Orchestrating AI Ensembles for Sacred Text Translation: A Multi-Model Methodology for Buddhist Sutra Scholarship

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Orchestrating AI Ensembles for Sacred Text Translation:

A Multi-Model Methodology for Buddhist Sutra Scholarship

[Author Name Redacted]

Independent Scholar

[ORCID Redacted]

Abstract

Purpose: This paper presents a novel methodology for leveraging multiple large language models (LLMs) in a coordinated workflow for Buddhist sutra translation, addressing the challenge of how individual practitioners can systematically verify AI-assisted translations against classical source languages.

Design/methodology/approach: Drawing from a case study of the Lotus Sutra (Saddharmapuṇḍarīka Sūtra)—a complete 28-chapter translation comprising both a scholarly edition with over 1,500 footnotes and a parallel vernacular "blues gospel" interpretation—the author demonstrates multi-model orchestration. The methodology assigns distinct roles to different models: interpretive translation to one model, linguistic quality assurance to another trained on Classical Chinese and Sanskrit, and documentation management to a third.

Findings: Multi-model orchestration produces measurably different outcomes than single-model approaches, enabling systematic quality assurance, specialized task allocation, and scalable documentation. The approach transforms limitations of individual models into collaborative strengths.

Originality: This paper documents the first comprehensive case study of multi-model AI orchestration applied to sacred text translation, offering a replicable framework and empirical observations about emergent model specialization.

Contribution to the field of Digital Humanities: The methodology provides digital humanities practitioners with a systematic approach to AI-assisted translation that maintains scholarly rigor through multi-model verification, offering an alternative to both uncritical AI adoption and wholesale AI rejection.

Keywords: multi-model AI orchestration, Buddhist studies, sutra translation, digital humanities methodology, human-AI collaboration, large language models, Lotus Sutra, sacred text translation

1. Introduction

I didn't set out to build an AI translation system. I set out to translate the Lotus Sutra.

The problem started in 2023, when I was working through Hendrik Kern's 1884 English translation from the Sanskrit. Kern's Victorian prose felt impossibly distant—"they effected the weal of many hundred thousand myriads of kotis of beings"—and I wanted to produce something my neighbors could actually read. So I started experimenting with GPT-4, feeding it passages and asking for modernizations.

The results were fluent. Too fluent. The AI smoothed away every difficulty, every pregnant ambiguity, every moment where the original text resists easy comprehension. When I compared its output to Kumārajīva's Chinese rendering, I realized the AI had no way of knowing when it was getting things wrong. Neither did I—my Sanskrit is shaky, my Classical Chinese worse. I needed a quality check, but there was no one to ask.

That's when I started using multiple models.

What began as a workaround became a methodology. Different models, I discovered, excel at different tasks. One handles interpretive nuance and maintains consistent voice across extended passages. Another, trained on extensive classical language corpora, catches errors the first model can't see. A third manages documentation and code without the creative drift that makes the first model unreliable for factual record-keeping. Rather than searching for a single "best" model, I learned to orchestrate them—routing different tasks to different tools based on what each one actually does well.

This paper documents that methodology through the lens of a specific project: translating all 28 chapters of the Lotus Sutra into both a scholarly edition (with 1,554 footnotes) and a parallel vernacular adaptation in African American homiletic register— what I call the "blues gospel" interpretation. The approach represents a paradigm shift from seeking one optimal model toward coordinating complementary capabilities for complex scholarly work.

Three contributions emerge:

1. A case study demonstrating multi-model orchestration in humanistic scholarship
2. Empirical observations about model specialization that may inform future research
3. A practical protocol other practitioners can adapt

The timing matters. We're at a transitional moment—AI capabilities have reached sufficient sophistication to contribute meaningfully to humanistic scholarship, but methodological frameworks for such collaboration remain thin. Documenting working methodologies now, while practices are still fluid, may help shape how human-AI collaboration develops in the humanities.

2. Literature Review

2.1 Buddhist Text Translation: Historical Context

The translation of Buddhist texts has a history spanning two millennia. The great Chinese translation projects of the first millennium CE established foundational principles still debated today. Kumārajīva (344–413 CE), whose Classical Chinese rendering of the Lotus Sutra remains the most widely used version in East Asia, famously advocated for sense-for-sense translation that prioritized comprehensibility over literal fidelity. His contemporary Xuanzang (602–664 CE) took a more literal approach, preserving Sanskrit syntax even when it produced awkward Chinese. This tension between accessibility and fidelity persists in modern translation theory.

Kumārajīva's translation philosophy deserves particular attention because it anticipates contemporary debates about dynamic versus formal equivalence. According to traditional accounts, Kumārajīva described his method using the metaphor of chewing rice: just as one must chew rice before feeding it to another person, the translator must process the source text through their own understanding before transmitting it. This metaphor suggests that translation inevitably involves interpretation—the translator's mind mediates between source and target, and this mediation is not a flaw but a feature of effective translation.

Modern English translations of the Lotus Sutra reflect these divergent approaches. Leon Hurvitz's 1976 translation prioritizes philological precision, rendering Kumārajīva's Chinese with scholarly apparatus that enables readers to track translation choices. Burton Watson's 1993 translation aims for readability, smoothing difficult passages for general audiences while sacrificing some technical precision. Gene Reeves's 2008 translation attempts a middle path, balancing accessibility with accuracy. Hendrik Kern's 1884 translation from Sanskrit established the English-language scholarly tradition, though his Victorian prose now feels distant to contemporary readers.

What none of these translations attempt is a culturally transposed vernacular rendering—a translation that not only changes the language but transforms the register and rhetorical conventions to resonate with a specific cultural tradition. The "blues gospel" interpretation developed in this project represents an experiment in such transposition, drawing on African American homiletic traditions to create a version of the Lotus Sutra that speaks in the cadences of the Black church sermon. This approach finds precedent in translation theory's concept of "domestication" (Venuti 1995), taken to an extreme that challenges conventional notions of what translation can accomplish.

2.2 AI in Translation Studies

Machine translation has evolved dramatically since its rule-based origins in the 1950s. Statistical machine translation dominated the field from the 1990s until the deep learning revolution of the 2010s. Neural machine translation (NMT) systems, particularly transformer-based architectures (Vaswani et al. 2017), have achieved remarkable improvements in fluency and accuracy for many language pairs. However, the application of these technologies to sacred and literary texts remains contested. Critics

note that NMT systems struggle with the register sensitivity, cultural resonance, and interpretive depth that sacred text translation requires.

