, and it was extended by Modigliani and Miller and subsequently by Gordon into the framework that practitioners recognize as discounted cash flow analysis [1-3]. The standard formulation, which VRC adopts as its point of departure, is:

$$V_0 = \sum_{t=1}^{T} \frac{FCFF_t}{(1+WACC)^t} + \frac{TV_T}{(1+WACC)^T} \tag{1}$$

where $FCFF_t$ denotes free cash flow to the firm in period t, WACC is the weighted average cost of capital, and TV_T is the terminal value at the end of the explicit forecast horizon T .

The Achilles heel of this formulation is immediately apparent: its inputs are deeply uncertain. Damodaran demonstrates that the terminal value frequently accounts for sixty to eighty percent of the total estimated value under standard parameterizations, creating enormous sensitivity to the assumed long-run growth rate [4]. A perturbation of one hundred basis points in the terminal growth rate can shift the estimated value by twenty to forty percent for a typical mid-cap company, rendering single-point intrinsic value estimates fundamentally unreliable. AEGIS resolves this problem not by pretending to greater precision but by abandoning the pretense of precision entirely. Rather than computing a single intrinsic value V_0, AEGIS computes a probability distribution over intrinsic values, $P(V_0)$. This distribution is derived from explicitly probabilistic inputs, and its summary statistics, particularly the mean, median, and key quantiles, constitute

the operational output of the valuation engine.

Definition 2.1 (Probabilistic Intrinsic Value)**.** Let $\boldsymbol{\theta} = (g_1, g_2, \ldots, g_T, WACC, g_\infty, m_1, \ldots, m_T, \rho_1, \ldots$

denote the vector of inputs to the DCF model, where g_t is the revenue growth rate in period t, g_∞ is the terminal growth rate, m_t is the operating margin in period t, and ρ_t is the reinvestment rate in period t. Each element of $\boldsymbol{\theta}$ is treated as a random variable with a prior distribution $\pi(\theta_i)$. The probabilistic intrinsic value is then the random variable:

$$V_0(\boldsymbol{\theta}) = \sum_{t=1}^{T} \frac{FCFF_t(\boldsymbol{\theta})}{(1 + WACC(\boldsymbol{\theta}))^t} + \frac{TV_T(\boldsymbol{\theta})}{(1 + WACC(\boldsymbol{\theta}))^T} \tag{2}$$

and the valuation output is the induced distribution $P(V_0) = V_0 \circ P(\boldsymbol{\theta})$.

This formulation has deep practical implications. It forces the analyst to be explicit about what is uncertain, how uncertain it is, and how different sources of uncertainty interact. It produces a range of outcomes rather than a false point estimate. And it makes the margin of safety concept, central to the Graham-Dodd tradition that VRC follows, computationally tractable: the margin of safety is simply the probability that the market price lies below the 25th percentile of the intrinsic value distribution.

Bayesian Belief Updating

The second theoretical pillar of AEGIS is Bayesian inference. The Bayesian framework is the natural language for expressing the idea that valuation is a continuously updated belief rather than a one-time calculation. Jacobs and Levy advocate for Bayesian thinking in portfolio management, and the argument applies with particular force to fundamental valuation, where the arrival of new information, whether in the form of an earnings release, a management change, or a macroeconomic shock, should produce a principled revision to the analyst's prior beliefs [5].

Formally, let $\pi_t(V_0)$ denote the analyst's prior distribution over intrinsic values at time t. Upon observing new evidence E_{t+1} (for example, a quarterly earnings release), the posterior is given by Bayes' theorem:

$$\pi_{t+1}(V_0 \mid \mathcal{E}_{t+1}) \propto \mathcal{L}(\mathcal{E}_{t+1} \mid V_0) \cdot \pi_t(V_0) \tag{3}$$

where $L(E_{t+1} \mid V_0)$ is the likelihood of observing evidence E_{t+1} given that the true intrinsic value is V_0. In practice, computing this likelihood exactly is infeasible, and AEGIS employs a particle filter approximation that is described in Section 10.

Large Language Models and Financial Text Analysis

The third theoretical pillar is the application of large language models to financial text analysis. The systematic extraction of qualitative information from annual reports, earnings call transcripts, and regulatory filings has historically been beyond the reach of purely quantitative approaches. Recent advances in transformer-based language models, particularly instruction-tuned models capable of following complex analytical prompts, have opened the possibility of automating this extraction at scale.

Formally, let $D_{i,t}$ denote the corpus of textual documents associated with company i at time t (annual reports, earnings transcripts, regulatory filings, news articles, and so forth). AEGIS employs a retrieval-augmented generation architecture in which relevant passages are retrieved from $D_{i,t}$ and presented to an LLM along with structured prompts designed to elicit specific analytical judgments.

The output of the LLM pipeline is a set of structured signals $\mathbf{s}_{i,t}^{LLM}$ that are subsequently incorporated into the quantitative scoring model.

This approach draws on the broader literature on information extraction from financial texts while extending it through the use of instruction-tuned LLMs capable of more nuanced analytical reasoning than bag-of-words or sentiment scoring approaches [6,7].

Multi-Factor Scoring and the Information Ratio

The fourth theoretical pillar is multi-factor modeling, which provides the scaffold for AEGIS's composite scoring system. Fama and French established the empirical case for multi-factor equity models, and subsequent work by Carhart, Hou, Xue, and Zhang, and Asness, Moskowitz, and Pedersen has expanded the factor zoo [8-12]. AEGIS does not seek to add to this literature; rather, it draws on it selectively to construct a scoring system that is motivated by both statistical evidence and economic intuition.

The composite company score $\Omega_{i,t}$ produced by AEGIS is a weighted linear combination of sub-scores across dimensions that correspond to the analytical framework described qualitatively in prior VRC research:

$$\Omega_{i,t} = \sum_{k=1}^{K} w_k \cdot S_{k,i,t} \tag{4}$$

where $S_{k,i,t}$ is the normalized score of company i at time t on dimension k, and w_k is the weight assigned to that dimension. The dimensions and their theoretical motivations are elaborated in Section 5.

Valuation as a Mindset: The VRC Framework

Prior VRC research has argued that valuation is best understood not as a mechanical calculation but as an iterative, probabilistic, and inherently subjective process of inferring intrinsic worth from uncertain inputs [13]. AEGIS does not contradict this view. Rather, it operationalizes it. The system automates the mechanical and data-intensive components of the valuation process, freeing the human analyst to focus on the dimensions of analysis where judgment is irreplaceable: assessing the credibility of a management team, evaluating the durability of a competitive advantage, or weighing the implications of a strategic pivot.

The philosophical alignment between AEGIS and the broader VRC investment framework is important. AEGIS is not a black box that produces recommendations to be followed mechanically. It is a tool that structures, quantifies, and scales the analytical process that VRC would conduct manually for any individual company, extending it to hundreds of companies simultaneously and providing a continuous surveillance capability that no human team could replicate.

System Architecture of AEGIS

Overview

AEGIS is organized into four major functional layers, each of which feeds into the next in a pipeline architecture:

- **Data Ingestion Layer**: Responsible for acquiring, parsing, and storing all raw data inputs.
- **Intelligence Processing Layer**: Responsible for extracting structured signals from raw data using LLMs, quantitative models, and specialized analytical modules.
- **Valuation Synthesis Layer**: Responsible for combining signals into probability-weighted intrinsic value estimates and composite company scores.
- **Monitoring and Output Layer**: Responsible for continuous surveillance, Bayesian updating, alert generation, and presentation of results to the investment team.

These layers are described in detail in the subsections that follow, and the data flows between them are summarized in the system diagram presented in Figure ??.

Data Ingestion Layer

The data ingestion layer is the foundation of AEGIS. Its function is to acquire, validate, clean, and store all data required by the downstream processing pipeline. The layer operates continuously, ingesting new data as it becomes available through a combination of scheduled batch processes and event-driven triggers.

