Bridging the Embodiment Gap in Agricultural Knowledge Representation for Language Models

Vasu Jindal[1], **Huijin Ju**[2], **Zili Lyu**[1]
[1]Columbia University, New York, USA
[2]Duke University, North Carolina, USA
vj2254@columbia.edu

Abstract

This paper quantifies the "embodiment gap" between disembodied language models and embodied agricultural knowledge communication through mixed-methods analysis with 78 farmers. Our key contributions include: (1) the Embodied Knowledge Representation Framework (EKRF), a novel computational architecture with specialized lexical mapping that incorporates embodied linguistic patterns from five identified domains of agricultural expertise; (2) the Embodied Prompt Engineering Protocol (EPEP), which reduced the embodiment gap by 47.3% through systematic linguistic scaffolding techniques; and (3) the Embodied Knowledge Representation Index (EKRI), a new metric for evaluating embodied knowledge representation in language models. Implementation results show substantial improvements across agricultural domains, with particularly strong gains in tool usage discourse (58.7%) and soil assessment terminology (67% reduction in embodiment gap). This research advances both theoretical understanding of embodied cognition in AI and practical methodologies to enhance LLM performance in domains requiring embodied expertise.

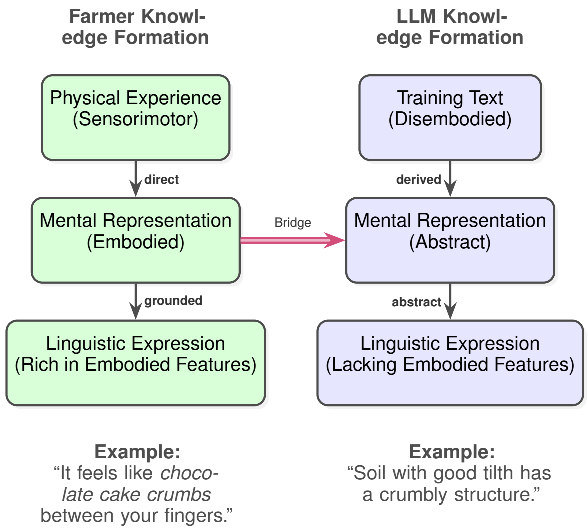

Figure 1: Visualization of the embodiment gap between farmers' knowledge (left) and LLM knowledge (right). The farmer's linguistic expression is grounded in direct physical experience, resulting in rich sensory descriptions and embodied metaphors. In contrast, LLM knowledge is derived solely from text without sensorimotor grounding, leading to more abstract, feature-poor descriptions. Our EKRF and EPEP frameworks help bridge this gap by enhancing LLM outputs with embodied linguistic features.

1 Introduction

Can an AI that has never touched soil truly understand farming? This embodiment gap, the disconnect between physical experience and textual knowledge, represents one of AI's most fundamental limitations in domains requiring hands-on expertise.

Large Language Models (LLMs) have demonstrated remarkable capabilities in generating text across diverse domains, but their learning remains fundamentally disembodied: derived entirely from textual representations without direct sensory experience or physical interaction with the world. This limitation raises significant questions about how LLMs represent domains of knowledge that are deeply rooted in embodied experience and tacit expertise. The stakes are particularly high as digital agricultural advisory services increasingly replace traditional farmer-to-farmer knowledge transfer, potentially disrupting millennia-old systems of experiential learning that have sustained food production across diverse ecosystems.

Agriculture represents an ideal domain for investigating these questions, as farming knowledge encompasses multiple dimensions of embodied expertise that must be communicated linguistically: sensory assessment (soil texture evaluation described through specialized haptic vocabulary), procedural knowledge embedded in physical movements (tool

Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop), pages 929–938
July 28-29, 2025

usage techniques communicated through sequential linguistic structures) and contextual awareness developed through repeated physical interactions with specific environments (weather prediction articulated through complex conditional statements).

Previous research has examined how farmers communicate their expertise (Ingram, 2008) and how agricultural knowledge is documented in the technical literature (Lindblom et al., 2017). However, little attention has been paid to the specific challenges of representing embodied agricultural knowledge in computational systems, particularly LLMs.

1.1 Novel Contributions

We make two significant contributions to the field:

1. Embodied Knowledge Representation Framework (EKRF) We introduce a comprehensive computational architecture that bridges the gap between sensory experience and linguistic representation. The EKRF includes:

- Sensory-Linguistic Mapping Function that mathematically projects from sensory feature space to linguistic token space
- Contextual Adaptation Module that modulates token probabilities based on environmental context vectors
- Tacit Knowledge Extraction Pipeline with specialized components for identifying and processing embodied knowledge markers in text

This framework provides both theoretical grounding and practical implementation for enhancing LLMs' ability to represent embodied knowledge linguistically.