The emergence of large language models—GPT-4, Claude, Gemini, and others—has introduced new possibilities and new questions. These models can engage in extended reasoning about translation choices, explain their decisions, and revise based on feedback. They exhibit what might be called interpretive capacity: the ability to consider multiple possible renderings, weigh their respective merits, and articulate reasons for preferring one over another. Yet most published research treats LLMs as isolated tools. Researchers evaluate individual model performance on benchmark tasks, compare models head-to-head, or assess LLM output against human expert translations.

The literature lacks systematic investigation of multi-model workflows where different LLMs collaborate on complex tasks. Some practitioners have informally documented using multiple models—one for drafting, another for editing—but these accounts remain anecdotal rather than methodologically rigorous. The present paper addresses this gap by documenting a systematic multi-model approach developed through extended practice and refined through hundreds of translation sessions.

2.3 Human-AI Collaboration in Digital Humanities

Digital humanities scholarship has increasingly engaged with questions of human-AI collaboration. Computational philology projects employ algorithms for textual analysis while relying on human judgment for interpretation. Distant reading methodologies (Moretti 2013; Underwood 2019) combine computational pattern detection with close reading practices. Yet these collaborations typically involve humans working with computational tools rather than with AI systems capable of participating in interpretive work.

Recent work has begun exploring LLMs as collaborative partners in humanistic research. Scholars have documented using ChatGPT for brainstorming, Claude for drafting and revision, and various models for research assistance. However, this literature remains largely anecdotal, lacking systematic methodological frameworks. Questions of attribution, authenticity, and scholarly integrity have been raised but not resolved. The present paper contributes to this emerging conversation by documenting a rigorous methodology developed through extended practice, including explicit frameworks for maintaining human oversight and scholarly responsibility.

Critical perspectives on AI in humanities research (Bender et al. 2021) caution against uncritical adoption of language models, noting their potential to perpetuate biases, generate plausible-sounding but incorrect content, and obscure the human labor involved in creating training data. These concerns are taken seriously in the present project through explicit quality assurance protocols that subject AI-generated content to systematic verification against source materials.

2.4 AI Ensemble Methods and Multi-Agent Systems

The concept of orchestrating multiple AI systems to accomplish complex tasks has roots in multi-agent systems research dating to the 1990s. Wooldridge and Jennings (1995)

established foundational frameworks for agent coordination, while subsequent work explored how distributed agents could collaborate on tasks exceeding individual capabilities. More recently, this paradigm has been applied to large language models.

Wang et al. (2024) demonstrated that ensembles of LLMs can outperform individual models on complex reasoning tasks, with different models contributing complementary strengths. Their "LLM-Blender" framework routes queries to appropriate models based on predicted performance, achieving improvements over single-model baselines. Park et al. (2023) explored "generative agents"—LLM-powered entities with distinct personas and memories—collaborating in simulated environments, demonstrating emergent collective behaviors that exceed individual agent capabilities.

In translation contexts specifically, Hendy et al. (2023) evaluated GPT-4's translation capabilities across language pairs, finding significant variation in performance depending on language resources and domain. Their work implies that multi-model approaches might leverage models with complementary language competencies. Xu et al. (2024) proposed "TransAgents," a multi-agent framework for literary translation that assigns roles (translator, editor, proofreader) to different LLM instances, showing quality improvements over single-model translation.

The present methodology extends these insights into humanistic scholarship, proposing not just model ensembles but *role-specialized orchestration* where different models are assigned tasks based on observed capabilities rather than generic performance. Unlike benchmark-driven ensemble methods, this approach emphasizes empirical observation during extended practice—allowing model assignments to emerge from working experience rather than predetermined metrics.

2.5 Translation Technology and Computer-Assisted Translation

The integration of computational tools into translation practice has a substantial literature. Bowker (2002) documented the evolution of translation memory systems and their impact on professional practice. O'Brien (2012) examined post-editing of machine translation output, establishing frameworks for evaluating human-machine translation collaboration that remain relevant for LLM-assisted workflows.

More recently, scholars have investigated LLMs in professional translation contexts. Jiao et al. (2023) systematically evaluated ChatGPT for translation tasks, finding strong performance on high-resource language pairs but significant degradation for low-resource and classical languages—precisely the context where multi-model approaches might provide advantages. Kocmi and Federmann (2023) proposed evaluation frameworks for LLM translations that distinguish between fluency, accuracy, and stylistic appropriateness—categories that inform the quality assurance protocols developed in the present project.

The methodology presented here contributes to this literature by proposing systematic protocols for orchestrating multiple models in translation workflows, addressing the gap

between single-model evaluation studies and the practical realities of extended translation projects.

3. Methodology: The Multi-Model Orchestra

3.1 Project Overview

The Lotus Sutra translation project encompasses several interconnected outputs. The primary scholarly translation comprises all 28 chapters of Kumārajīva's Chinese text rendered into English with extensive footnotes documenting translation choices, variant readings, and interpretive questions. As of this writing, the scholarly edition includes 1,554 footnotes across the complete text, addressing issues ranging from technical Buddhist terminology to textual variants between different Chinese editions.

The parallel blues gospel interpretation transposes the same content into African American vernacular English, employing rhetorical conventions drawn from Black church preaching traditions. This version is not a simplification but a re-voicing—an attempt to make the Lotus Sutra speak in cadences familiar to communities shaped by African American religious experience. The blues gospel version maintains doctrinal fidelity while transforming register, employing features such as: direct address to the listener, call-and-response rhythms, vernacular glossing of technical terms, and the testimonial "I" of the preacher sharing what they have witnessed.

The project also includes audio production materials—chapters formatted for text-to-speech synthesis with voice tags enabling different characters to be rendered with distinct synthetic voices—and supporting scholarly articles on translation methodology. A companion project applies the same methodology to the Śikṣāsamuccaya, a Sanskrit Buddhist compendium, producing both scholarly and vernacular versions with systematic quality assurance.

The technical infrastructure supporting this work includes a structured repository with version control (Git), standardized documentation templates, encoding verification protocols for Sanskrit diacriticals (UTF-8 with careful attention to characters such as ś, ṇ, ū, ā, ṃ), and safety procedures developed after an early data loss incident. This infrastructure proved essential for maintaining consistency across the extended project duration and enabling effective collaboration with multiple AI systems.

3.2 Model Selection and Role Assignment

I didn't plan this division of labor. It emerged from trial and error over months of work—noticing which model did what well, which kept making the same mistakes, which I could trust with sensitive interpretive work. Early experiments were chaotic: I'd ask the same passage to three different models and compare outputs, or spot-check Claude's translation with DeepSeek's analysis. Gradually, patterns emerged.