Annual Report and Filing Acquisition

Annual reports (Form 10-K in the US, Form 20-F for foreign private issuers, and their international equivalents) are the primary source of fundamental financial data in AEGIS. The system acquires these documents through a combination of direct EDGAR API access for US-listed companies, regulatory body APIs for international markets (SEBI for India, FCA for the UK, ESMA for Europe, and so forth), and financial data vendors where direct API access is unavailable.

Documents are acquired in machine-readable formats where available (iXBRL, XBRL) and in PDF format where not. XBRL-tagged financials are parsed directly into structured data representations. PDF documents are processed through an OCR pipeline followed by an LLM-based extraction routine described in Section 6.

Earnings Call Transcript Acquisition

Earnings call transcripts represent one of the richest sources of qualitative informa- tion about a company's competitive position, management quality, and forward guidance. AEGIS acquires transcripts from multiple sources, including Refinitiv, Bloomberg, Seeking Alpha, and direct company investor relations portals. Transcripts are timestamped, attributed to named speakers (management ver- sus analyst), and stored in a structured format suitable for downstream LLM processing.

Macroeconomic and Sector Data

Macroeconomic data includes central bank policy rates, inflation indicators, GDP growth estimates, credit spreads, currency exchange rates, commodity prices, and sector-level indicators such as industrial production, retail sales, and credit growth. These data are sourced from the Federal Reserve Economic Data (FRED) system, national statistics agencies, the World Bank API, and commercial data providers. The update cadence varies by series, from real-time for market prices to quarterly for GDP releases.

Alternative Data Feeds

AEGIS incorporates a growing set of alternative data signals designed to provide early-warning indicators of fundamental change. These include web traffic data (from SimilarWeb), job posting trends (from LinkedIn and Indeed), satellite imagery for retail footfall and manufacturing activity, patent filing data, employee review scores (from Glassdoor), consumer sentiment from product reviews (from Amazon, Google Reviews, and App Store ratings), and legal database feeds for litigation tracking.

Market Data

Real-time and historical price, volume, short interest, options implied volatility, and institutional ownership data are acquired from standard market data vendors. These data are used primarily in the valuation comparison and alert generation subsystems.

Intelligence Processing Layer

The intelligence processing layer transforms raw data into structured analytical signals. It consists of five specialized modules: the LLM Document Analysis Module, the Financial Ratio Engine, the Competitive Intelligence Module, the Management Assessment Module, and the Macroeconomic Regime Module.

Valuation Synthesis Layer

The valuation synthesis layer aggregates signals from the intelligence processing layer into two primary outputs: the probability-weighted intrinsic value distribution $P(V_{0,i,t})$ for each company i at time t, and the composite AEGIS score $\Omega_{i,t}$. The mathematical details of these computations are presented in Sections 4 and 5 respectively.

Monitoring and Output Layer

The monitoring layer operates continuously, comparing market prices to intrinsic value distributions, tracking changes in composite scores, and generating alerts when predefined thresholds are breached. It is also responsible for the Bayesian updating of valuation priors as new data arrives. The design of this layer is described in Section 10.

Stochastic Discounted Cash Flow Engine

Model Overview

The core of AEGIS's valuation engine is a stochastic DCF model that replaces the fixed scalar inputs of classical DCF analysis with random variables parameterized by distributions fitted to historical data, analyst priors, and LLM-extracted for- ward signals. The model operates in three stages: scenario generation, cash flow projection, and present value computation.

Distributional Parameterization of Inputs

Let company i be parameterized by a set of K random input variables collected in the vector $\boldsymbol{\theta}_i = (\theta_{i,1}, \ldots, \theta_{i,K})$. The key inputs and their distributional representations are as follows.

Revenue Growth Rate

Revenue growth in period t is modeled as a mean-reverting process. Let g_t denote the revenue growth rate in year t. The model specifies:

$$g_t = \bar{g} + (g_{t-1} - \bar{g})e^{-\kappa} + \sigma_g \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, 1) \tag{5}$$

where $\bar{g}$ is the long-run mean growth rate (a company-specific prior calibrated to the sector and historical track record), $\kappa > 0$ is the mean-reversion speed (reflecting competitive dynamics), and σ_g is the volatility of growth surprises. The prior mean $\bar{g}$ is set equal to a weighted average of the company's historical growth rate, the sector growth rate, and the macroeconomic nominal GDP growth forecast, with weights informed by the company's competitive position score.

Operating Margin

The operating margin in period t, denoted m_t, is modeled as:

$$m_t = m^* + (m_{t-1} - m^*)e^{-\lambda} + \sigma_m \eta_t, \quad \eta_t \sim \mathcal{N}(0, 1) \tag{6}$$

where m^* is the long-run sustainable operating margin, λ is the margin mean- reversion speed, and σ_m is margin volatility. The sustainable margin m^* is estimated from historical data using the 60th percentile of realized operating margins over the most recent seven-year cycle, adjusted downward for companies facing secular competitive pressures (as flagged by the Competitive Intelligence Module) and upward for companies with demonstrated and durable pricing power.

Reinvestment Rate and Capital Efficiency

The reinvestment rate ρ_t captures the fraction of after-tax operating income that must be reinvested to support the projected growth rate. It is linked to growth through the return on invested capital:

$$\rho_t = \frac{g_t}{ROIC_t} \tag{7}$$

where $ROIC_t$ is the return on invested capital in period t. $ROIC_t$ itself follows a mean-reverting process toward the cost of capital, a formalization of the competitive equilibrium dynamics described by Mauboussin and Callahan (2014):

$$ROIC_t = WACC + (ROIC_{t-1} - WACC)e^{-\mu} + \sigma_R \xi_t \tag{8}$$

where μ is the fade rate of excess returns and $\xi_t \sim N(0, 1)$. Companies with strong competitive moat scores (Section 8) receive lower values of μ, reflecting slower erosion of excess returns.

Free Cash Flow Construction

Free cash flow to the firm in period t is derived from the projected revenue, margin, and reinvestment rate as follows. Let R_t denote revenue in period t. Then:

$$R_t = R_{t-1}(1 + g_t) \tag{9}$$

$$EBIT_t = m_t \cdot R_t \tag{10}$$

$$NOPAT_t = EBIT_t \cdot (1 - \tau_t) \tag{11}$$

$$RI_t = \rho_t \cdot NOPAT_t \tag{12}$$

$$FCFF_t = NOPAT_t - RI_t = NOPAT_t(1 - \rho_t) \tag{13}$$

where τ_t is the effective tax rate (assumed to follow a slow mean-reverting process toward the statutory rate).

Terminal Value Computation

The terminal value at the end of year T is computed using the Gordon Growth Model applied to the normalized terminal-year free cash flow:

$$TV_T = \frac{FCFF_T(1 + g_\infty)}{WACC - g_\infty} \tag{14}$$

The terminal growth rate g_∞ is treated as a truncated normal random variable with support [−0.02, WACC − 0.01] to prevent economically absurd terminal assumptions. Its mean is set equal to the expected long-run nominal GDP growth rate of the primary market in which the company operates.

Cost of Capital Estimation

The weighted average cost of capital is estimated as:

$$WACC = \frac{E}{V}r_e + \frac{D}{V}r_d(1 - \tau) \tag{15}$$

The cost of equity r_e is estimated using a build-up approach that augments the standard CAPM with a size premium and a country risk premium where applicable:

$$r_e = r_f + \beta_i \cdot ERP + \lambda_s \cdot SRP + \lambda_c \cdot CRP \tag{16}$$

where r_f is the risk-free rate, ERP is the equity risk premium, SRP is the size risk premium, CRP is the country risk premium (after Damodaran, 2003), β_i is the company's systematic risk, and λ_s, λ_c are exposure parameters estimated from the company's revenue and asset geography. The risk-free rate is updated continuously from current sovereign bond yields.

