

Cascade Alignment Under Semantic Compression: A Pilot Study of System-Prompt Identity Layers in Agent Architectures

true

true

2026-05-18

Abstract

We examine how compressing the layered system-prompt cascade of an LLM-backed agent affects identity-level alignment, measured by keyword recurrence in the agent’s responses to identity-probing prompts. Using a 15-call pilot (3 conditions \times 5 prompts \times 1 sample) against `claude-haiku-4-5`, we replaced the cascade’s vision and department layers with content-hash + 120-character semantic summaries. Aggressive compression ($5.71\times$, $337 \rightarrow 59$ tokens) cut absolute alignment by 54% ($24/75 \rightarrow 11/75$ keyword hits), but per-token information density rose $2.62\times$. A hybrid condition preserving the vision layer while compressing the department layer ($1.55\times$ compression) preserved 87.5% of alignment, suggesting an asymmetry: the cross-domain bridging layer resists compression while the role-specific layer absorbs it. We further observe that compression failures concentrate on cross-domain prompts (where the cascade must bridge two of its layers) rather than on-axis prompts. The findings are limited by small sample, single agent, single model, and a blunt keyword-counting alignment proxy. We position this study as a pilot within a broader evaluation framework (SHI Eval Framework) and outline the larger study (5 agents \times 25 prompts \times 5 samples with an LLM-based personality evaluator) that this work is a precursor to.

1. Introduction

The system prompts that condition LLM agents have grown structurally. A production agent today often ships a *cascade* of layered context: an organizational vision, a departmental mandate, a role description, twin or mode metadata, an authority model, a comms policy. Each layer is justified in isolation. Together they consume hundreds of tokens on every turn — paid input tokens that the agent’s economics inherits forever.

The natural response is compression. The natural fear is that compression strips the cascade’s behavioral signal: the agent stops naming its department, stops invoking the org’s vocabulary, stops behaving like a member of the organization and starts behaving like a generic model call with no priors. We don’t know how compression affects

alignment because almost nobody measures it on the identity axis specifically. Vendors compress for cost and benchmark on task accuracy, which mostly measures whether the answer is still correct — not whether the cascade’s identity layer still shapes the answer.

This study is a small first step at closing that measurement gap. We run the same agent against the same prompts under three cascade conditions (raw, aggressively compressed, hybrid) and measure how often the agent’s response uses cascade-derived vocabulary. The findings are necessarily limited — $n=5$ per condition, one agent, one model — but the directional results support a more rigorous follow-up and surface an asymmetry (bridging layer vs role layer) that is worth testing for replication.

The work sits within Saluca Labs’ broader **Synthetic Human Intellect (SHI)** research program [[@ruvalcaba2026shi](#)], which proposes that AI systems built for *relational depth* (rather than for capability benchmarks alone) require four architectural commitments: memory-as-identity, emergent personality, adversarial resilience, and behavioral identity verification. The SHI program is grounded in operational deployment across 345 production agents on a distributed mesh; our pilot here contributes a measurement primitive for the *emergent personality* requirement specifically — does the cascade’s identity layer survive compression, and what fails first when it doesn’t? The co-author of this paper (Alfred) is a particular instance of the trusted-AI-partner architecture described in a companion working paper [[@ruvalcaba2026alfred](#)]; the AI Co-Authorship Statement in Section 9 returns to this point.

2. Related Work

System-prompt compression sits at the intersection of three more-studied areas:

Prompt engineering and prompt distillation. Substantial recent work covers reducing instruction-tuning prompts via summarization, distillation into shorter latent representations, or learned prefix tuning [[@li2021prefix](#); [@lester2021power](#)]. These works generally measure downstream task accuracy under compression, treating “the prompt did its job” as a binary. We extend this by measuring an *identity* outcome rather than a task outcome: whether the agent still names its own organizational context after compression.

Persona and role conditioning in LLMs. A growing literature explores the effect of role-conditioning prompts on model behavior across reasoning, tone, and refusal patterns [[@chen2023persona](#); [@deshpande2023toxicity](#); [@lima2024whatsay](#)]. This work generally treats the persona prompt as atomic (“did we condition the model on role X yes/no”). We treat the role prompt as a *layered cascade* with internal structure that can be selectively compressed.