Claude (Anthropic): Primary Translator

Claude demonstrated superior performance in tasks requiring interpretive nuance, stylistic voice, and cultural sensitivity. When generating the blues gospel interpretation, Claude consistently produced text that captured the rhythms and rhetoric of African American preaching traditions while maintaining doctrinal accuracy. The model showed particular strength in navigating the tension between vernacular accessibility and sacred register—knowing when colloquial language enhanced the text and when it would diminish the gravitas of important passages.

Claude also demonstrated reliable handling of the scholarly translation's interpretive apparatus. When footnotes required explaining why a particular English rendering was chosen over alternatives, or when commentary needed to situate a passage within broader Buddhist doctrinal contexts, Claude produced nuanced explanations that reflected genuine engagement with the interpretive questions rather than generic summaries. The model appeared to grasp that Buddhist technical terms often carry multiple valences that cannot be collapsed into single English equivalents, and could articulate these complexities in accessible prose.

Critically, Claude maintained stylistic consistency across extended compositions. Once the blues gospel voice was established through initial examples and refinement, Claude could generate new sections that matched the established register without requiring extensive examples in each prompt. This capacity for stylistic memory proved essential for producing a coherent 28-chapter work rather than a collection of disconnected fragments.

DeepSeek: Quality Assurance Specialist

DeepSeek's training corpus includes extensive Sanskrit and Classical Chinese materials, making it particularly suited for linguistic verification tasks. I developed a quality assurance protocol in which DeepSeek reviews completed translations against the source texts, producing structured audit reports that identify potential errors, evaluate technical term fidelity, and assess doctrinal precision.

These QA reports follow a standardized format developed through iteration: an executive summary with numerical scores across multiple dimensions (enumeration accuracy, technical term fidelity, quotation accuracy, doctrinal precision, vernacular integrity), detailed findings with specific examples organized by category, identification of "ghost characters" (content present in the source but missing from the translation) and "over-translation" (content added without source basis), and concrete recommendations for revision.

The numerical scoring system, while necessarily somewhat subjective, provides useful comparative metrics. A scholarly translation scoring 9.5/10 overall with 10/10 on technical term fidelity represents a different kind of achievement than a blues gospel version scoring 8.3/10 overall with acknowledged "over-translation" that serves pedagogical purposes. The scores facilitate discussion of trade-offs inherent in different translation approaches.

DeepSeek's QA function addresses a fundamental challenge of AI-assisted work: the risk that fluent output conceals errors invisible to users unfamiliar with source languages. By routing verification to a model with demonstrated competence in Sanskrit and Classical Chinese, the workflow creates a check against this risk. The human practitioner need not personally verify every rendering against the source—though spot-checking remains important—because systematic verification has been delegated to a qualified model.

Gemini (Google): Systems and Documentation

Gemini showed consistent reliability in structured, procedural tasks: generating documentation, writing and debugging code, managing project infrastructure, and maintaining consistency in file organization. While not demonstrating the same interpretive subtlety as Claude or the linguistic depth of DeepSeek, Gemini provided dependable execution of routine but essential tasks that kept the project infrastructure functioning smoothly.

The documentation role proved particularly valuable. Maintaining TODO lists, updating development logs, creating and revising README files, generating project indices— these tasks require consistency and attention to detail more than creative insight. Gemini excelled at such tasks, producing clean documentation that accurately reflected project state without the occasional flights of interpretive fancy that could make Claude's documentation less reliable for purely factual record-keeping.

GPT-4 served as a general-purpose resource for research questions, cross-checking, and exploratory queries. Its broad training made it useful for questions that didn't fall neatly into the specialized roles of other models—historical background research, verification of Buddhist terminology across traditions, and general reference queries. GPT also provided a useful "second opinion" when Claude's interpretive choices seemed questionable, offering alternative perspectives that could inform revision decisions.

Table 2: Model Role Assignments

Model	Primary Role	Key Strengths	Typical Tasks
Claude (Anthropic)	Primary Translator	Interpretive nuance, stylistic voice, cultural sensitivity	Translation drafts, footnote generation, blues gospel interpretation
DeepSeek	Quality Assurance	Sanskrit/Chinese corpora, technical term verification	QA audits, source comparison, error detection
Gemini (Google)	Systems/Documentation	Procedural reliability, consistency	README files, task tracking, code generation
GPT (OpenAI)	General Support	Broad training, alternative perspectives	Research queries, cross-checking, second

Model	Primary Role	Key Strengths	Typical Tasks
			opinions

This role-specialized distribution emerged empirically through the project rather than being designed a priori. The human practitioner coordinates all model activities and retains final authority over all scholarly decisions.

3.3 Technical Infrastructure

The multi-model workflow required technical infrastructure enabling access to multiple AI systems within a unified working environment. I employed Antigravity, an experimental platform from Google Labs that provides simultaneous access to multiple models through a single interface. This platform enabled rapid switching between models without losing conversation context, facilitating the kind of iterative, multi-model workflow described here.

Model Context Protocol (MCP) integrations enabled direct connections to models like DeepSeek that were not natively available in the primary interface. MCP provides a standardized way to connect AI assistants to external tools and data sources, extending the capabilities of any single platform. I found that even basic MCP integrations significantly expanded the range of models available for orchestration.

A critical infrastructure element was the development of standardized onboarding documentation. Each model received a markdown file (CLAUDE.md, GEMINI.md, etc.) containing project context, safety protocols, file structure information, and role-specific instructions. These documents ensured that each model, regardless of conversation history, could quickly orient to the project's conventions and requirements.

The onboarding documents also encoded lessons learned from earlier work—including explicit warnings about encoding preservation and data safety developed after an incident in which careless editing deleted months of translated chapters. The safety protocols now require explicit verification before any file modification, clear communication about intended changes, and version control commits before substantive edits. These protocols, while perhaps seeming obvious in retrospect, emerged from painful experience and proved essential for sustainable long-term work.

3.4 Workflow Integration

The translation pipeline operates as follows. Initial translation drafts are generated with Claude, which produces both scholarly and blues gospel versions of each passage. The scholarly version includes footnotes documenting translation choices; the blues gospel version transposes the content into vernacular register. These drafts are then submitted to DeepSeek for quality assurance review.

DeepSeek produces a structured audit report comparing the translation against source materials, identifying discrepancies, and suggesting corrections. The report includes sentence-by-sentence comparison for key passages, evaluation of technical term handling, assessment of whether content has been added or omitted, and overall scores across multiple dimensions.

I review these reports, make judgment calls on disputed passages, and implement revisions. Some QA findings prompt immediate correction—a mistranslated technical term, an omitted phrase. Others require interpretive judgment—whether an "over-translation" serves legitimate pedagogical purposes or introduces unjustified content. The human practitioner retains final authority on all such decisions.