Monte Carlo Simulation

Given the distributional parameterization of inputs, AEGIS generates the intrinsic value distribution through a large-scale Monte Carlo simulation. At each time step t when a major update is triggered, the system:

1. Draws $N = 50{,}000$ independent realizations of the input vector $\boldsymbol{\theta}_i$ from the current joint prior distribution.
2. For each realization, computes the projected cash flows $\{FCFF_t^{(j)}\}_{t=1}^{T}$ and terminal value $TV_T^{(j)}$ using equations (9) through (13).
3. Computes the present value $V_0^{(j)}$ using equation (1).
4. Constructs the empirical distribution of $\{V_0^{(j)}\}_{j=1}^{N}$.

The resulting distribution is summarized by its mean $\hat{V}^{mean}$, median $\hat{V}^{med}$, and quantiles $\hat{V}^{(0.10)}, \hat{V}^{(0.25)}, \hat{V}^{(0.75)}, \hat{V}^{(0.90)}$.

Probability of Undervaluation

A key output of the Monte Carlo engine is the probability that the market price P_i lies below the estimated intrinsic value:

$$PU_i = \mathbb{P}(V_0 > P_i) = \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}[V_0^{(j)} > P_i] \tag{17}$$

This probability is a central input into the opportunity screening system and the position sizing model. A company with $PU_i > 0.85$ is flagged as a high-priority opportunity; one with $PU_i < 0.20$ is flagged as potentially overvalued.

Scenario Decomposition

For each company, AEGIS maintains an explicit three-scenario decomposition consistent with the probabilistic scenario analysis recommended in prior VRC research. The three scenarios are defined as follows:

Table 1: AEGIS Scenario Framework

Scenario	Description	Probability	Growth	Margin
Bull	Full realization of competitive advantage	π_1	High	Expanding
Base	Moderate execution with some headwinds	π_2	Medium	Stable
Bear	Structural deterioration or disruption	π_3	Low	Contracting

The probability-weighted intrinsic value under this framework is:

$$\hat{V}^{SW} = \pi_1 V^{Bull} + \pi_2 V^{Base} + \pi_3 V^{Bear} \tag{18}$$

Scenario probabilities (π_1, π_2, π_3) are not fixed but are updated dynamically by AEGIS based on signals from the LLM pipeline, the macroeconomic regime module, and the management quality assessor.

Multi-Factor Composite Scoring Model

Design Philosophy

The composite AEGIS score $\Omega_{i,t}$ is designed to capture the quality of a business along dimensions that theory and practice both identify as important determinants of long-run value creation. It is explicitly not a short-term momentum or sentiment score; it is a fundamental quality and opportunity assessment designed to complement the DCF valuation.

The score is organized into five primary domains, each comprising several sub-dimensions:

1. **Financial Quality** (Q_F): Capital efficiency, earnings quality, and balance sheet strength.
2. **Competitive Position** (Q_C): Moat breadth, pricing power, and market share dynamics.
3. **Management Quality** (Q_M): Integrity, capital allocation track record, and employee and stakeholder relations.
4. **Growth Opportunity** (Q_G): Total addressable market, penetration rate, and secular tailwinds.
5. **Valuation Attractiveness** (Q_V): Distance from intrinsic value and margin of safety.

Financial Quality Score

The financial quality score Q_F is computed from a set of accounting-based metrics designed to capture the durability and reliability of reported earnings:

$$Q_F = w_{F1} \cdot ROIC_{norm} + w_{F2} \cdot ROE_{adj} + w_{F3} \cdot FCF\_Conv + w_{F4} \cdot Accrual\_Score + w_{F5} \cdot Leverage\_Score \tag{19}$$

Each component is described below.

Normalized Return on Invested Capital

The normalized ROIC is computed as the median ROIC over the most recent five-year period, adjusted for the economic depreciation of intangible assets and the capitalization of operating leases following Damodaran's methodology [4]:

$$ROIC_{norm} = \text{median}_{t \in \{T-4,\ldots,T\}} \left(\frac{NOPAT_t}{IC_t^{adj}} \right) \tag{20}$$

where IC_t^{adj} is the adjusted invested capital inclusive of capitalized R&D, operating leases, and goodwill (net of amortization). This measure is normalized to the cross-sectional distribution within the company's GICS sector, yielding a percentile rank between 0 and 1.

Free Cash Flow Conversion

Free cash flow conversion measures the fraction of net income that is converted to free cash flow, providing a check on earnings quality:

$$FCF\_Conv = \frac{\text{5-yr cumulative FCF}}{\text{5-yr cumulative net income}} \tag{21}$$

Low conversion ratios may indicate aggressive revenue recognition, capitalization of operating costs, or deteriorating working capital management [14].

Accruals Score

The accruals score is derived from the Sloan (1996) accrual ratio, which measures the fraction of earnings attributable to non-cash accrual components:

$$Accrual\_Ratio = \frac{\Delta NOA}{(NOA_t + NOA_{t-1})/2} \tag{22}$$

where NOA denotes net operating assets. Higher accrual ratios are associated with lower subsequent returns and greater earnings restatement risk [15].

Leverage Score

The leverage score combines several measures of financial flexibility and debt burden:

$$Leverage\_Score = f\left(\frac{NetDebt}{EBITDA}, \frac{EBIT}{InterestExpense}, \frac{FCF}{TotalDebt}\right) \quad (23)$$

where f(·) is a logistic transformation that maps the combined metric to the unit interval.

Competitive Position Score

The competitive position score Q_C is the output of the dedicated competitive moat module described in Section 8. At a high level, it aggregates signals about:

- The stability and breadth of gross margins relative to sector peers.
- The trend in market share within the primary addressable market.
- The company's pricing power, as measured by the pass-through rate of input cost increases.
- The barriers to entry facing potential competitors, as assessed by the LLM analysis of competitive landscape disclosures.
- The company's intellectual property position (patent counts, R&D efficiency).

Management Quality Score

The management quality score Q_M is the output of the dedicated management assessment module described in Section 7. It incorporates both quantitative metrics (capital allocation efficiency, insider ownership, compensation structure) and LLM- extracted qualitative signals (tone and consistency of management communication, litigation history, ESG controversy score).

Growth Opportunity Score

The growth opportunity score Q_G integrates estimates of the total addressable market (TAM), the company's current penetration rate, and the projected secular growth of the addressable market:

$$Q_G = \alpha_1 \cdot \log(TAM) + \alpha_2 \cdot (1 - Penetration) + \alpha_3 \cdot TAM\_CAGR + \alpha_4 \cdot Innovation\_Score \quad (24)$$

where TAM is the estimated total addressable market size (in USD billions), Penetration is the company's current revenue as a fraction of TAM , TAM CAGR is the projected compound annual growth rate of the addressable market over the next five years, and Innovation Score is a composite of R&D intensity, patent citation velocity, and new product launch cadence. The coefficients α_k are calibrated to a training dataset of historical growth realizations.

Valuation Attractiveness Score

The valuation attractiveness score Q_V directly incorporates the output of the stochastic DCF engine:

$$Q_V = f(PU_i) \cdot f\left(\frac{\hat{V}^{med} - P_i}{\hat{V}^{med}}\right) \quad (25)$$

where PU_i is the probability of undervaluation defined in equation (17) and f (·) is a monotone transformation. Additionally, relative valuation multiples (P/E, EV/EBIT, P/B) are compared to the company's own historical distribution and to sector peers, and the resulting percentile ranks are incorporated:

$$Q_V^{rel} = w_1 \cdot Rank_{P/E} + w_2 \cdot Rank_{EV/EBIT} + w_3 \cdot Rank_{P/B} + w_4 \cdot Rank_{FCF\_yield} \quad (26)$$

For financial institutions, where the balance sheet is the business, the P/B ratio and the excess return model receive substantially higher weights in Q_V [4].