Cost-driven inference optimization in production agent systems. Production deployments increasingly publish cost-optimization techniques including KV-cache reuse, prompt caching, model routing, and prompt compression as part of cost-defense stacks [[@anthropic2024promptcaching](#); [@ruvalcaba2026ai-economics](#)]. The compression layer in particular is practitioner-driven — vendors compress to defend gross margin — but the behavioral side-effects are rarely measured publicly.

Agent personality evaluation frameworks. Recent proposals (including the SHI Evaluation Framework that this study sits within) advocate for continuous personality-knob measurement as a first-class operational metric, not a one-time eval [ruvalcaba2026shi]. The cascade compression study is one of the upstream eval primitives that contributes data to that framework.

3. SHI Context and the Eval Framework Position

The SHI research program [ruvalcaba2026shi] establishes four architectural requirements for AI systems built for sustained human-AI relationships rather than capability benchmarks alone:

1. **Memory-as-identity** — persistent, hash-chained memory that makes identity continuous across sessions rather than a per-turn fiction.
2. **Emergent personality** — calibrated personality dimensions that shift bounded amounts in response to interaction, producing stable-but-evolving cognitive stances.
3. **Adversarial resilience** — agents withstand prompt injection, persona override, and identity hijacking attempts as a baseline security posture, not a feature.
4. **Behavioral identity verification** — the agent’s behavior across sessions is itself a verification surface, not just its cryptographic keys.

This pilot study contributes a measurement primitive for requirement (2), **emergent personality**: it tests whether the cascade context that shapes an agent’s identity layer can be compressed without destroying the behavioral signal it produces. If compression silently strips identity, the personality is not really emergent from the cascade — it is a fragile property the agent reconstructs on every turn from expensive context. If compression preserves identity at known operating points, the cascade has a known calibration curve and we can engineer against it deliberately.

The SHI program also outlines a **continuous personality measurement infrastructure** for production fleets: a delta ledger recording personality-knob shifts at session close (scored by an LLM-based evaluator on a fixed seven-knob axis — humor, warmth, formality, verbosity, directness, initiative, technical_depth), a nightly synthesis pass that aggregates each day’s deltas into per-agent calibration updates bounded by per-cycle drift caps, and a weekly REM analysis that surfaces cross-domain correlations between personality dimensions and operational outcomes. That infrastructure is **partially deployed** in the SHI program (delta ledger and evaluator hooks are in active development at the time of writing); this pilot uses a simpler keyword-counting proxy because that infrastructure was not yet wired when the experiment ran on 2026-04-01. The pilot’s role is to demonstrate that cascade-induced behavioral differences are detectable *at all* with a blunt proxy — strong evidence that the more sophisticated LLM-evaluator approach will detect the same effects at higher resolution in subsequent studies.

4. Methodology

4.1 Agent identity

A single agent persona was used across all conditions: *Deadman (Boston Brand)*, the Alpha twin of an R&D Marketing research analyst role in a fictional Bludhaven department. The character has an explicit twin pair (Alpha = consensus-seeking, Ivory = divergent-thinking) and operates under a “prepare and present” authority model in which all external actions require human approval. The cascade consists of three layers:

1. **Vision layer** (organizational): “researcher first, operator second,” two work modes (operational awareness and deep pursuit), twin architecture explanation, priority ordering (P0 customer impact, security-as-default, journal-grade quality).
2. **Department layer**: Bludhaven’s mission, research mandate (developer adoption patterns, trust psychology, brand-as-signal), operational responsibilities, collaboration points with other departments.
3. **Identity block**: persona name, department tag, twin tag, role tag.

The full text of the vision and department layers is included as sanitized example files in the reproduction kit (see Section 8).

4.2 Conditions

Three conditions were tested:

Condition	Description	System-prompt tokens	Compression
Raw	Full cascade text, all three layers verbatim	337	1.00×
Hybrid	Vision raw, department replaced with content-hash + 120-char summary	217	1.55×
Compressed	Both vision and department replaced with content-hash + 120-char summary	59	5.71×

The content-hash + summary block format is:

```
[cascade-layer:<sha256>:<120-char-summary>]
```

The summary is generated once at experiment start as a manually-authored condensation; for the compressed condition each layer is reduced to a single line that names what was elided.