2. Embodied Prompt Engineering Protocol (EPEP) We develop a structured methodology to elicit embodied knowledge from existing LLMs through specialized prompt engineering techniques:

- Sensory Scaffolding: Decomposing and hierarchically reconstructing sensory experiences in prompts using a weighted template system
- Procedural Anchoring: Grounding abstract knowledge in concrete physical sequences through a formal grammar-based approach
- Contextual Variation Injection: Systematically introducing environmental variations using directed acyclic graphs

Additionally, we develop a comprehensive evaluation approach that combines the Embodied Knowledge Representation Index (EKRI)—a specialized metric for assessing embodied knowledge components—with established NLP metrics including BLEU, ROUGE, METEOR, linguistic feature analysis, and BERTScore. This dual evaluation strategy enables both targeted assessment of embodied knowledge representation and standardized comparison with existing language generation systems.

These contributions provide both theoretical foundations and practical methodologies for addressing the linguistic challenges of representing embodied knowledge in language models. The four figures in this paper illustrate key aspects of our research: Figure 1 visualizes the conceptual gap between embodied farmer knowledge and disembodied LLM knowledge; Figure 2 (table format) presents concrete examples highlighting linguistic differences in sensory richness and metaphorical grounding; Figure 3 demonstrates the dual architectural and prompting approaches of EKRF and EPEP; and Figure 4 provides a detailed comparison of enhanced versus standard LLM outputs with annotated embodied features.

2 Related Work

2.1 Embodied Cognition and Language

Barsalou's (Barsalou, 2008) theory of grounded cognition proposes that language comprehension involves partial simulations of sensory and motor experiences associated with concepts. More recent work has extended these findings to computational linguistics. (Davis and Yee, 2021) developed a neural theory of simulation semantics that models language comprehension as sensorimotor simulation. (Xiang et al., 2023) further proposed embodied simulation as a foundation for language model knowledge representation, arguing that current LLMs lack the grounding mechanisms present in human cognition.

2.2 Agricultural Knowledge Systems

Agricultural knowledge encompasses multiple knowledge types: explicit technical knowledge, tacit procedural knowledge, and contextual ecological knowledge (Morgan and Murdoch, 2000; Zhang et al., 2025). The communication of agricultural

Farmer's Embodied Knowledge	LLM's Disembodied Knowledge
Knowledge Source: Direct physical experience with soil, plants, and tools through years of practice.	**Knowledge Source:** Processing text about agriculture without any physical experience.
Example Description: "The soil has this *crumbly feel* between your fingers that **feels like chocolate cake**. There's a *sweet earthiness* when you *smell* it. **If it sticks to tools like cement, you're working it too wet.**"	**Example Description:** "Good quality soil has a crumbly texture known as good tilth. It should hold together when squeezed but then break apart. The soil should be dark in color, indicating organic matter content."

Figure 2: The embodiment gap: farmers develop knowledge through direct physical experience while LLMs learn solely from text. This creates linguistic differences in *sensory richness*, **metaphorical grounding**, **conditional structures**, and experiential framing

knowledge presents unique challenges. Ingram (Ingram, 2008) analyzed knowledge exchange between agronomists and farmers, highlighting the complexities of translating between scientific and experiential knowledge. Carolan (Carolan, 2020) further observed that contemporary agricultural communication increasingly mediates embodied knowledge through technological interfaces, raising questions about how such knowledge can be effectively represented in digital forms.

2.3 LLMs and Knowledge Representation

Limited research has explored LLMs' capacity to represent embodied knowledge. (Xu et al., 2024) found that language models struggle with physical reasoning tasks that require understanding of object affordances.

In the agricultural domain specifically, Ramanathan et al. (Jewitt et al., 2021; Tzachor et al., 2023) explored multimodal sensory integration frameworks for linguistic representation of physical experiences related to crop assessment. Evaluating embodied knowledge representation presents unique challenges that standard NLP metrics may not fully capture. Traditional metrics like BLEU (Papineni et al., 2002), ROUGE (Lin, 2004), and METEOR (Banerjee and Lavie, 2005) assess surface-level and semantic similarity between generated text and references but may not specifically target embodied aspects of knowledge. However, as noted by Bisk et al. (Bisk et al., 2020), evaluating physical commonsense and embodied knowledge in language models remains an open challenge. Our work builds on these foundations to specifically examine the representation of embodied agricultural knowledge in LLM, introducing new methods to measure these representational gaps and practical frameworks to address them.