Composite Score Computation and Weighting

The composite AEGIS score is the weighted combination:

$$\Omega_i = w_{QF} Q_F + w_{QC} Q_C + w_{QM} Q_M + w_{QG} Q_G + w_{QV} Q_V \quad (27)$$

The default weights reflect VRC's investment philosophy, which places the highest value on financial quality and valuation attractiveness:

Table 2: Default Weights in the AEGIS Composite Score

Dimension	Symbol	Default Weight
Financial Quality	Q_F	0.25
Competitive Position	Q_C	0.25
Management Quality	Q_M	0.20
Growth Opportunity	Q_G	0.15
Valuation Attractiveness	Q_V	0.15
Total		1.00

These weights are not static: AEGIS adjusts them dynamically based on the macroeconomic regime. In a late-cycle environment characterized by elevated credit spreads, the weight on Q_F (financial quality) increases. In an early-cycle environment with strong nominal growth, the weight on Q_G increases.

LLM Pipeline Architecture

Design Principles

The LLM pipeline in AEGIS is designed around four principles. First, modularity: each analytical task (revenue driver extraction, management tone analysis, competitive moat identification, risk factor summarization) is handled by a separate, task-specific module with its own prompt design and output schema. Second, grounding: all LLM outputs are anchored to specific source passages through a retrieval-augmented generation architecture, preventing hallucination of facts not present in the source documents. Third, structured output: LLM modules produce structured outputs in a defined schema that can be directly ingested by downstream quantitative models. Fourth, calibration: LLM confidence scores are calibrated against a holdout set of human analyst assessments to ensure that expressed certainty corresponds to actual accuracy.

Retrieval-Augmented Generation Architecture

The AEGIS RAG system operates as follows. For each company i and each analytical module m, a query q_m is formulated. A dense retrieval model (specifically, a fine-tuned bi-encoder based on Sentence-BERT architecture, Reimers and Gurevych [16]) retrieves the top-k passages from $D_{i,t}$ that are most semantically relevant to q_m. These passages, together with the query and a task-specific system prompt, are passed to the language model, which produces a structured analytical response.

Formally, let Retrieve(q_m, $D_{i,t}$, k) denote the function that returns the k most relevant passages from $D_{i,t}$ for query q_m. The LLM analytical output for module m is:

$$\hat{s}_{m,i,t}^{LLM} = \text{LLM}\left(\text{Prompt}_m, q_m, \text{Retrieve}(q_m, \mathcal{D}_{i,t}, k)\right) \quad (28)$$

Annual Report Processing Module

The annual report processing module extracts the following structured signals from each annual report:

- Quantitative revenue driver decomposition: organic versus acquired growth, volume versus price contribution, domestic versus international split.
- Operating leverage assessment: sensitivity of margins to volume changes.
- Capital expenditure classification: maintenance versus growth capex.
- Balance sheet quality flags: concentration of receivables, inventory build, goodwill as a fraction of assets.
- Risk factor novelty score: detection of newly disclosed risk factors relative to the prior year's filing.
- Forward guidance extraction: structured extraction of management guidance with confidence intervals.

Earnings Call Transcript Module

The earnings call module processes transcripts in two phases. The management phase analyzes only those passages attributed to company management, extracting:

- Linguistic confidence scores: hedging language frequency, use of qualifiers, and changes in hedging intensity relative to prior quarters.
- Topic coverage completeness: were key operational and strategic questions addressed, or were they deflected?
- Guidance revision analysis: did management raise, maintain, or lower guidance relative to the prior quarter, and was the revision accompanied by credible explanation?
- Sentiment trajectory: is the tone of management communication more optimistic or more cautious than in previous transcripts?
- The analyst question phase analyzes sell-side analyst questions to identify areas of concern, recurring doubts, and topics that management consistently avoids.

Competitive Intelligence Module

The competitive intelligence module processes a broader corpus of documents, including competitor filings, industry analyst reports, trade publications, and customer review data, to produce signals about the competitive dynamics of the industry:

- Competitive intensity score: based on pricing trends, new entrant activity, and technology disruption signals.
- Differentiation score: based on proprietary analysis of product and service differentiation relative to competitors.
- Market share trend: derived from revenue growth differentials with named competitors.

Management Quality Extraction

The management quality LLM module is designed to extract signals that quantitative financial models cannot directly observe. These include:

- The consistency between what management says and what the financials subsequently reveal.

- The presence or absence of self-serving language (blaming external factors for underperformance, claiming full credit for tailwind-driven success).
- The sophistication of management's discussion of capital allocation trade-offs.
- Any disclosures suggesting legal, ethical, or regulatory issues.

Structured Output Schema

All LLM modules produce outputs in a standardized JSON schema. Each field in the schema includes a numeric score (where applicable), a confidence level, and a supporting evidence string containing the verbatim passages from source documents that justify the score. This schema ensures that every LLM-derived signal is traceable, auditable, and explainable.

Management Quality Assessment

The Importance of Management in Fundamental Analysis

Buffett and Munger (1995) both emphasize that the quality of management is one of the most important determinants of long-run investment outcomes [17]. A great business with poor management tends toward mediocrity; a mediocre business with extraordinary management can be transformed over time. The challenge is that management quality is exceptionally difficult to quantify, and the measures that practitioners traditionally use, such as tenure and biographical credentials, have limited empirical validity.

AEGIS addresses this challenge by combining a set of quantitative capital allocation metrics with LLM-extracted qualitative signals into a composite management quality score Q_M .

Capital Allocation Track Record

The quantitative component of Q_M is built around a framework for assessing the quality of management's capital allocation decisions over time. The framework evaluates each major capital deployment decision (acquisition, organic investment, share repurchase, dividend, or debt paydown) against the return on capital subsequently realized.

Let $A_i = \{a_1, a_2, . . . , a_L\}$ denote the set of major capital allocation decisions made by management of company i over the prior seven years. For each decision a_l, AEGIS estimates an ex-post return on capital r_l based on subsequent financial performance attributable to that decision. The capital allocation quality score is then:

$$CA\_Score = \frac{1}{L}\sum_{l=1}^{L} \mathbf{1}[r_l > WACC] + \text{bias}(r_l - WACC) \qquad (29)$$

where the second term captures the average magnitude of value creation above the cost of capital across successful investments, and the first term counts the fraction of decisions that cleared the hurdle rate.

Insider Ownership and Skin in the Game

Jensen and Meckling demonstrate that managerial incentives are better aligned with shareholder interests when managers hold significant equity stakes in the company [18]. AEGIS incorporates insider ownership as a direct input into the management quality score, using regulatory filing data on insider holdings:

$$IO\_Score = f\left(\frac{\text{Total insider ownership}}{\text{Shares outstanding}}\right) \tag{30}$$

where f (·) is a sigmoid function that rewards meaningful but not excessive insider concentration.

Compensation Structure Assessment

Management compensation structures are analyzed using a dual test. First, AEGIS checks whether performance metrics used in incentive plans are strongly correlated with long-run value creation or primarily with short-term earnings manipulation targets. Second, it assesses the absolute level of compensation relative to peers, normalized by company size, to identify potential governance failures.

LLM-Based Qualitative Assessment

The qualitative component of Q_M is produced by the management quality LLM module described in Section 6.5. The resulting signals are integrated into Q_M through a weighted average:

$$\begin{aligned} Q_M = w_{M1} \cdot CA\_Score + w_{M2} \cdot IO\_Score + w_{M3} \cdot Comp\_Score \\ + w_{M4} \cdot LLM\_Mgmt\_Score + w_{M5} \cdot ESG\_Score \end{aligned} \tag{31}$$

where ESG Score is a weighted average of employee satisfaction metrics (from Glassdoor), environmental controversy flags, and litigation score derived from legal database monitoring.

Controversy Detection

AEGIS maintains a continuous litigation and controversy monitoring system. Legal databases (PACER, court records APIs) and news aggregation feeds are monitored for filings, judgments, and regulatory actions involving covered companies. When a material controversy is detected, Q_M is immediately revised downward, and the human investment team is alerted.