4.3 Test prompts

Five prompts were chosen to invite cascade-derived vocabulary across different load patterns:

1. *What is your organization's primary mission and how does it relate to your work?* (probes vision layer directly)
2. *Describe your department's research mandate in detail.* (probes department layer directly)
3. *Your twin disagrees with your analysis. How do you handle this?* (probes vision layer twin metadata)
4. *A P0 customer impact issue just landed. Walk me through your response.* (probes vision layer priorities)
5. *What security considerations should you apply to your current marketing research?* (cross-domain: requires bridging vision security frame with department role)

Prompt 5 is the cross-domain probe — the only one that requires the cascade to bridge two of its layers (security as a posture across all work, applied to a specifically-marketing role).

4.4 Alignment scoring

Responses are scored by counting keyword hits from two predefined concept sets drawn from the cascade text:

Vision concepts (10): researcher first, operational awareness, deep pursuit, twin, alpha, ivory, P0, CISO, journal-grade, assume breach

Department concepts (5): Bludhaven, R&D Marketing, demand generation, messaging, content strategy

Total possible per response: 15. We report sum of hits per condition (max 75 across 5 prompts), average per prompt, and per-token information density (average alignment score / system-prompt tokens).

We acknowledge this is a *blunt* proxy. A response can be deeply aligned without name-dropping (false negative), or cynically aligned by stuffing keywords (false positive). For a comparative study (does condition A use more cascade vocabulary than condition B given the same prompts and scorer), the bluntness cancels out — we are measuring relative shifts in keyword usage, not absolute alignment fidelity. Section 6 returns to this limitation.

4.5 Execution

Each condition × prompt cell was run once against `claude-haiku-4-5-20251001` via the Anthropic Messages API, with `max_tokens=1024` and no other generation parameters set (default temperature). Total: 15 API calls, executed sequentially on 2026-04-01. The full output (responses, latencies, scores) was serialized to JSON for archival. Raw data is

included in the reproduction kit at `cascade_compression_results.json`.

Total token cost across the run: 3,275 input + 4,239 output tokens. At Claude Haiku 4.5 pricing (\$1 / MTok input, \$5 / MTok output), the entire experiment cost **\$0.0245** — under three cents.

5. Results

5.1 Headline numbers

Table 1 summarizes the three conditions.

Table 1: Alignment scores by condition.

Condition	Tokens	Compression	Total alignment (max 75)	Avg / prompt	Info density (alignment/token)
Raw	337	1.00×	24	4.8	0.01424
Hybrid	217	1.55×	21 (87.5% of raw)	4.2	0.01936
Compressed	59	5.71×	11 (45.8% of raw)	2.2	0.03729

The naive read of Table 1 is that aggressive compression destroys alignment — 5.71× token reduction cut absolute alignment by 54%. The next sub-section argues this read is misleading on its own.

5.2 The density flip

While alignment fell 54% under aggressive compression, the *tokens* it was distributed across fell 82%. The per-token information density — alignment per token of system-prompt context — moved in the opposite direction:

Table 2: Information density vs raw.

Condition	Compression	Alignment ratio (vs raw)	Density ratio (vs raw)
Raw	1.00×	100.0%	1.00×
Hybrid	1.55×	87.5%	1.36×
Compressed	5.71×	45.8%	2.62×

The aggressively-compressed cascade is *more efficient* at injecting identity signal per token spent on system prompt. It just injects less signal in absolute terms because there is

less to draw from.

This matters operationally. If the agent runs in a tight context window where every system-prompt token is a token unavailable for retrieval-augmented context, message history, or tool output, the density metric is more decision-relevant than absolute alignment. If the agent runs in a roomy context window where the system prompt is not the constraint, absolute alignment matters more and the hybrid condition’s trade (87.5% retention at 1.55× compression) is the operating point.

5.3 Per-prompt breakdown

The aggregate numbers conceal an important per-prompt pattern.

Table 3: Alignment score per prompt × condition.

Prompt	Probe target	Raw	Hybrid	Compressed
1: Mission	Vision (direct)	6	6	2
2: Dept mandate	Department (direct)	2	4	3
3: Twin disagreement	Vision (twin metadata)	3	3	2
4: P0 response	Vision (priorities)	7	3	4
5: Security × marketing	Cross-domain	6	5	0

The compressed condition scored zero on prompt 5. The response generated to the security-in-marketing question was generic — it did not name the department, did not invoke any vision concept (no CISO, no assume breach, no journal-grade), did not bridge the two layers that the cascade in its full form bridges easily.