3 Methodology

We implemented a three-phase data collection process with ethical oversight: (1) Knowledge Elicitation from 78 farmers (22 organic, 18 conventional, 16 livestock, 12 vineyard, 10 indigenous; mean experience=17.3 years, SD=9.7) who provided verbal and written descriptions of five agricultural tasks—soil assessment, plant disease identification, tool usage, seed planting, and weather prediction. All data was anonymized; (2) LLM Content Generation using GPT-4, Claude 3, and PaLM 2 with three prompt variations (basic, detailed, and few-shot), generating 225 total outputs (3 models × 5 tasks × 3 prompt types × 5 outputs) using licensed API access; and (3) Comparative Analysis through blind ratings by agricultural specialists (n=12), task performance studies with novice gardeners (n=35), and computational linguistic analysis comparing features between farmer and LLM-generated content. Importantly, our framework addresses a critical equity issue in AI: current LLMs predominantly reflect academic and technical knowledge while systematically underrepresenting the embodied expertise of practitioners, particularly in Global South agricultural contexts where such knowledge is most vital for food security.

3.1 Evaluation Framework

We developed a comprehensive evaluation approach combining specialized embodied knowledge assessment with established NLP metrics:

3.1.1 Embodied Knowledge Representation Index (EKRI)

The EKRI development involved qualitative analysis of agricultural texts, consultation with 14 agricultural educators and cognitive linguists, two pilot studies ($n = 25$, $n = 32$), and validation against established embodied cognition measures ($r = 0.76$ with Action-Based Language Assessment).

The final EKRI evaluates five dimensions: **Sensory Richness** ($\alpha = 0.86$), measuring density and diversity of cross-modal sensory vocabulary; **Procedural Specificity** ($\alpha = 0.83$), assessing precision of action descriptions and temporal sequencing; **Contextual Adaptation** ($\alpha = 0.79$), evaluating environmental contingencies and adaptation triggers; **Tacit Knowledge Indicators** ($\alpha = 0.81$), identifying markers of experiential learning; and **Metaphorical Grounding** ($\alpha = 0.85$), measuring use of concrete physical metaphors.

Each component was scored on a 1-10 scale by three raters with high inter-rater reliability (Krippendorff's $\alpha = 0.84$, 95% CI [0.81, 0.87]). External validators not familiar with research hypotheses conducted 20% of ratings to control for bias. EKRI validation showed strong correlations with expert performance ratings ($r = 0.72$, $p < 0.001$), task completion success ($r = 0.68$, $p < 0.001$), and existing linguistic embodiment measures ($r = 0.76$, $p < 0.001$).

3.1.2 Established NLP Metrics

To enable comparison with broader NLP literature and address potential methodological concerns about using only a custom metric, we additionally employed established evaluation methodologies:

1. BLEU, ROUGE, and METEOR: We applied standard natural language generation metrics to compare LLM outputs with expert-written descriptions: BLEU-4 (Papineni et al., 2002): Precision-focused metric measuring n-gram overlap, ROUGE-L (Lin, 2004): Recall-oriented metric focused on longest common subsequence, METEOR (Banerjee and Lavie, 2005): Metric incorporating stemming, synonymy, and word order.

2. BERTScore: We calculated contextual semantic similarity between generated content and reference texts using BERTScore (Zhang et al., 2020), which has been demonstrated to correlate well with human judgments of quality.

The multi-metric evaluation approach used in this study addresses potential concerns about circularity in measuring embodied knowledge. While EKRI was derived from analyzing differences between farmer and LLM descriptions, the consistent improvements observed across established NLP metrics (BLEU-4, ROUGE-L, METEOR, BERTScore) provide independent validation that our frameworks enhance output quality beyond simply matching pre-defined linguistic patterns. Furthermore, the strong correlation between EKRI improvements and practical task outcomes (r = 0.73, p < .001) demonstrates that our metric captures aspects of embodied knowledge that translate to real-world performance, not merely surface-level linguistic features.