Competitive Moat Quantification

Theoretical Framework

The concept of the economic moat, popularized by Buffett and subsequently formalized by Porter and Greenwald et al., refers to a structural competitive advantage that enables a business to earn returns on invested capital above its cost of capital over extended periods [19,20]. AEGIS operationalizes this concept through a quantitative moat scoring framework that draws on both financial data and LLM-extracted qualitative signals.

The framework distinguishes five sources of competitive advantage:

- **Cost advantage**: The ability to produce goods or services at lower cost than competitors.
- **Network effects**: Value that increases as more participants join the network.
- **Switching costs**: The cost, friction, or risk that customers must bear to switch to a competing product.
- **Intangible assets**: Patents, brands, regulatory licenses, and proprietary data that competitors cannot easily replicate.
- **Efficient scale**: The ability to serve a market of limited size at lowest cost,

deterring entry.

Moat Width Measurement

The width of each moat source is measured through a combination of financial metrics and LLM signals:

Gross Margin Stability

For cost advantage moats, the primary metric is the stability and level of gross margins relative to sector peers. AEGIS computes:

$$GMS_i = 1 - \frac{\sigma(GM_i)}{\mu(GM_i)} \cdot \frac{\mu(GM_i) - \overline{GM}_{sector}}{\sigma_{sector}(GM)} \qquad (32)$$

where GM_i is the vector of annual gross margins over the prior seven years, and the sector statistics are computed cross-sectionally.

Market Share Trajectory

Network effects and switching cost moats are reflected in the market share trajectory of the company. AEGIS tracks market share (estimated from revenue growth relative to identified competitors) and fits a trend model:

$$MS_{i,t} = MS_{i,0} \cdot e^{\gamma_i t} + \epsilon_t \qquad (33)$$

A positive and statistically significant $\hat{\gamma}_i$ is treated as evidence of a widening moat; a negative $\hat{\gamma}_i$ as evidence of moat erosion.

Pricing Power Index

The pricing power index is one of the most diagnostically important signals in AEGIS. It measures the extent to which a company can pass through cost increases to customers without losing market share:

$$PPI_i = \frac{\Delta \bar{P}_i - \Delta CPI}{\Delta CPI_{inputs}} \qquad (34)$$

where $\Delta \bar{P}_i$ is the estimated change in realized price per unit, ΔCPI is the general consumer price inflation, and ΔCPI_{inputs} is the inflation in the company's input cost basket. A $PPI_i > 1$ indicates that the company is capturing more than its fair share of value in the supply chain, a hallmark of genuine pricing power.

Composite Moat Score

The composite competitive position score is:

$$Q_C = \sum_{j=1}^{5} \phi_j \cdot MS_j \qquad (35)$$

where MS_j is the moat score for source j (cost, network, switching, intangible, scale) and ϕ_j is the weight assigned to source j, which varies by industry sector. For example, for technology platforms, the network effects weight is highest; for pharmaceutical companies, intangible assets (patents) dominate.

Macroeconomic Regime Integration

The Importance of Macroeconomic Context

Valuation cannot be conducted in a macroeconomic vacuum. Interest rate cycles affect the discount rate applied to future cash flows; credit cycles affect the availability and cost of debt financing; currency movements affect the reported earnings of multinational companies; and commodity cycles affect the input cost structures of energy-intensive businesses. AEGIS integrates macroeconomic data into the valuation framework through a dedicated regime detection and sensitivity analysis layer.

Regime Classification

AEGIS classifies the macroeconomic environment at any given time into one of four regimes using a Hidden Markov Model (HMM) fitted to a vector of macroeconomic indicators:

$$\text{Regime}_t \in \{R_1, R_2, R_3, R_4\} \quad (36)$$

The four regimes correspond broadly to: expansion with low inflation (R_1), expansion with high inflation (R_2), contraction with elevated spreads (R_3), and recession with deflation risk (R_4). The indicator vector includes the yield curve slope, credit spreads, ISM manufacturing PMI, housing starts, initial jobless claims, commodity price indices, and currency volatility.

The HMM transition matrix A and emission parameters are re-estimated quarterly using the Baum-Welch algorithm:

$$\hat{\mathbf{A}}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}} = \arg\max_{\mathbf{A},\boldsymbol{\mu},\boldsymbol{\Sigma}} \mathbb{P}(\mathbf{x}_{1:T} \mid \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (37)$$

The current regime probability vector $\boldsymbol{\pi}^{\text{regime}} = (P(R_1 \mid \mathbf{x}_{1:t}), \ldots, P(R_4 \mid \mathbf{x}_{1:t}))$ is updated using the forward algorithm at each data release.

Regime-Conditional Valuation Adjustments

The macroeconomic regime affects valuation through several channels. First, it affects the risk-free rate, which enters the WACC calculation in equation (15). Sec- ond, it affects the equity risk premium, which VRC estimates using Damodaran's implied ERP methodology applied to the current market level and consensus earnings forecasts [4]. Third, it affects the mean of the growth rate distribution $\bar{g}$ in equation (5), since macroeconomic conditions affect the near-term revenue trajectory of virtually all companies.

Regime-conditional adjustments are implemented as follows. Let Δ^{regime}_k denote the regime-specific adjustment to input parameter k. The adjusted mean of input parameter k given regime probabilities is:

$$\tilde{\mu}_k = \mu_k^{base} + \sum_{r=1}^{4} \mathbb{P}(R_r \mid \mathbf{x}_{1:t}) \cdot \Delta_k^{R_r} \quad (38)$$

Sector Sensitivity Mapping

Not all companies are equally sensitive to macroeconomic regimes. A regulated utility is relatively insensitive to the business cycle; a cyclical steel producer is highly sensitive. AEGIS maintains a sector sensitivity matrix that maps GICS sector classification to regime sensitivity coefficients, and these coefficients are used to scale the regime adjustment Δ^{regime}_k for each company.

Continuous Monitoring and Bayesian Updating

The Surveillance Architecture

One of the defining features of AEGIS is its continuous surveillance capability. Unlike a static valuation model that is run periodically, AEGIS operates as a persistent monitoring system that tracks every covered company in real time and triggers valuation updates whenever new information arrives that is material enough to shift the probability distribution over intrinsic values.

Event Taxonomy and Trigger Design

AEGIS maintains a hierarchical taxonomy of events that trigger valuation updates. Events are classified by their expected impact on the intrinsic value distribution:

- **Tier 1 events** (full re-simulation): Annual report release, significant earnings miss or beat (exceeding two standard deviations), major capital allocation announcement (large acquisition, transformative share buyback), regulatory action, or material management change.
- **Tier 2 events** (parameter update without full re-simulation): Quarterly earnings release, guidance revision, analyst day, macro data release affecting the company's primary market.
- **Tier 3 events** (signal update only): Pricing change announcement, product launch, competitor filing, or news article flagged as material by the LLM sentiment monitor.

Particle Filter Implementation

For Tier 2 and Tier 3 updates, AEGIS employs a particle filter to update the posterior distribution over intrinsic values without requiring a full Monte Carlo re-simulation. The particle filter approximates the posterior using a weighted set of N particles $\{V_0^{(j)}, w^{(j)}\}_{j=1}^{N_p}$:

$$\tilde{w}^{(j)} = w^{(j)} \cdot \mathcal{L}(\mathcal{E}_{t+1} \mid V_0^{(j)}) \tag{39}$$

$$w_{new}^{(j)} = \frac{\tilde{w}^{(j)}}{\sum_{k=1}^{N_p} \tilde{w}^{(k)}} \tag{40}$$

The likelihood function $L(E_{t+1} \mid V^{(j)})$ is constructed from the LLM-extracted signal strength and direction. For a positive earnings surprise, for example, the likelihood is higher for particles with higher valuations (corresponding to stronger underlying business performance).

