SUMATOSOFT

Enterprise Software Development Vendor Evaluation Methodology

1. The vendor selection process

The vendor selection process consists of six high-level steps. Detailed explanations are provided in the following sections.

1.1 Build a candidate pool from proof-heavy sources.
We start by identifying companies based on publicly available enterprise-grade evidence. See section 2.

1.2 Apply enterprise-grade inclusion criteria.
Only vendors that meet the minimum enterprise requirements proceed to scoring. See section 3.

1.3 Remove vendors with weak or biased signals.
Vendors that cannot be assessed reliably are excluded from further evaluation. See section 4.

1.4 Score vendors using an enterprise-weighted rubric (100 points).
Qualified vendors are evaluated using a standardized enterprise scoring model. See section 5.

1.5 Verify claims using a “show your work” approach.
All claims are verified against public, consistent evidence before affecting the score. See section 6.

1.6 Normalize results for comparison.
Evaluation inputs and outputs are standardized to enable fair comparison across vendors. See section 7.

2. Where to source candidates

2.1 Review platforms and vendor directories

We use review platforms and directories as an initial discovery layer.

Primary sources include:

- Clutch and G2 (service pages and reviews where available).
- DesignRush and GoodFirms (used for discovery, then verified via primary materials).

From these sources, we extract:

- Review volume and recency.
- Repeated strengths and risks (delivery, communication, quality).
- Typical project size, industries served, and engagement models.

2.2 Company primary materials (highest weight)

Company-owned materials are the primary source for understanding what a vendor actually delivers.

We review:

- Service landing pages and capability descriptions.
- “How we work,” delivery playbooks, and governance descriptions.
- Security, privacy, and compliance pages.
- Case studies with defined scope, constraints, and outcomes.

From these materials, we extract:

- Core offering and areas of specialization.
- Delivery model (locations, overlap, team structure).
- Evidence quality (level of detail, metrics, stakeholder complexity).

2.3 Public proof of enterprise work

We use independent public proof to confirm enterprise-scale relevance.

Sources include:

- Named Client lists (when publicly disclosed).
- Detailed case studies and references in press releases.
- Partner and ecosystem directories, such as AWS Partner Network, Microsoft Solutions Partner, Google Cloud Partner.

From this proof, we extract:

- Enterprise brand credibility.
- Types of programs delivered (platforms, modernization, managed run + change).
- Signals of regulated or high-compliance delivery context.

2.4 Certifications and compliance registries

Where applicable, we use certification and compliance registries to validate security and compliance claims.

Sources include:

- ISO certificate registries (certificate-level validation).
- SOC 2 report availability statements with scope and auditor disclosure.
- Industry-specific compliance claims tied to delivery context (e.g., healthcare, payments).

From these sources, we extract:

- Certification presence and scope.
- Maturity indicators such as documented policies, secure SDLC, and audit readiness.

2.5 Engineering footprint and hiring signals

Engineering footprint and hiring data are used as supporting signals of depth and scale.

We review:

- Open-source presence on GitHub (organization-level activity, published tooling).
- Engineering blogs and architecture write-ups.
- Job descriptions indicating platform, cloud, security, SRE, and data roles.

From these signals, we extract:

- Breadth of real-world engineering practices.
- Investment in platform engineering, DevOps, QA automation, and SRE.

3. Inclusion criteria (enterprise-grade)

A company is included in the evaluation only if it meets all criteria below.
These rules ensure that evaluated vendors are enterprise-relevant and comparable.

3.1 Enterprise relevance (scope and complexity)

We include vendors that demonstrate delivery experience involving at least one of the following:

- Core business platforms or mission-critical systems.
- Complex integrations (legacy systems, data platforms, third-party ecosystems).
- Multi-team or multi-stream delivery at the program level.
- Long-term support and system evolution (run + change).

Signals we look for:

- Platform ownership rather than isolated feature delivery.
- Evidence of stakeholder, dependency, and risk complexity is typical for enterprises.

3.2 Evidence quality (proof over claims)

We include vendors only when public proof is sufficient to assess them objectively.