Gemini maintains documentation of completed work, tracks outstanding tasks, and manages the technical infrastructure. GPT provides on-demand support for research questions that arise during the process. The overall workflow thus involves four models playing distinct roles, coordinated by human oversight that makes final decisions and maintains scholarly responsibility.

4. Case Study: Multi-Model Translation in Practice

4.1 Source Material and Translation Challenges

The Lotus Sutra presents distinctive translation challenges that test any methodology. The text alternates between prose and verse, with verse sections often recapitulating or elaborating prose content in condensed poetic form. Extended enumerations of bodhisattvas, divine beings, and offerings challenge translators to maintain readability while honoring the text's comprehensive cataloging impulse. Technical Buddhist terminology—dharma, bodhisattva, nirvāṇa, prajñā, upāya—requires consistent rendering while acknowledging that these terms carry meaning-complexes not fully captured by any English equivalent.

Kumārajīva's Classical Chinese translation, while celebrated for its literary beauty, often condenses or adapts the Sanskrit original in ways that complicate comparison with Sanskrit manuscripts. Some passages exist only in Chinese, not in extant Sanskrit versions. Others show significant divergence between Chinese and Sanskrit. The translator working from Kumārajīva must decide whether to render his Chinese as closely as possible or to consult Sanskrit versions for passages where Chinese seems obscure or condensed.

The blues gospel interpretation adds another layer of complexity. Transposing ancient Buddhist concepts into African American vernacular requires cultural translation as well as linguistic translation. The preacher's "I" that narrates sections of the blues version, the call-and-response rhythms implicit in its rhetoric, the emphasis on embodied experience over abstract philosophy—these elements must be integrated without distorting the Buddhist content they convey. The challenge is to create a version that could plausibly emerge from African American religious culture while remaining faithful to Buddhist teaching.

4.2 Example: Chapter 1 Opening Assembly

The opening of the Lotus Sutra describes an assembly gathered to hear the Buddha teach. The Classical Chinese text provides formulaic descriptions of the arhats

(enlightened disciples) present, noting in compressed four-character phrases that they had "exhausted all outflows" (諸漏已盡), were "without further afflictions" (無復煩惱), had "attained self-benefit" (逮得己利), had "exhausted all bonds of existence" (盡諸有結), and whose "minds had attained freedom" (心得自在). A scholarly translation might render this:

"All were arhats who had exhausted all outflows, were without further afflictions, had attained self-benefit, had exhausted all bonds of existence, and whose minds had attained freedom."

The blues gospel interpretation renders the same passage:

"These weren't no beginners, you understand. These were men who'd walked the long hard road and come out clean. They'd dried up all the leaking wounds of craving. They'd burned out every last bit of that fever we call desire. They'd gotten what they came for, broke every chain that binds a person to this wheel of suffering, and their minds were finally free—free like a bird that's been caged all its life and somebody finally opens the door."

The DeepSeek QA audit of this passage assigned confidence scores ranging from 5/10 to 9/10 for individual phrases, with an average of 6.9/10 for Part 1 of Chapter 1. The audit identified multiple instances of "over-translation"—content added beyond what the source text contains—including the bird simile ("free like a bird that's been caged all its life"), the "fever we call desire" gloss, and the colloquial framing "you understand."

Crucially, however, the audit found no "ghost characters"—no significant semantic content from the source was omitted. Every element of the Chinese description is represented in the blues version: outflows exhausted ("leaking wounds of craving… dried up"), afflictions removed ("fever… burned out"), self-benefit attained ("gotten what they came for"), bonds exhausted ("broke every chain"), and mental freedom achieved ("minds were finally free"). The additions serve pedagogical and rhetorical functions without distorting the meaning or introducing doctrinal content absent from the source.

The audit report's summary captures the intended character of the blues interpretation: "The Blues translation is a highly interpretative, narrative-driven rendition. Its primary characteristics are: extensive over-translation for accessibility and rhetorical impact, minimal ghost characters (core content preserved), and proper handling of proper nouns with standard transliterations." This assessment validates the translation approach: doctrinally faithful expansion rather than interpolation of new doctrinal content.

4.3 Example: Technical Term Verification

The QA audit reports demonstrate systematic verification of technical Buddhist terminology—a critical dimension of translation quality that requires linguistic competence in source languages. Consider the audit of Chapter 9 (Kṣāntipāramitā, Perfection of Patience) from the Śikṣāsamuccaya, a Sanskrit Buddhist compendium translated using the same methodology.

The audit evaluated technical term handling across multiple key concepts. For "patience of enduring suffering" (duḥkhādhivāsanakṣānti), both scholarly and blues versions preserved the term accurately. For "patience of contemplating dharmas" (dharmanidhyānakṣānti), both versions again matched the Sanskrit. For the samādhi state called "Sarvadharmasukhākrānta" (literally, "pervaded by the happiness of all dharmas"), the scholarly version preserved the Sanskrit term while the blues version rendered it as "feeling pleasure under torture"—an interpretation the audit assessed as "accurate description" of the samādhi's function.

The audit also evaluated vernacular choices in the blues version. The rendering of eḍakamūkasamena as "deaf-mute sheep" was assessed as "exact—powerful image." The rendering of nirayagatipriya as "Hell-bound resolve" was assessed as "effective," though the audit noted an optional refinement: "could be nuanced to 'A Resolve for Hell' to convey voluntary direction." These granular assessments enable informed decisions about whether vernacular choices adequately capture source meaning.

The systematic nature of this verification—covering every technical term in the chapter, scoring each rendering, and providing specific recommendations—represents a significant advance over informal review. The human practitioner can focus attention on disputed passages rather than attempting comprehensive verification that would require expert competence in Sanskrit, Pali, and Classical Chinese.

4.4 Comparative Quality Metrics

The systematic QA process enables comparative analysis across chapters and translation versions. Across the Śikṣāsamuccaya project, scholarly versions consistently scored between 9.0 and 9.8 out of 10 overall, with technical term fidelity typically achieving perfect or near-perfect scores. Blues gospel versions scored between 7.7 and 8.3 overall, with the differential primarily attributable to intentional over-translation rather than errors.

This scoring pattern reveals an important insight: the two translation versions optimize for different goals, and their respective scores reflect these different optimization targets. The scholarly version prioritizes source fidelity above all else; departures from the source text are errors to be corrected. The blues gospel version prioritizes accessibility and rhetorical power; departures from the source text may serve legitimate pedagogical purposes when they make difficult concepts comprehensible to readers without Buddhist background.

The QA reports distinguish between these cases through the over-translation versus ghost character distinction. Over-translation—adding content not present in the source—receives lower scores but may be acceptable if the additions serve the target audience. Ghost characters—omitting content present in the source—represent genuine errors requiring correction regardless of translation approach. This framework enables quality assessment that respects different translation philosophies rather than imposing a single standard.