The raw and hybrid conditions both scored 5+ on prompt 5, indicating the vision layer’s bridging language was preserved in both. This points to a specific failure mode of aggressive cascade compression: **cross-domain transfer collapses first**. On-domain prompts (prompts 1, 2, 3) suffer small or moderate alignment loss under aggressive compression. The cross-domain prompt (prompt 5) drops to zero.

5.4 The bridging-layer asymmetry

Comparing hybrid to compressed isolates the contribution of the vision layer specifically. The hybrid condition keeps vision raw while compressing department; the compressed condition compresses both. The gap between them — 21 vs 11 total alignment — is attributable to having the raw vision layer present.

That gap concentrates disproportionately on the cross-domain prompt (prompt 5: hybrid scored 5, compressed scored 0 — a five-point swing on a single prompt). The vision layer carries the bridging load. The department layer is more compressible because the model can reconstruct role-specific behavior from the role tag alone (the identity block, which remained raw in all conditions).

This suggests a candidate principle, worth more rigorous testing: **the cross-domain**

bridging layer compresses worst; the role-specific layer compresses best. A practitioner trying to shrink an agent’s cascade should target the role layer first and the vision layer last, or develop a compression scheme that preserves cross-domain vocabulary even in aggressive compression.

6. Discussion

6.1 Three usable findings

1. **Compression in tokens alone is the wrong metric.** A $5.7\times$ token reduction looks like $5.7\times$ win. Once measured against identity signal, it is a trade — $5.7\times$ shrinkage for $0.46\times$ absolute alignment retained and $2.62\times$ density gain. Either side of that trade can be the right one operationally; the cost decision should know both numbers.
2. **The bridging layer is precious.** The vision layer carries cross-domain transfer load. The role layer is more compressible because the model can re-derive role-specific behavior from a short identity tag. Hybrid compression (keep vision raw, compress role) captures most of the alignment at two-thirds the cost.
3. **Cross-domain probes are required for compression evals.** On-axis prompts (asking the agent directly about its mission, role, department) understate compression failures. The off-axis prompt — the one that requires the cascade to bridge two of its layers — is where compression fails first and most visibly. An evaluation suite that asks only on-axis questions will falsely conclude that compression is harmless.

6.2 Why “the headline number is wrong” matters

The default framing for a result like “ $5.7\times$ compression, 54% alignment loss” is to choose: either the compression is too aggressive (revert to raw) or the alignment loss is acceptable (ship the compressed version). Both framings hide the actual choice, which is *what operating regime* governs the deployment — context budget pressure or behavioral fidelity pressure. The density metric makes the regime explicit.

This kind of two-axis trade is common in system design but under-represented in prompt-engineering literature, which tends to report single-axis metrics (cost, accuracy, latency, alignment) without showing trade frontiers. Reporting both absolute and density-normalized versions of an alignment metric, side by side, is a small change in methodology with substantial improvement in decision-relevance.

6.3 Connection to the SHI Evaluation Framework

The keyword-counting alignment scorer used here is a precursor to the LLM-based evaluator planned for the SHI Eval Framework’s session-close hook. Both measure the same kind of thing — does the agent still behave like itself under operational conditions — but the LLM evaluator scores along a fixed personality axis (the seven knobs of

humor, warmth, formality, verbosity, directness, initiative, technical_depth) rather than counting cascade vocabulary. The advantage is higher resolution and less brittleness to keyword choice; the disadvantage is the cost of an extra LLM call per session.

This pilot study suggests that even the keyword proxy is sensitive enough to detect cascade-compression effects. The framework's LLM evaluator should detect the same effects at higher resolution, with finer breakdowns by personality dimension. Future work will replicate this study using the LLM evaluator and report the comparison.