Required evidence includes:

- Case studies with defined scope, technical context, and outcomes.
- Named industries and realistic delivery constraints.
- Traceable references such as Client quotes, press releases, or partner pages, where available.

We do not rely on:

- Pure marketing statements.
- Unverifiable “enterprise experience” claims.

3.3 Delivery maturity (predictability and governance)

We include vendors that describe delivery as a repeatable system.

We look for:

- Documented delivery methodology and governance structure.
- Project controls such as reporting cadence, risk management, and escalation paths.
- QA and engineering standards, including testing discipline and release practices.

Signals extracted:

- Consistency between “how we work” descriptions and case studies.
- Evidence of controlled delivery beyond individual teams.

3.4 Security baseline (minimum enterprise bar)

We include vendors that demonstrate a credible baseline security posture.
We look for:

- Security and privacy documentation describing SDLC practices.
- Evidence of audit readiness through certifications or equivalent process proof.
- Ability to work within restricted environments and enterprise security reviews.

3.5 Architectural and technical depth

We include vendors that show real technical capability beyond generic “full-stack” claims.

We look for evidence of:

- Architecture and platform engineering competency.
- Cloud and data engineering depth relevant to enterprise systems.
- Design for scalability, resilience, and long-term maintainability.

Signals are taken from:

- Case studies, architecture write-ups, and technical blogs.
- Descriptions of senior technical roles involved in delivery.

3.6 Scale and delivery model transparency

We include vendors that can clearly explain how delivery works at scale.

We look for clarity on:

- Team locations and on/near/offshore mix, including overlap hours.
- Team ramp-up approach and continuity mechanisms.
- Engagement models such as project delivery, dedicated teams, or managed programs.
- Opaque or vague delivery models are not considered enterprise-ready.

4. Exclusion rules (enterprise-grade)

We remove companies from the evaluation when any of the conditions below apply. These rules protect the integrity and comparability of results.

4.1 Pay-to-play visibility

We exclude vendors when evaluation visibility depends on financial placement.

Signals include:

- Sponsored positions or paid ranking influence.
- Lead-generation listings are presented as an independent evaluation.
- Disproportionate exposure unsupported by proof-heavy materials.

4.2 Thin or unverifiable proof

We exclude vendors when public evidence does not support an objective assessment.

Signals include:

- Case studies without scope, constraints, technical context, or outcomes.
- Absence of Client identification across all materials.
- Broad capability statements without supporting detail.

4.3 Unclear delivery ownership

We exclude vendors when responsibility for delivery cannot be determined.

Signals include:

- Opaque subcontracting or partner dependency.
- Vague delivery descriptions covering locations, roles, or governance.
- Network-style presentation with inconsistent execution ownership.

4.4 Enterprise risk signals without mitigation evidence

We exclude vendors when recurring risk patterns appear without counter-signals.

Signals include:

- Repeated review themes around missed deadlines, weak QA, or high churn.
- Lack of visible security posture documentation or SDLC practices.
- Absence of operational ownership for support, incidents, or production systems.

4.5 Misaligned company type for this list

We exclude companies whose primary operating model does not align with enterprise delivery.

Excluded categories include:

- Freelancer marketplaces or talent brokers without delivery governance.
- Design or marketing studios with limited enterprise engineering depth.
- Narrow boutiques unable to scale to program-level enterprise work.

5. Scoring rubric

We evaluate qualified vendors using a 100-point enterprise-weighted scoring model. Each scoring area reflects a decision dimension commonly used in enterprise vendor selection.
Scores are assigned based on verifiable public evidence and consistency across sources.

5.1 Scoring structure and weights

The total score consists of the following weighted areas:

- Security and compliance readiness - 25 points.
- Delivery predictability and governance - 20 points.
- Technical depth for target scope - 20 points.
- Relevant enterprise cases - 15 points.
- Scale and ramp capability - 10 points.
- Commercial fit and transparency - 10 points.
- Weights reflect typical enterprise risk priorities.

5.2 Scoring approach

For each area, we apply the same evaluation logic:
Identify evidence signals defined for the category.

- Check signal consistency across independent sources.
- Assign a score within the allowed range based on signal strength.
- Scores represent evidence quality and maturity signals, not guarantees of future performance.