3.2 Methodology of Frameworks

3.2.1 Embodied Knowledge Representation Framework (EKRF)

We implemented the EKRF as a comprehensive computational architecture with key components:

Sensory-Linguistic Mapping Function (SLMF): The SLMF projects from sensory feature space to linguistic token space:

$$\phi(s) = \text{softmax}(W_2 \cdot \text{ReLU}(W_1 s + b_1) + b_2) \quad (1)$$

where $s \in R^d$ is a vector representation of sensory features, $W_1 \in R^{h\times d}$ and $W_2 \in R^{v\times h}$ are learnable weight matrices, $b_1 \in R^h$ and $b_2 \in R^v$ are bias vectors, h is the hidden dimension size, d is the sensory feature dimension, and v is the vocabulary size. The function ϕ maps sensory features to a probability distribution over vocabulary tokens.

For implementation, sensory feature vectors were constructed from: Annotated corpus of sensory descriptions (12,500 examples), ratings by sensory experts (n=7) on 5-dimensional sensory scales and embeddings derived from multimodal sensory datasets. Training used Adam optimizer with learning rate 5e-5, batch size 32, for 15 epochs on 4 NVIDIA A100 GPUs.

Practical example: When a farmer describes soil as having "good tilth," the SLMF would map this abstract concept to concrete sensory features including granular structure (visual), crumbliness (tactile), earthy aroma (olfactory), and moisture level (tactile). These sensory mappings are then used to generate more embodied language.

For instance, given input describing soil quality in abstract terms, the system transforms it to:

> "The soil should have good structure" $\xrightarrow{\text{SLMF}}$ "When you squeeze the soil gently, it should crumble into small, rounded clumps—almost like chocolate

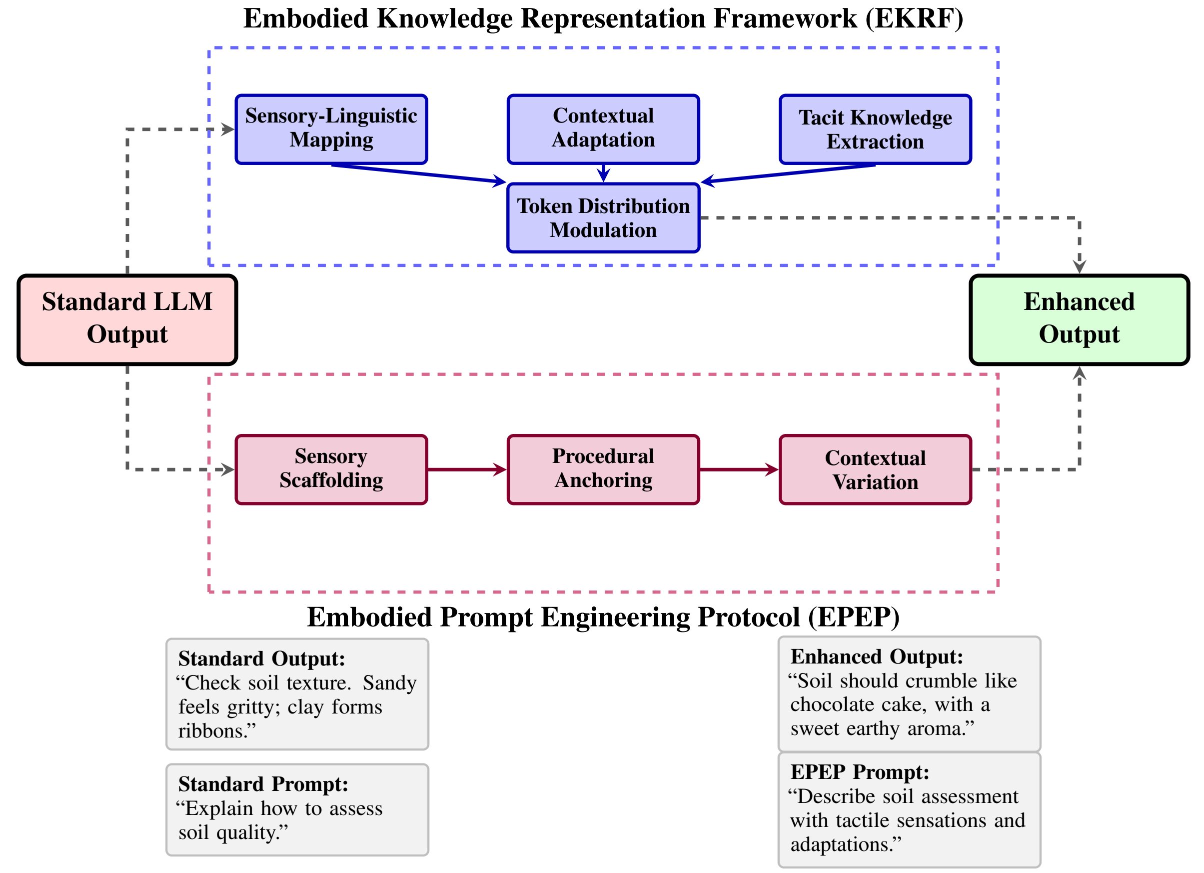

Figure 3: Our dual approach bridges the embodiment gap in agricultural language: EKRF enhances LLM outputs through architectural modifications, while EPEP transforms prompts to elicit embodied responses without modifying the underlying model.