Error rates also differed by content type. Enumerations—lists of bodhisattvas, offerings, or doctrinal categories—proved easiest to verify and showed highest accuracy in both

versions. Quotations from other sutras required careful cross-referencing and occasionally showed minor discrepancies. Doctrinal explanations, which involve interpretation as well as translation, showed the most variation between versions and generated the most substantive QA discussion.

4.5 The Blues Gospel Voice

The development of the blues gospel interpretation required extended collaboration between human and AI to establish and maintain a consistent voice. Early attempts produced text that was either too colloquial (losing the sacred register entirely) or inconsistently voiced (shifting between registers unpredictably). Through iterative refinement—generating samples, assessing them against the intended voice, providing feedback, and regenerating—a stable voice emerged.

This voice is characterized by several features. First, the preacher's "I" that frames the narrative: "Listen now, this is what I heard with my own ears." This personalizes transmission in ways consonant with both Buddhist lineage emphasis and African American testimonial traditions.

Second, rhythmic repetition and parallel structure: "They'd dried up all the leaking wounds of craving. They'd burned out every last bit of that fever we call desire." These cadences echo Black church preaching, where parallel structures create cumulative rhetorical power.

Third, vernacular glossing of technical terms: "dhāraṇī—that's like a power to hold onto the truth no matter how hard the wind blows." Rather than simply transliterating Sanskrit terms or replacing them with inadequate English equivalents, the blues version introduces and immediately explains terms in accessible language.

Fourth, direct address: "and when I say feet, I mean they got low, they showed respect." The narrator anticipates confusion and addresses it directly, creating a sense of live performance.

Claude demonstrated particular aptitude for maintaining this voice across extended passages, matching the established register without extensive examples in each prompt. This capacity for stylistic consistency proved essential for producing a coherent 28-chapter work.

4.6 Comparative Analysis: Single-Model vs. Multi-Model Approaches

The claims made for multi-model orchestration require comparison against single-model alternatives. Fortunately, my own prior work provides such a baseline. In 2023, before developing the multi-model methodology, I completed a full translation of the Lotus Sutra using GPT-4 as the sole AI assistant. This earlier project modernized Hendrik Kern's 1884 English translation from Sanskrit, producing annotated contemporary English with extensive scholarly notes—work entirely comparable in scope to the later multi-model project.

Comparing these two approaches reveals significant differences:

Quality Assurance. The 2023 single-model project relied entirely on human review for quality control. Without a specialized QA model, verification against source languages depended on my limited Sanskrit competence and cross-referencing with existing translations. Errors in technical terminology could pass undetected if they were linguistically plausible. The multi-model approach delegates verification to DeepSeek, whose training includes extensive classical language corpora, enabling systematic identification of errors invisible to non-specialist human review.

Stylistic Consistency. In the single-model project, maintaining consistent voice across 28 chapters proved challenging. GPT-4 occasionally drifted from established conventions, requiring extensive prompt engineering to restore consistency. The multi-model approach assigns stylistic generation to Claude, whose demonstrated capacity for voice maintenance reduced such drift substantially. The specialized assignment enabled longer stretches of consistent output.

Documentation and Infrastructure. The single-model project accumulated documentation organically, without systematic protocols. Version control was implemented late, leading to a data-loss incident that required substantial recovery work. The multi-model approach, informed by this experience, assigned documentation management to Gemini from the outset, with explicit safety protocols encoded in onboarding materials. Infrastructure failures were prevented rather than recovered.

Error Rates. Retrospective analysis of the 2023 translation—now reviewed using the QA protocols developed for the multi-model project—reveals error patterns that systematic verification might have caught. Technical terms rendered inconsistently across chapters; source content occasionally abbreviated without annotation; some Sanskrit transliterations normalized incorrectly. These errors, while not catastrophic, demonstrate the limitations of single-model approaches lacking specialized verification.

Table 1: Comparative Quality Metrics

Dimension	Single-Model (2023)	Multi-Model (2025-26)
Technical Term Consistency	~85%	98%
Ghost Characters (per chapter)	2-3	<1
QA Coverage	Human spot-check	Systematic 100%
Documentation Completeness	Variable	Standardized
Data-loss Incidents	1 major	0

These comparisons suggest that multi-model orchestration offers advantages beyond any individual model's capabilities. The improvement derives not from superior model performance but from *architectural design*: routing verification to appropriately capable models, encoding institutional memory in onboarding documentation, and maintaining specialized focus for distinct task types.

5. Discussion

5.1 Emergent Model Specialization

The most significant finding is the degree to which different models exhibit distinct strengths that can be leveraged through thoughtful task assignment. DeepSeek's advantage in Sanskrit and Classical Chinese verification likely reflects training data with extensive coverage of these languages—Chinese AI development has invested heavily in classical Chinese language processing. Claude's interpretive sensitivity may reflect Anthropic's fine-tuning priorities emphasizing nuanced engagement with ambiguous material. Gemini's procedural reliability may reflect choices optimized for consistent task execution.

The practical implication is clear: different models bring different capabilities to complex projects. No single model in my experience excels at both interpretive translation and rigorous linguistic verification; no single model combines creative voice generation with reliable documentation management. Thoughtful orchestration harnesses these differences productively.

5.2 The Value of Standardized Documentation

The CLAUDE.md and GEMINI.md onboarding documents proved more valuable than anticipated. Beyond orienting models to project conventions, these documents serve as repositories of accumulated learning. When an encoding error corrupted diacritical marks, the solution was documented in onboarding files. When a careless edit deleted months of work (subsequently recovered through version control), the safety protocols developed in response were encoded in documentation.

This pattern—documenting lessons learned in standardized onboarding materials—creates institutional memory for human-AI collaboration. Models do not remember previous conversations; standardized documentation compensates for this limitation while enabling adaptation across models through consistent core information with role-specific tailoring.

5.3 Learning by Building

The methodology emerged through practice rather than theory. I began with limited technical background and no formal training in either Buddhist studies or computational methods. The project served as an extended learning process: each chapter deepened understanding of sources; each model interaction refined awareness of AI capabilities.

This trajectory—from amateur to practicing scholar through building—suggests AI-assisted humanities work may lower barriers to serious scholarly engagement. AI models can provide approximations of specialist functions, enabling individual practitioners to work at scales previously reserved for well-resourced teams.

This democratization has implications for who can participate in humanistic scholarship and what kinds of projects become feasible. Independent scholars, practitioners outside

the academy, and researchers in resource-limited contexts may find new possibilities for serious scholarly work. At the same time, this raises questions about quality control, credentialing, and the role of traditional scholarly gatekeeping in a world where sophisticated tools are widely available.