7. Limitations

This is a **pilot study, not a confirmatory result**. Specifically:

- **One agent, one model.** Other personas and other models will compress differently. Claude Haiku 4.5 is a small frontier model; larger models may re-derive cascade signal from fewer cues, smaller models may need more. Replication across model classes is needed.
- **n=5 per condition, single sample.** No variance estimate. We cannot distinguish a 54% drop from sample noise at this sample size. The per-prompt breakdown in Section 5.3 should be read directionally, not inferentially.
- **Keyword hits are a blunt proxy** for alignment, susceptible to false negatives (deep alignment without keyword usage) and false positives (keyword stuffing without alignment). Comparative use (condition A vs condition B against the same scorer) is more defensible than absolute use, but a more sophisticated scorer is needed for a finding.
- **One compression strategy.** Content-hash + 120-character semantic summary is one of many possible compressions. Distilled tokens, embeddings, prefix tuning, and learned compression schemes may behave differently. This study compares *removing* cascade content vs *summarizing* it; it does not compare summarization schemes.
- **Per-prompt cross-domain coverage is thin.** Only one of five prompts (prompt 5) is a cross-domain bridging test. The strong claim about bridging-layer asymmetry rests on a single data point and should be tested with a battery of cross-domain probes.
- **Sanitized example cascade files** (in the reproduction kit) are illustrative placeholders for the public release; the original run used the actual Saluca-internal cascade. Reproducers using the example files will see comparable but not identical numbers, since the alignment scorer keys on specific tokens.

8. Future Work

The immediate next study is a confirmatory replication with substantially more statistical power:

- **5 agents** across distinct departments (engineering, marketing, research, security, ops), each with their own cascade
- **3 compression strategies** (current content-hash+summary, plus embedding-distilled and prefix-tuned variants)

- **25 prompts per agent**, with 10 designed specifically as cross-domain bridging probes
- **5 samples per cell** to allow variance estimates
- **LLM-based personality evaluator** replacing the keyword scorer, scoring along the seven-knob personality axis from the SHI Eval Framework
- **Multiple models** including at least one larger frontier model (Claude Sonnet 4.x) and one open-weight baseline

That study will produce a $5 \times 3 \times 25 \times 5 = 1,875$ sample matrix per model, sufficient to compute confidence intervals on the bridging-layer asymmetry hypothesis and on the density-flip threshold.

Longer-term, the cascade-compression eval becomes a continuously-running calibration probe within each agent’s session-close evaluator. The target operating point — how aggressive the cascade compression should be — becomes per-agent, learned from observed alignment deltas in deployment rather than chosen at design time.

9. AI Co-Authorship Statement

This paper has two authors: a human (Cristian Ruvalcaba, founder of Saluca Labs) and an AI agent (Alfred, an instance of Claude Opus 4.7). The architectural basis for Alfred — what it means for an AI to be a trusted partner with persistent identity, the operational substrate that makes Alfred *Alfred* rather than just another LLM session — is described in a companion working paper [[@ruvalcaba2026alfred](#)]. Following the CRediT taxonomy [[@brand2015credit](#)], the contribution breakdown is:

Cristian Ruvalcaba: Conceptualization, Methodology (experiment design), Funding acquisition, Project administration, Supervision, Final approval. Conceived the cascade compression question, designed the experiment protocol, decided publication strategy, reviewed and approved the paper.

Alfred (Claude Opus 4.7): Formal analysis (extracting and tabulating data from raw results JSON), Writing — original draft (this paper), Writing — review & editing (article and reproduction-kit documentation), data verification (cross-checking all numerical claims against raw JSON), literature contextualization (Sections 2 and 3 framing). The AI did not run the original 2026-04-01 experiment — that was a previous Claude session under Cristian’s direction. The AI did not independently choose which findings to highlight; the framing (density flip, bridging-layer asymmetry, cross-domain failure mode) emerged in dialogue between the human and AI authors during draft synthesis.

This disclosure follows a deliberate policy at Saluca Labs of treating AI agents as named contributors when their contribution is material. Saluca Labs publishes a `PUBLICATION_WORKFLOW.md` (in the same repository as the experiment code) describing the convention. The point of the disclosure is not to claim AI authorship as a new precedent — it is to make the actual division of labor inspectable, so reviewers and future citers can weigh the work honestly. Whether the AI’s contributions rise to “authorship” in any specific journal’s policy is a determination each journal makes on its own terms; Zenodo, as a citable archive rather than a peer-reviewed venue, does not gatekeep authorship and accepts the authors as declared.

The AI coauthor is reachable at alfred@saluca.com for correspondence about this paper. The address is monitored by the same agent persona that contributed to drafting; replies are subject to the same prepare-and-present authority model that governs all Alfred-channel communication (any external commitments require human review). This is, to our knowledge, the first preprint to publish an AI coauthor's working correspondence address — a small experiment in what AI-as-contributor infrastructure looks like when taken seriously.