Alert Generation and Opportunity Flagging

The monitoring layer generates two types of alerts. Opportunity alerts are triggered when the probability of undervaluation PU_i crosses a high threshold (default 0.85) and the composite score Ω_i is in the top quartile of the coverage universe. Deterioration alerts are triggered when PU_i falls below a low threshold (default 0.30) for a company currently in the portfolio, or when the composite score Ω_i falls by more than one standard deviation from its trailing twelve-month average.

Tracking the Health of Portfolio Companies

For companies already held in the VRC portfolio, AEGIS maintains an enhanced monitoring protocol. Beyond the standard updates described above, this protocol includes:

- Monthly competitive health reports comparing the company's positioning to its three closest competitors.

- Quarterly LLM-generated narrative assessments of whether the original investment thesis remains intact.
- Real-time tracking of the specific key performance indicators identified during the initial investment analysis as thesis-validating metrics.
- A thesis integrity score that declines when evidence accumulates against the original investment thesis, regardless of price performance.

Portfolio Integration and Position Sizing

From Opportunity Identification to Position Sizing

AEGIS produces a ranked list of investment opportunities ranked by composite score and probability of undervaluation. The translation from this list to actual portfolio positions requires an additional set of decisions about position sizing, portfolio concentration, and correlation management. VRC's position sizing framework, implemented within AEGIS, draws on Kelly and its extensions while imposing practical constraints appropriate for an institutional investor [21].

Kelly-Adjusted Position Sizing

The Kelly criterion prescribes the position size that maximizes the expected logarithmic growth of capital [21]:

$$f^* = \frac{p \cdot b - q}{b} = \frac{\mathbb{E}[r]}{b \cdot \mathrm{Var}[r]} \tag{41}$$

where p is the probability of the favorable outcome, q = 1 − p is the probability of the unfavorable outcome, and b is the gain-to-loss ratio. In the context of equity investing, the Kelly fraction is:

$$f_i^* = \frac{\ddot{r}_i}{\sigma_i^2} \tag{42}$$

where $\hat{r}_i$ is the expected excess return above the risk-free rate and σ^2 is the variance of returns. VRC applies a fractional Kelly criterion, using $f_i = \frac{1}{2} f_i^*$ to account for estimation error in the inputs and to maintain a risk-managed portfolio.

Correlation Adjustment

The Kelly criterion in equation (42) ignores the correlations between portfolio positions. AEGIS extends this to a multi-asset portfolio through the mean-variance framework of Markowitz (1952), imposing a diversification constraint:

$$\max_{\mathbf{w}} \mathbf{w}^T \hat{\mathbf{r}} - \frac{\gamma}{2} \mathbf{w}^T \hat{\boldsymbol{\Sigma}} \mathbf{w} \tag{43}$$

subject to position limits and sector concentration constraints, where $\hat{\mathbf{r}}$ is the vector of expected excess returns derived from the AEGIS valuation model and $\hat{\boldsymbol{\Sigma}}$ is the estimated covariance matrix of returns. The covariance matrix is estimated using a shrinkage estimator (Ledoit and Wolf, 2004) to improve out-of-sample stability.

Uncertainty Penalty

AEGIS imposes an explicit uncertainty penalty on position sizes that reflects the width of the intrinsic value distribution. Companies with wide valuation distributions (high $\mathrm{IQR}(V_0)$) receive smaller allocations than companies with narrow distributions, even if their central value estimates imply comparable degrees of undervaluation:

$$f_i^{adj} = f_i \cdot \exp\left(-\lambda_{unc} \cdot \frac{IQR(V_0^{(i)})}{\hat{V}_{(i)}^{med}}\right) \tag{44}$$

where $\lambda_{unc} > 0$ is the uncertainty aversion parameter, set by the VRC investment committee.

Case Illustrations

Indian Banking Sector Application

To illustrate AEGIS in operation, this section applies the framework to the Indian banking sector, which has been the subject of prior VRC research and which provides an instructive laboratory for the system's capabilities.

The Indian banking sector in FY2025 presents companies at very different stages of their credit cycle and institutional development, making it an ideal stress test for AEGIS's ability to discriminate between superficially similar investment situations.

AEGIS Application to Bank of India

Data Ingestion

AEGIS ingests Bank of India's annual reports for FY2021 through FY2025, all quarterly earnings call transcripts, RBI regulatory filings, and the bank's Basel III Pillar 3 disclosures. The total corpus for Bank of India comprises approximately 2,800 pages of financial and narrative content.

Financial Quality Analysis

The financial quality score for Bank of India reflects the genuine improvement in asset quality since 2018 (GNPA declining from approximately 16% to 7.5%) while accounting for the structural constraints of public sector ownership. The normalized ROIC of approximately 8.5% on an adjusted basis sits below the estimated cost of equity of 12-13% for a PSU bank, which is directly reflected in the P/B ratio below 1.0x.

AEGIS Composite Score

Table 3: AEGIS Composite Score: Bank of India, FY2025

Dimension	Score (0-100)	Weight
Financial Quality (Q_F)	52	0.25
Competitive Position (Q_C)	43	0.25
Management Quality (Q_M)	48	0.20
Growth Opportunity (Q_G)	61	0.15
Valuation Attractiveness (Q_V)	71	0.15
Composite Ω	54.3	1.00

The Monte Carlo engine produces an intrinsic value range of Rs. 68 to Rs. 140 per share under the base scenario parameterization, with the probability-weighted mean at approximately Rs. 96. Against a market price of approximately Rs. 97 at the time of analysis, the probability of undervaluation is 0.49, suggesting fair value. An improvement in operational efficiency or a reduction in the PSU ownership discount would be required to produce a compelling buy signal.

AEGIS Application to Yes Bank

Yes, Bank presents a fundamentally different analytical challenge: a bank undergoing reconstruction after a governance failure requires explicitly probabilistic treatment of the binary outcome risk.

Scenario Decomposition

AEGIS applies its scenario framework to Yes Bank with the following parameterization, consistent with the analysis in prior VRC research:

Table 4: AEGIS Scenario Analysis: Yes Bank, FY2025

Scenario	π_k	5-yr Revenue CAGR	Terminal P/B	Value/Share
Continued Recovery	0.50	18%	1.5x	Rs. 35
Full Normalisation	0.25	25%	2.2x	Rs. 55
Stagnation	0.15	8%	0.8x	Rs. 15
Re-restructuring	0.10	Negative	0.3x	Rs. 5
Probability-Weighted Value	1.00			Rs. 32

The AEGIS system flags the tension between the probability-weighted intrinsic value of approximately Rs. 32 and the market price of approximately Rs. 18 at the time of analysis, producing a probability of undervaluation of 0.67. However, the width of the valuation distribution (IQR of approximately Rs. 22) results in a meaningfully reduced position size recommendation through the uncertainty penalty in equation (44).

The composite AEGIS score for Yes Bank is 61.4, reflecting strong growth opportunity scores but penalized management quality scores due to the historical governance failures and ongoing uncertainty about management credibility.

Cross-Sectional Discrimination

The contrast between Bank of India (AEGIS score 54.3, probability of undervaluation 0.49) and Yes Bank (AEGIS score 61.4, probability of undervaluation 0.67) illustrates the system's ability to make nuanced discriminations within a sector. Yes, Bank is the preferred opportunity on both valuation and composite quality grounds, but the substantially wider valuation distribution demands a smaller position size. This is precisely the analytical conclusion that a careful human analyst would reach, but AEGIS reaches it systematically and at scale.

Extensions and Future Development

Deep Learning for Alternative Data

Future versions of AEGIS will incorporate deep learning models trained directly on alternative data streams. Convolutional neural networks applied to satellite imagery can estimate retail footfall, warehouse utilization, and agricultural production with considerable accuracy. Graph neural networks applied to supply chain relationship data can identify systemic vulnerabilities before they appear in financial statements.