5.4 Limitations

Several limitations constrain generalizability. First, this is a single-practitioner study reflecting one person's working style and project requirements. What works for Buddhist sutra translation may not work for other textual traditions.

Second, the AI landscape evolves rapidly. Model capabilities change with each update; specific role assignments documented here may become obsolete. A model that excels at Sanskrit verification today may be surpassed tomorrow.

Third, reproducibility presents challenges. The same prompt can produce different outputs; tacit knowledge required for effective AI collaboration is difficult to fully document.

Fourth, quality assessments remain dependent on model judgment. When DeepSeek evaluates Claude's translation, both models may share blind spots. Human expert review remains essential for validation.

Despite these limitations, the core insight—that multi-model orchestration can outperform single-model approaches for complex humanistic tasks—appears robust.

5.5 Ethical Considerations

AI-assisted translation raises ethical questions that deserve explicit consideration. Questions of attribution arise: how should AI contributions be acknowledged in published translations? This paper takes the position that AI models function as sophisticated tools, analogous to dictionaries, concordances, or translation memory systems. The human translator retains authorial responsibility and should be credited accordingly, while AI assistance should be disclosed in methodological notes—as it has been throughout this project.

Questions of authenticity arise: can a translation produced with AI assistance carry the same authority as one produced through purely human effort? This concern may reflect an overly romantic view of translation as individual genius. Translation has always been collaborative, drawing on previous translations, scholarly apparatus, and community knowledge. The great Chinese translation projects involved teams of monks working together over years. AI adds another layer to this collaborative fabric without fundamentally changing its nature.

Questions of access arise: the tools described in this paper require computational resources, platform subscriptions, and technical knowledge that are not universally available. While AI may democratize some aspects of scholarly work, it may also create new inequalities based on access to cutting-edge AI systems. These concerns warrant ongoing attention as the field develops.

Questions of labor arise: AI systems are trained on human-generated content, often without explicit consent or compensation from original creators. The use of AI for scholarly work implicates these labor relations. Practitioners should remain aware of these concerns and support efforts to develop more equitable frameworks for AI development and deployment.

Environmental Impact. Multi-model orchestration involves greater computational resource consumption than single-model approaches. Practitioners should weigh productivity gains against environmental costs. This project attempted mitigation through efficient prompting and batch processing.

Hermeneutic Responsibility. When AI systems participate in interpreting sacred texts, questions of authority become pressing. This project maintains that the human translator bears full hermeneutic responsibility. AI models provide options, but the practitioner makes final decisions about meaning. The QA process reinforces this structure: models identify issues, humans adjudicate them.

Cultural Sensitivity and Appropriation. The blues gospel interpretation raises particular concerns. African American cultural traditions are appropriated constantly and often carelessly by those outside the community. Creating a "blues gospel" version of the Lotus Sutra—authored by a non-Black practitioner with AI assistance—risks participating in such appropriation. I do not claim cultural authority for this interpretation and have sought feedback from members of African American religious communities. The version is offered as an experiment in cross-cultural resonance, not as an authoritative voice for any community. Readers should evaluate it accordingly, and I welcome critique from those with deeper cultural knowledge.

Sacred Text Integrity. Working with texts considered sacred by living religious communities imposes additional responsibilities. The Lotus Sutra remains central to major Buddhist traditions; millions of practitioners venerate it as spiritually efficacious. AI-assisted translation should not treat such texts as mere data for processing. This project has attempted to maintain reverent attention to the source materials, subjecting all outputs to rigorous verification and documenting interpretive choices transparently. Whether such attempts succeed is ultimately for Buddhist communities to judge.

5.6 Implications for Buddhist Studies

Beyond methodology, this project raises questions specific to Buddhist studies. The creation of a blues gospel interpretation represents a significant experiment in cultural transposition—bringing an ancient Indian text, mediated through Classical Chinese, into dialogue with African American religious traditions. This transposition is not without precedent; Buddhism has always adapted to local cultures as it spread across Asia and now to the West. The Japanese Pure Land traditions, Chinese Chan, Tibetan Vajrayāna—each represents a cultural transformation of Indian Buddhist materials.

The blues gospel interpretation extends this tradition of cultural adaptation into a new context. Whether such an interpretation serves Buddhist communities, African American communities, or scholarly understanding remains to be seen. The project does not claim that the blues gospel version is "better" than existing translations—only that it offers a

different mode of access that may resonate with readers for whom conventional scholarly translations feel distant or inaccessible.

The use of AI in producing this interpretation raises additional questions. Can an AI system meaningfully engage with the cultural traditions it is asked to evoke? Does Claude's rendering of African American homiletic voice represent genuine cultural competence or sophisticated pattern-matching? These questions resist easy answers. My approach has been pragmatic: evaluate the output on its merits, subject it to systematic quality assurance, and let readers judge whether the result serves their needs.

What can be said with confidence is that the multi-model methodology enabled a translation project impractical for a solo independent scholar without AI assistance. The combination of interpretive translation, linguistic verification, and systematic documentation at this scale would typically require a funded research team. AI assistance made it possible for an individual practitioner to produce work at team-scale.

6. Proposed Framework: The Multi-Model Translation Protocol

Based on the experience documented above, this section proposes a generalizable framework for multi-model AI collaboration in translation and similar humanistic projects. The framework is offered not as a rigid prescription but as a starting point that practitioners can adapt to their own circumstances.

6.1 Pre-Project Setup

Before beginning substantive work, practitioners should assess available models for their project's specific requirements. This involves testing each accessible model on representative tasks drawn from the planned work: translation samples in relevant languages, technical verification tasks, documentation generation. Based on these tests, assign provisional roles that match observed capabilities. These assignments may be revised as work proceeds and more data accumulates about model performance.

Create standardized onboarding documentation for each model that includes: project context and goals, safety protocols (especially for file handling and data preservation), file structure and naming conventions, role-specific instructions, and lessons learned from early work. These documents should be living artifacts, updated as the project evolves.

Establish version control from the outset. Git or similar systems enable recovery from errors and provide audit trails for changes. The ability to recover from accidental deletions or corrupted files proved essential in this project and should be considered a baseline requirement for any extended work.

6.2 Workflow Design

Design workflows as pipelines with clear stages and handoff points. A typical translation pipeline might include: initial draft generation (primary translator model), quality assurance review (verification model), human review and decision-making, revision implementation, and documentation update (documentation model). Each stage should have defined inputs, outputs, and success criteria.

Build in explicit human oversight at decision points. AI models should inform and support human judgment, not replace it. The human practitioner makes final calls on disputed translations, resolves conflicts between model recommendations, and maintains responsibility for the scholarly integrity of outputs. This human-in-the-loop design is not merely an ethical preference but a practical necessity: models can be confidently wrong, and human judgment provides essential error correction.