5.3 Category-level scoring logic

Each scoring category follows a structured assessment:

- Strong, consistent, and specific signals result in higher scores.
- Partial or generic signals result in mid-range scores.
- Weak or marketing-level signals result in lower scores.
- Scoring is conservative when evidence is limited.

5.4 Deduction principles

Deductions are applied when:

- Claims lack supporting evidence.
- Signals conflict across sources.
- Key enterprise indicators are missing or unclear.
- Deductions reflect risk exposure relevant to enterprise buyers.

5.5 Interpretation of scores

Scores enable relative comparison across vendors.

They support:

- Shortlisting.
- Identification of strengths and risk areas.
- Structured discussion during the RFP and interview stages.
- Scores are directional indicators designed to complement deeper evaluation.

6. How we verified claims (certifications, case studies, references)

We treat marketing statements as hypotheses.
A claim affects the score only after validation through credible, consistent evidence.

6.1 Certifications and security assertions

We verify security and compliance claims using explicit, scope-aware signals.

We check:

- Certification statements with defined scope, entity, and locations.
- Certificate identifiers, issuing bodies, and validity periods are published.
- Public registries and partner compliance pages, where applicable.

Signals extracted:

- Certification presence and coverage.
- Indicators of audit readiness and documented security programs.
- Claims without scope detail are scored conservatively.

6.2 Case studies (depth test)

We evaluate each case study based on evidence depth.

We assess:

- Context: industry, scale, constraints, stakeholder complexity.
- Scope: ownership boundaries and delivery responsibility.
- Technology and architecture: stack, deployment model, integrations, data flows.
- Outcomes: measurable results and operational impact where available.
- Case studies lacking delivery detail do not strengthen scores.

6.3 Client references and public proof

We use independent public proof to validate enterprise relevance.

Stronger signals include:

- Named Clients are linked to specific work.
- Long-term engagement indicators, such as multi-phase programs.
- Consistent Client statements across sources.

Weaker signals include:

- Logo lists without a delivery context.
- Unspecified enterprise claims without traceable proof.

6.4 Consistency cross-check across sources

We cross-check claims across multiple materials.

We compare:

- Services pages, delivery methodology descriptions, and case studies.
- Review patterns highlighting recurring strengths or risks.

- Hiring signals indicating real operational practices.

When discrepancies appear, priority is given to:

- Primary materials with concrete detail.
- Independent sources confirming the same signals.

6.5 Evidence-strength downgrades

When important claims cannot be validated:

- The claim is marked as unverified.
- The claim does not influence differentiation.
- The score reflects only confirmed evidence.

This approach preserves comparability across vendors.

7. Normalize results for fair comparison

We normalize evaluation inputs and outputs to ensure consistent, comparable results across vendors.

7.1 Service taxonomy normalization

We map vendor offerings to a consistent enterprise service taxonomy.

Standard categories include:

- Enterprise modernization programs.
- Data and AI platforms.
- Digital product engineering.
- Managed engineering programs (run + change).

This mapping ensures services are compared at the same abstraction level.

7.2 Delivery model normalization

We standardize how delivery models are described and compared.

We normalize:

- Headquarters and delivery locations.
- Onshore, nearshore, and offshore mix.
- Overlap hours and collaboration model.

This allows delivery structures to be assessed consistently across vendors.

7.3 Evidence strength normalization

We align evidence signals to common strength levels.

We normalize:

- Case study depth and specificity.
- Presence of named Clients and measurable outcomes.
- Technical and architectural detail.

Evidence strength is evaluated relative to the same criteria for all vendors.

7.4 Risk flag normalization

We standardize how risk indicators are recorded and interpreted.

Normalized risk flags include:

- Key-person dependency.
- Unclear security posture.
- Vague support or production ownership model.

Risk flags inform score interpretation and shortlisting decisions.

This normalization step ensures that final scores support fair enterprise comparisons and practical shortlisting.

SUMATOSOFT

Thank you for your time!

Any questions? Drop us a line!

Headquarters

One Boston Place, Suite 2602
Boston, MA 02108, United States

Other ways to get in touch

info@sumatosoft.com
sumatosoft.com