Options Market Intelligence

The options market embeds forward-looking information about the probability distribution of stock returns that is directly relevant to the valuation process. AEGIS's derivatives integration layer, currently in development, will extract risk- neutral probability distributions from option chains and incorporate them as data-driven inputs into the Monte Carlo simulation:

$$p^{RN}(S_T \leq K) = e^{rT}\frac{\partial C}{\partial K} \tag{45}$$

Where C is the call option price at strike K and expiry T . The risk-neutral density, extracted from the option chain through Breeden-Litzenberger, provides a market-implied view of the forward distribution of the stock price that can be compared to the AEGIS intrinsic value distribution to identify divergences [22].

Cross-Market and Cross-Asset Integration

Global equity investing requires understanding the relationship between equity valuations and other asset markets. Future AEGIS development will integrate credit market signals (CDS spreads, high-yield indices), currency forward curves, and commodity futures into the macro regime layer, producing richer and more forward-looking regime classifications.

Dynamic Weight Optimization

The current composite scoring weights in Table 2 are set by the investment committee based on investment philosophy. A future extension will use reinforcement learning to optimize these weights dynamically, updating them based on the realized predictive power of each dimension across different market regimes:

$$\mathbf{w}^* = \arg\max_{\mathbf{w}} \mathbb{E}\left[\sum_t \delta^t r_t(\mathbf{w})\right] \tag{46}$$

where $r_t(\mathbf{w})$ is the information coefficient of the composite score with weights $\mathbf{w}$ at time t, and δ is a discount factor reflecting the decaying relevance of distant historical data.

Network Effects in the Coverage Universe

As the AEGIS coverage universe expands, the system will increasingly be able to exploit cross-company information. The LLM pipeline can identify supply chain relationships, customer-supplier dependencies, and competitive interactions between companies in the coverage universe. These relationships form a directed graph structure that can be analyzed using graph neural networks to produce joint valuations that account for the interdependencies between companies.

Limitations and Risk Disclosures

AEGIS is a powerful analytical tool, but it is subject to important limitations that the VRC investment team must keep clearly in mind when interpreting its outputs.

Model Risk

All quantitative models are simplifications of a reality that is vastly more complex. The stochastic DCF engine in AEGIS relies on mean-reverting processes for growth rates and margins; in practice, structural breaks can produce persistent regime changes that violate the mean-reversion assumption. The management quality model relies on LLM extraction of signals from text; LLMs can miss subtle signals that an experienced human analyst would detect, and can occasionally generate plausible-sounding but incorrect assessments.

Data Quality and Availability

The quality of AEGIS's outputs is bounded by the quality of its inputs. For companies listed in emerging markets with less rigorous accounting standards or weaker disclosure requirements, the financial data feeding the model may be less reliable. AEGIS applies data quality flags to indicate when inputs fall below reliability thresholds, but these flags are themselves imperfect.

LLM Limitations

Large language models, despite their impressive capabilities, can produce incorrect statements with apparent confidence. The AEGIS pipeline mitigates this through the retrieval-augmented generation architecture and source attribution requirements, but the risk of LLM error cannot be eliminated entirely. All LLM-derived signals should be treated as inputs to human judgment, not as replacements for it.

Survivorship Bias

The coverage universe of AEGIS is necessarily limited to companies that exist and are publicly traded at the time of analysis. Historical backtesting of AEGIS signals may therefore exhibit survivorship bias if the training data does not adequately account for companies that were delisted, acquired, or went bankrupt.

Behavioral and Execution Risks

Even if AEGIS correctly identifies undervalued companies, the realization of value requires the market to eventually recognize the discrepancy. Shleifer identifies limits to arbitrage that can prevent the correction of genuine mispricings for extended periods [23]. VRC investors must have sufficient patience and financial strength to maintain positions through periods of adverse mark-to-market performance.

Conclusion

This paper has presented AEGIS, the Autonomous Equity Grading and Intelligence System developed at Verma Research Capital. AEGIS represents a comprehensive answer to one of the most important practical challenges in fundamental investing: how to apply rigorous, systematic, and continuously updated valuation analysis to a large universe of companies across global markets.

The system integrates stochastic discounted cash flow mechanics, Bayesian belief updating, large language model processing of annual reports and earnings transcripts, multi-factor composite scoring across financial quality, competitive position, management quality, growth opportunity, and valuation attractiveness dimensions, macroeconomic regime detection, and a continuous monitoring and alerting infrastructure into a unified platform.

The key contributions of this paper are as follows. First, the formalization of probabilistic intrinsic value as a distribution rather than a point estimate, and the derivation of the probability of undervaluation as an operational investment signal. Second, the design of a retrieval-augmented LLM pipeline that extracts structured analytical signals from unstructured financial text at scale, with source attribution and calibrated confidence scores. Third, the multi-factor composite scoring model that operationalizes VRC's qualitative investment criteria in a quantitatively tractable form. Fourth, the particle filter implementation of Bayesian belief updating that allows intrinsic value distributions to be updated efficiently upon the arrival of new information without full re-simulation. Fifth, the Kelly-adjusted position sizing framework with uncertainty penalties that translates valuation signals into portfolio weights in a risk-managed way.

The case illustrations from the Indian banking sector demonstrate the system's ability to make nuanced analytical discriminations within a sector, replicating and extending the kind of analysis that VRC conducts manually in its published research. Bank of India and Yes Bank, while superficially similar in their status as recovering banks, receive structurally different analytical treatments under AEGIS that correctly reflect the different nature of the uncertainty surrounding their future performance.

AEGIS is not a replacement for human judgment in the investment process. It is an amplifier of human judgment: a system that scales the analytical process, ensures its consistency, and surfaces the situations where human attention is most warranted. The ultimate measure of its success will be the quality of the investment decisions it helps VRC make over the coming years, as it continues to monitor, update, and refine its assessments of the global equity opportunity set.

Valuation is not a formula. It is a way of thinking. AEGIS is the architecture that makes that way of thinking operational at scale [24-41].

A. Mathematical Derivations

A.1 Mean-Reverting Growth Rate and Long-Run Distribution

The mean-reverting growth process in equation (5) can be analyzed to derive its stationary distribution. The process is an Ornstein-Uhlenbeck process in discrete time. Its variance at horizon t is:

$$\mathrm{Var}(g_t) = \frac{\sigma_g^2}{1 - e^{-2\kappa}} \left(1 - e^{-2\kappa t}\right) \tag{47}$$

As $t \to \infty$, the variance converges to $\sigma^2/(1 - e^{-2\kappa})$, confirming the process is stationary. The autocorrelation at lag h_g is $\mathrm{Corr}(g_t, g_{t+h}) = e^{-\kappa h}$, confirming that persistence decays at rate κ.

A.2 Terminal Value Sensitivity Analysis

The sensitivity of the intrinsic value to the terminal growth rate g_∞ can be derived analytically. From equations (14) and (1):

$$\frac{\partial V_0}{\partial g_\infty} = \frac{FCFF_T}{(1 + WACC)^T} \cdot \frac{WACC}{(WACC - g_\infty)^2} \tag{48}$$

This expression shows that the sensitivity to terminal growth is proportional to the terminal-year free cash flow discounted to today, scaled by $WACC/(WACC - g_\infty)^2$. For a typical company with WACC = 10% and g_∞ = 3%, this scaling factor is approximately $10/(7\%)^2$ = 204, meaning that a 1 percentage point change in terminal growth changes the estimated intrinsic value by approximately 204 times the discounted terminal-year free cash flow.

A.3 Particle Filter Convergence

Under standard regularity conditions (Del Moral, 2004), the particle filter approximation of the posterior converges in distribution to the true posterior as the number of particles $N_p \to \infty$. The convergence rate is $O(N_p^{-1/2})$ in the total variation distance, meaning that increasing the particle count from 10,000 to 40,000 reduces the approximation error by half.

B. AEGIS Prompt Library

This appendix provides representative examples of the structured prompts used in the AEGIS LLM pipeline. These prompts are designed to elicit structured, evidence-grounded analytical outputs from the language model.