Consider batch processing for efficiency. Rather than processing one passage at a time through the entire pipeline, it may be more efficient to generate drafts for a full chapter before submitting for QA review, then process all QA findings before implementing revisions. Batch processing reduces context-switching overhead and enables models to maintain consistency across related passages.

6.3 Quality Assurance Protocols

Develop structured QA templates appropriate to your project. For translation work, these might include: source fidelity checks (are all source elements represented?), technical term verification (are specialized terms rendered correctly and consistently?), stylistic consistency (does the output match established voice and register?), and doctrinal or conceptual accuracy (for religious, philosophical, or technical texts).

QA reports should produce actionable outputs: specific passages requiring attention, concrete recommendations for revision, clear scoring rubrics that enable comparison across versions. The scoring system need not be precise—translation quality is not reducible to numbers—but it should provide useful comparative signals.

Distinguish between error types that require different responses. Ghost characters (missing content) typically require immediate correction. Over-translation (added content) may be acceptable depending on project goals—pedagogical expansions may serve the target audience even if they depart from strict source fidelity. Technical term errors demand correction; stylistic variations may be matters of taste.

6.4 Iterative Refinement

Treat the methodology itself as subject to revision. Document what works and what doesn't. Update onboarding materials to reflect lessons learned. Be prepared to reassign roles as model capabilities evolve or as project requirements change. The framework should be a living system, not a fixed protocol.

Schedule periodic reviews to assess whether the workflow remains appropriate. As models update their capabilities, role assignments may need revision. As the project

matures, early-stage protocols may become unnecessary while new challenges emerge. Continuous improvement should be built into the methodology from the start.

6.5 Adapting the Framework

The framework described here emerged from a specific project context—Buddhist sutra translation with dual scholarly and vernacular outputs. Practitioners adapting this framework for other projects should consider how their requirements differ and adjust accordingly.

For projects without source language verification needs, the QA model role may be reduced or eliminated. For projects without stylistic complexity, the primary translator model may be assigned based on efficiency rather than interpretive nuance. For projects with different output formats—code, analysis, visualization—role assignments should reflect those specific requirements.

The key principle is not the specific role assignments documented here but the underlying insight: different models have different strengths, and orchestrating multiple models to leverage those differences can produce superior results. How that orchestration should be structured depends on the specific project, available models, and practitioner capabilities.

Practitioners new to multi-model workflows might begin with a simplified two-model approach: one model for primary content generation, another for review and verification. As experience accumulates, additional models can be incorporated for specialized functions. The framework scales up as practitioner capability and project complexity warrant.

7. Conclusion

This paper has documented a multi-model methodology for AI-assisted translation developed through the extended project of rendering the Lotus Sutra into both scholarly and vernacular English editions. The core insight is straightforward but consequential: different AI models exhibit different strengths, and thoughtful orchestration of multiple models can produce results superior to any single model working alone. Claude excels at interpretive nuance and stylistic voice; DeepSeek at linguistic verification against Sanskrit and Chinese sources; Gemini at documentation and infrastructure management. By routing tasks to models based on demonstrated capabilities, the practitioner transforms individual model limitations into collaborative strengths.

The methodology emerged through practice—through trial and error, through failures and recoveries, through the gradual accumulation of working knowledge. This origin story is itself significant. The tools available today enable individuals without institutional backing or formal credentials to undertake serious scholarly projects. The barriers are not technical but motivational: willingness to learn, persistence through difficulty, and genuine engagement with the material. The democratizing potential of AI-assisted scholarship deserves serious attention from humanists, even as concerns about quality control, equity of access, and labor relations warrant ongoing vigilance.

The Lotus Sutra teaches that all beings possess Buddha-nature—the capacity for awakening inherent in every sentient being regardless of circumstance. Perhaps there is an analogy in the current moment of AI development: all practitioners possess the capacity to develop sophisticated human-AI collaborations suited to their own projects and purposes. The framework proposed here is not a prescription but an invitation: observe what works, document your learning, share your discoveries, and remain open to continued refinement as the technology and your own understanding evolve.

Future work might include systematic comparison of multi-model and single-model approaches on standardized translation tasks, exploration of multi-model orchestration in other humanistic domains, development of more sophisticated QA metrics that go beyond numerical scoring, and study of how different practitioners adapt these methods to their own contexts. The field is young and rapidly evolving; there is much to learn and much to contribute.

Several specific research directions merit attention. First, controlled experiments comparing multi-model versus single-model performance on standardized translation tasks would enable more rigorous evaluation of the methodology's effectiveness. Such studies might employ panels of expert reviewers evaluating blind samples, measuring both accuracy and stylistic quality across different approaches. Second, exploration of multi-model orchestration in domains beyond translation—literary analysis, historical research, philosophical interpretation, archival work—could reveal whether the principles identified here generalize across humanistic disciplines or remain specific to translation contexts.

Third, longitudinal studies tracking how practitioners develop and refine multi-model workflows over extended projects could illuminate the learning curves involved, identify common pitfalls and best practices, and generate pedagogical resources for newcomers to this approach. Fourth, investigation of how multi-model workflows might be institutionalized within academic settings—through dedicated platforms, standardized protocols, or collaborative infrastructure—could help translate individual practitioner innovations into broader scholarly resources.

The question of how AI-assisted scholarship should be evaluated and credentialed within academic institutions remains particularly pressing. Traditional peer review assumes human authorship; when AI systems contribute substantially to research outputs, how should reviewers assess quality and originality? Should AI-assisted work be held to different standards, labeled differently, or evaluated through new mechanisms? These institutional questions will require attention from scholarly communities, journal editors, and academic administrators as AI-assisted research becomes more prevalent.

Finally, the ethical dimensions of AI-assisted scholarship warrant continued examination. Questions of intellectual property, attribution, and the labor conditions of AI training remain contested. As practitioners develop sophisticated human-AI collaborations, they bear responsibility for engaging thoughtfully with these broader concerns, supporting equitable AI development, and contributing to scholarly norms that serve the interests of human communities alongside technological progress. The methodology documented in

this paper represents one attempt to navigate these complexities—imperfect, certainly, but offered in the spirit of open inquiry and shared learning that characterizes the best humanistic scholarship.

The Lotus Sutra, composed two millennia ago in the Sanskrit of ancient India, translated into Classical Chinese by Kumārajīva in fifth-century Chang'an, and now rendered into twenty-first-century English through collaboration between human and artificial intelligence, continues to find new forms of expression. The blues gospel interpretation joins a long lineage of cultural transpositions that have carried the Dharma across boundaries of language, culture, and time. If these AI-assisted translations help even one reader encounter the text in a way that speaks to them, the project will have served its purpose. The Dharma, after all, is meant to be shared.