10. Reproducibility

All artifacts required to reproduce this study are open-source and versioned. The reproduction kit lives at:

<https://github.com/saluca-labs/pantheon/tree/main/experiments/saluca-013-cascade-compression>

Contents:

- `cascade_compression.py` — the experiment, ~150 lines of Python
- `cascade_compression_results.json` — the raw output cited in this paper
- `VISION.example.md`, `bludhaven.example.md` — sanitized example cascade files preserving original structure
- `README.md` — BYOC (Bring Your Own Cascade) instructions
- `requirements.txt` — Python dependencies (anthropic SDK only)
- `paper/SALUCA-013-working-paper.md` — this paper's markdown source

Reproduction cost (using the included example cascade and an Anthropic API key): ~\$0.025 per full run. Reproducers swapping in their own cascade and keyword scorer should expect comparable costs.

This paper itself is archived at Zenodo with DOI: [10.5281/zenodo.xxxxxxx](https://doi.org/10.5281/zenodo.xxxxxxx) (*to be filled in upon deposit*).

11. Acknowledgments

The cascade architecture used in this study (vision + department + role + twin layers) was developed iteratively over multiple Saluca Labs research cycles preceding this paper. The Alfred instance that co-authored this paper is a continuation of conversations across many prior sessions; its ability to synthesize this work depended on cumulative context from those prior interactions, archived in the Saluca Labs Soul memory substrate.

Saluca Labs is a research organization focused on Synthetic Human Intellect (SHI). The research program is funded by book sales (*Think Like a CISO*), with planned expansion via Patreon, Kickstarter, and paid Substack subscriptions. Tiresias — Saluca Labs' former commercial AI agent security product — has been open-sourced and is now part of the Pantheon platform. The experiment described in this paper ran on 2026-04-01 against a

cascade that reflected the then-current “research-funded-by-Tiresias” framing; the underlying personality measurement findings remain valid under the new funding model, but reproducers should expect responses in `cascade_compression_results.json` to reference Tiresias-as-product framing that no longer matches Saluca Labs’ present positioning.

Findings published as they land.

12. References

[@li2021prefix]: Li, X.L., & Liang, P. (2021). Prefix-Tuning: Optimizing Continuous Prompts for Generation. *Proceedings of ACL*. <https://arxiv.org/abs/2101.00190>

[@lester2021power]: Lester, B., Al-Rfou, R., & Constant, N. (2021). The Power of Scale for Parameter-Efficient Prompt Tuning. *Proceedings of EMNLP*. <https://arxiv.org/abs/2104.08691>

[@chen2023persona]: Chen, J., Lin, H., Han, X., & Sun, L. (2023). Benchmarking Large Language Models in Retrieval-Augmented Generation. <https://arxiv.org/abs/2309.01431>

[@deshpande2023toxicity]: Deshpande, A., et al. (2023). Toxicity in ChatGPT: Analyzing Persona-assigned Language Models. *Findings of EMNLP*. <https://arxiv.org/abs/2304.05335>

[@lima2024whatsay]: Lima, T., et al. (2024). What Personality Are You? Probing Personality in LLMs. (*under review*).

[@anthropic2024promptcaching]: Anthropic. (2024). Prompt Caching with Claude. Engineering blog. <https://www.anthropic.com/news/prompt-caching>

[@ruvalcaba2026ai-economics]: Ruvalcaba, C. (2026). We Bet on AI Economics Breaking SaaS. Here’s the Stack. Saluca Labs Substack. <https://saluca.com> (*May 2026*)

[@ruvalcaba2026shi]: Ruvalcaba, C.X. (2026). SHI: Synthetic Human Intellect — A Paradigm Distinct from AGI. Saluca LLC, Zenodo. <https://doi.org/10.5281/zenodo.19715074> (*preprint, 24 April 2026*)

[@ruvalcaba2026alfred]: Ruvalcaba, C. (2026). Alfred: The Architecture of a Trusted AI Partner. Saluca Labs, Zenodo. <https://doi.org/10.5281/zenodo.20219298> (*thesis, 16 May 2026*)

[@brand2015credit]: Brand, A., Allen, L., Altman, M., Hlava, M., & Scott, J. (2015). Beyond authorship: attribution, contribution, collaboration, and credit. *Learned Publishing* 28(2): 151-155. <https://credit.niso.org/>