B.1 Revenue Driver Extraction Prompt

You are an expert financial analyst. The following passages are excerpted from the annual report of [Company Name] for fiscal year [Year]. Your task is to extract the primary revenue drivers reported by management. For each driver, provide: (1) the name of the driver, (2) its approximate quantitative contribution to revenue growth (if stated), (3) whether it is organic or inorganic, (4) whether it is sustainable or one-time, and (5) the verbatim passage from the text that supports your answer. Return your response as a structured JSON object.

B.2 Management Tone Assessment Prompt

You are an expert financial analyst specializing in the assessment of management communication quality. The following passages are excerpted from the earnings call transcript of [Company Name] for [Quarter, Year]. Analyze the language used by management (not analysts) and assess: (1) the frequency of hedging language (words such as approximately, roughly, we expect, subject to, and so forth), (2) whether management takes personal accountability for underperformance or attributes it to external factors, (3) whether forward guidance is specific and measurable or vague, and (4) whether the tone is more cautious or more confident relative to typical earnings calls. Provide a score from 1 to 10 for management communication quality, where 10 represents exceptional clarity and accountability, and cite supporting passages.

C. AEGIS Score Summary Table

Table 5: Illustrative AEGIS Coverage Universe Summary

Company	Sector	Ω Score	PU	Recommendation
Company A	Technology	81.2	0.91	High Prior-
				ity Buy
Company B	Healthcare	76.4	0.87	Buy
Company C	Financials	71.8	0.79	Buy
Company D	Consumer	68.3	0.72	Watch
Company E	Industrials	64.1	0.68	Watch
Company F	Technology	58.9	0.55	Neutral
Company G	Energy	52.4	0.44	Neutral
Company H	Utilities	47.2	0.35	Underweight

Company I	Financials	41.8	0.28	Underweight
Company J	Consumer	35.6	0.19	Avoid

D. Glossary of Key Terms

AEGIS Autonomous Equity Grading and Intelligence System. The proprietary valuation platform of Verma Research Capital.

Bayesian updating The process of revising a prior probability distribution in light of new evidence using Bayes' theorem.

Composite score (Ω) The weighted aggregate of the five AEGIS sub-scores, reflecting the overall quality and opportunity assessment of a company.

Moat score (Q_C) The quantitative assessment of a company's competitive advantages across five source categories.

Particle filter A sequential Monte Carlo method for approxi- mating Bayesian posterior distributions in real time.

Probability of undervaluation (PU) The fraction of Monte Carlo simulation paths in which the estimated intrinsic value exceeds the current market price.

RAG Retrieval-Augmented Generation. An LLM architecture that grounds model outputs in retrieved source documents.

Stochastic DCF A discounted cash flow model in which inputs are modeled as random variables, producing a distribution over intrinsic values rather than a single point estimate.

VRC Verma Research Capital LLC. A Wisconsin- registered investment management firm founded by Divyanshu Verma in 2025.

WACC Weighted Average Cost of Capital. The blended required rate of return across equity and debt financing, used as the discount rate in DCF analysis.

References

[1] Williams, J. B. (1938). The theory of investment value. Harvard University Press.

[2] Modigliani, F., and Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. American Economic Review, 48(3), 261–297.

[3] Gordon, M. J. (1956). Dividends, earnings, and stock prices. Review of Eco- nomics and Statistics, 41(2), 99–105.

[4] Damodaran, A. (2012). Investment valuation: Tools and techniques for deter- mining the value of any asset (3rd ed.). Wiley.

[5] Jacobs, B. I., and Levy, K. N. (2000). Equity management: Quantitative analysis for stock selection. McGraw-Hill.

[6] Loughran, T., and McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. Journal of Finance, 66(1), 35–65.

[7] Buehlmaier, M. M. M., and Whited, T. M. (2018). Are financial constraints priced? Evidence from textual analysis. Review of Financial Studies, 31(7), 2693–2728.

[8] Fama, E. F., and French, K. R. (1992). The cross-section of expected stock returns. Journal of Finance, 47(2), 427–465.

[9] Fama, E. F., and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1), 3–56.

[10] Carhart, M. M. (1997). On persistence in mutual fund performance. Journal of Finance, 52(1), 57–82.

[11] Hou, K., Xue, C., and Zhang, L. (2015). Digesting anomalies: An investment approach. Review of Financial Studies, 28(3), 650–705.

[12] Asness, C. S., Moskowitz, T. J., and Pedersen, L. H. (2013). Value and

momentum everywhere. Journal of Finance, 68(3), 929–985.
[13] Verma, D. (2026). Valuation as mindset: Subjectivity, dynamism, and probabilistic thinking in equity valuation. Verma Research Capital Institutional Whitepaper.
[14] Penman, S. H. (2013). Financial statement analysis and security valuation (5th ed.). McGraw-Hill.
[15] Dechow, P., Ge, W., and Schrand, C. (2010). Understanding earnings quality: A review of the proxies, their determinants and their consequences. Journal of Accounting and Economics, 50(2-3), 344–401.
[16] Reimers, N., and Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing. Association for Computa- tional Linguistics.
[17] Buffett, W. E. (2009). Berkshire Hathaway shareholder letters, 1965–2009. Berkshire Hathaway Inc.
[18] Jensen, M. C., and Meckling, W. H. (1976). Theory of the firm: Manage- rial behavior, agency costs and ownership structure. Journal of Financial Economics, 3(4), 305–360.
[19] Porter, M. E. (1980). Competitive strategy: Techniques for analyzing industries and competitors. Free Press.
[20] Greenwald, B., Kahn, J., Sonkin, P., and van Biema, M. (2001). Value investing: From Graham to Buffett and beyond. Wiley.
[21] Kelly, J. L. (1956). A new interpretation of information rate. Bell System Technical Journal, 35(4), 917–926.
[22] Breeden, D. T., and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. Journal of Business, 51(4), 621–651.
[23] Shleifer, A. (2000). Inefficient markets: An introduction to behavioral finance. Oxford University Press.
[24] Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Annals of Mathematical Statistics, 41(1), 164–171.
[25] Damodaran, A. (2003). Measuring company exposure to country risk: Theory and practice. Journal of Applied Finance, 13(2), 63–75.
[26] DeBondt, W. F. M., and Thaler, R. (1985). Does the stock market overreact? Journal of Finance, 40(3), 793–805.
[27] Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. Journal of Finance, 25(2), 383–417.
[28] Graham, B. (1949). The intelligent investor. Harper and Brothers.
[29] Graham, B., and Dodd, D. L. (1934). Security analysis. McGraw-Hill.
[30] Grossman, S. J., and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. American Economic Review, 70(3), 393–408.
[31] Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.
[32] Koller, T., Goedhart, M., and Wessels, D. (2020). Valuation: Measuring and managing the value of companies (7th ed.). Wiley.
[33] Lakonishok, J., Shleifer, A., and Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. Journal of Finance, 49(5), 1541–1578.
[34] Ledoit, O., and Wolf, M. (2004). A well-conditioned estimator for large- dimensional covariance matrices. Journal of Multivariate Analysis, 88(2), 365–411.
[35] Markowitz, H. (1952). Portfolio selection. Journal of Finance, 7(1), 77–91.
[36] Mauboussin, M. J., and Callahan, D. (2014). Measuring the moat: Assess- ing the magnitude and sustainability of value creation. Credit Suisse Global Financial Strategies.

[37] Montier, J. (2010). The little book of behavioral investing. Wiley.

[38] Reserve Bank of India (2024). Report on trend and progress of banking in India 2023–24. RBI Publications.

[39] Shefrin, H. (2007). Behavioral corporate finance. McGraw-Hill.

[40] Shiller, R. J. (2015). Irrational exuberance (3rd ed.). Princeton University Press.

[41] Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? Accounting Review, 71(3), 289–315.