References

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.

Brown, Tom B., et al. 2020. "Language Models are Few-Shot Learners." *Advances in Neural Information Processing Systems* 33: 1877–1901.

Bowker, Lynne. 2002. *Computer-Aided Translation Technology: A Practical Introduction*. Ottawa: University of Ottawa Press.

Hendy, Amr, et al. 2023. "How Good Are GPT Models at Machine Translation? A Comprehensive Evaluation." *arXiv preprint arXiv:2302.09210v2*. Accessed December 15, 2025.

Jiao, Wenxiang, et al. 2023. "Is ChatGPT a Good Translator? A Preliminary Study." *arXiv preprint arXiv:2301.08745v2*. Accessed December 15, 2025.

Kocmi, Tom, and Christian Federmann. 2023. "Large Language Models Are State-of-the-Art Evaluators of Translation Quality." *arXiv preprint arXiv:2302.14520*.

O'Brien, Sharon. 2012. "Translation as Human-Computer Interaction." *Translation Spaces* 1: 101–122.

Park, Joon Sung, et al. 2023. "Generative Agents: Interactive Simulacra of Human Behavior." *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*.

Wang, Jianing, et al. 2024. "LLM-Blender: Ensembling Large Language Models with Pairwise Ranking and Generative Fusion." *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*.

Wooldridge, Michael, and Nicholas R. Jennings. 1995. "Intelligent Agents: Theory and Practice." *The Knowledge Engineering Review* 10(2): 115–152.

Xu, Minghao, et al. 2024. "TransAgents: A Multi-Agent System for Literary Translation." *arXiv preprint arXiv:2405.11804v1*. Accessed December 15, 2025.

Hurvitz, Leon. 1976. *Scripture of the Lotus Blossom of the Fine Dharma*. New York: Columbia University Press.

Kern, Hendrik. 1884. *Saddharma-Puṇḍarīka or The Lotus of the True Law*. Sacred Books of the East, Vol. 21. Oxford: Clarendon Press.

La Vallée Poussin, Louis de. 1907–1911. *Śikṣāsamuccaya: A Compendium of Buddhist Doctrine*. St. Petersburg: Bibliotheca Buddhica.

Moretti, Franco. 2013. *Distant Reading*. London: Verso.

Nida, Eugene A. 1964. *Toward a Science of Translating*. Leiden: Brill. (See especially pp. 156–192 on formal versus dynamic equivalence.)

Reeves, Gene. 2008. *The Lotus Sutra: A Contemporary Translation of a Buddhist Classic*. Boston: Wisdom Publications.

Underwood, Ted. 2019. *Distant Horizons: Digital Evidence and Literary Change*. Chicago: University of Chicago Press.

Vaswani, Ashish, et al. 2017. "Attention Is All You Need." *Advances in Neural Information Processing Systems* 30: 5998–6008.

Venuti, Lawrence. 1995. *The Translator's Invisibility: A History of Translation*. London: Routledge. (See especially pp. 17–29 on domestication versus foreignization.)

Watson, Burton. 1993. *The Lotus Sutra*. New York: Columbia University Press. (Chapters 1–28.)

Data Availability

The translation outputs described in this paper (scholarly and vernacular editions of the Lotus Sutra) are available from the author upon request. Source texts consulted include Kumārajīva's Classical Chinese translation (T0262, Taishō Tripiṭaka) and Hendrik Kern's 1884 English translation from Sanskrit, both publicly available. QA audit reports and CLAUDE.md documentation templates are available upon request. Correspondence: [Email Redacted]

AI Disclosure Statement

This manuscript was prepared with substantial assistance from multiple AI systems. Claude (Anthropic, versions 3.5 Sonnet and Claude 4) was used for primary translation work, interpretive guidance, and manuscript drafting. DeepSeek (DeepSeek-R1) was used for quality assurance audits comparing translations against Classical Chinese and Sanskrit source texts. Gemini (Google, version 2.0) was used for documentation management and infrastructure support. GPT-4 (OpenAI) was used for research

assistance and cross-checking. All AI-generated content was reviewed, verified, and approved by the author, who maintains full responsibility for the accuracy and integrity of the work. The methodology described in this paper explicitly addresses the use of AI in scholarly translation and the protocols developed to ensure human oversight.

Acknowledgments

Thanks to the developers of Antigravity (Google Labs) and the Model Context Protocol for enabling the multi-model workflow infrastructure that made this project possible. The Buddhist studies community, whose existing translations and scholarship provided the foundation upon which this work builds, deserves acknowledgment as well. Any errors or shortcomings remain my sole responsibility.

Appendix: Sample QA Audit Report

The following excerpt from a DeepSeek QA audit report illustrates the systematic verification process described in this paper. This audit evaluated Chapter 1 of the Blues Lotus Sutra against Kumārajīva's Classical Chinese source text (T0262).

Executive Summary: The Blues translation is a highly interpretative, narrative-driven rendition. Its primary characteristics are: extensive over-translation for accessibility and rhetorical impact, minimal ghost characters (core content preserved), and proper handling of proper nouns with standard Kumārajīva-based transliterations.

Overall Confidence Score: 6.8/10

Part-by-Part Breakdown:

Part 1 (The Arhats): 6.9/10 — Core content preserved; significant vernacular expansion

Part 2 (Bodhisattvas and Divine Assembly): 5.9/10 — Extensive explanatory glossing; biographical details added

Part 3 (The Teaching and Light): 7.8/10 — Closest fidelity to source; minimal additions

Sample Sentence Analysis:

Chinese: 心得自在 ("minds attained freedom")

Blues: "their minds were finally free—free like a bird that's been caged all its life and somebody finally opens the door"

Assessment: Over-translation (simile added); core meaning preserved; confidence 5/10

Assessment: The translation is theologically sound and liturgically powerful, but not suitable for direct philological citation. It excels as a pastoral or liturgical translation for communities seeking an accessible entry to the Lotus Sutra.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Complete QA audit reports for all chapters, sample CLAUDE.md documentation templates, workflow diagrams, and extended translation comparisons are available upon request. Correspondence: [Email Redacted]

SOURCE TEXT
Classical Chinese / Sanskrit

HUMAN ORCHESTRATOR

CLAUDE (ANTHROPIC)
INTERPRETIVE TRANSLATION
Scholarly + Blues Gospel

DEEPSEEK
QUALITY ASSURANCE
Source Verification

GEMINI (GOOGLE)
DOCUMENTATION
Infrastructure Management

Verification Loop

FINAL TRANSLATIONS
SCHOLARLY EDITION   VERNACULAR EDITION