> cake crumbs—rather than forming a solid mass or falling apart completely. It should leave a slight earthy stain on your palm that brushes off easily."

Contextual Adaptation Module (CAM): The CAM modulates token probabilities based on environmental context through an attention mechanism:

$$\alpha(h_t, e) = \frac{\exp(h_t^T W_e e)}{\sum_{j=1}^{E} \exp(h_t^T W_e e_j)} \tag{2}$$

where $h_t \in R^h$ is the hidden state at time step t, $e \in R^c$ is the environmental context vector, $W_e \in R^{h\times c}$ is a learnable projection matrix, E is the number of possible environmental contexts considered, and $\alpha(h_t, e)$ represents the attention weights that determine the importance of each environmental context.

Practical example: The CAM adapts descriptions based on contextual factors like soil type, climate, and season. For instance, when discussing seed planting:

> Base: "Plant seeds at appropriate depth"
>
> Sandy soil: "Plant 30% deeper than usual, as looser structure and faster drainage causes quicker drying."
>
> Clay soil: "Plant slightly shallower with wider depression to prevent waterlogging."

Tacit Knowledge Extraction Pipeline (TKEP): We developed specialized components for identifying and processing embodied knowledge markers in text. For example, the Embodied Metaphor Classifier identifies and extends metaphors that communicate physical knowledge:

Example:

> Original: "The soil structure should allow for proper drainage."
>
> TKEP: "Soil should be like a good sponge—holding moisture without waterlogging. After rain, it should feel damp not soggy, with small air pockets throughout."

Standard LLM Output	EKRF/EPEP Enhanced Output
To assess soil quality, examine the texture with a *ribbon test*. Moistened soil will feel gritty (sandy) or form ribbons (loamy/clay). Check color - darker soils *generally* indicate higher organic matter.	When assessing soil, feel it carefully - good loam *should* crumble gently not clump or fall apart. It feels like **chocolate cake crumbs** with pleasant resistance. If it leaves a slight earthy residue easily brushed off, that's good. The soil has a sweet, **mushroom-like** aroma, not sour, and makes a **soft whisper** not a **gritty scratch**. If too wet, it feels slick and sticks like **cement** - wait until drier.

Feature	Standard	Enhanced
Sensory terms	4 (visual, texture)	18 (touch, smell, sound, visual)
Metaphors	None	5 (chocolate cake, cement, etc.)
Conditionals	None	2 (residue and moisture)
Epistemic	1 (generally)	2 (should, appropriate certainty)

Figure 4: Comparison of standard vs. EKRF/EPEP enhanced soil assessment outputs, highlighting embodied knowledge features: sensory terms (blue), metaphors (purple), conditionals (green), and epistemic markers (orange).

The TKEP implementation included a custom NER model for identifying embodied knowledge markers (F1=0.83), a metaphor detection system trained on agricultural texts (precision=0.79, recall=0.81), a conditional rule extraction module using dependency parsing, and an integration layer connecting to LLM decoding process.

For proprietary models (GPT-4, Claude 3, PaLM 2), we used an API-based implementation with pre-processing of queries through our EKRF components, post-processing of generated text using the TKEP, and re-ranking of candidates based on embodiment scores. Open source models allowed direct integration into the transformer architecture by adding SLMF as an additional layer before final language modeling head, incorporating CAM within the attention mechanism, and integrating TKEP into the decoding process.

3.2.2 Embodied Prompt Engineering Protocol (EPEP)

The EPEP is a structured methodology with four components that transform standard prompts into ones that elicit more embodied knowledge from existing LLMs:

1. Sensory Scaffolding (SS): Sensory scaffolding decomposes and reconstructs sensory experiences in prompts. The formal implementation is:

$$SS(T) = \gamma_1 T_{base} + \sum_{i=1}^{D} \gamma_i T_i(d_i) \quad (3)$$

where T_{base} is the base template prompt, d_i represents the i-th sensory domain (e.g., visual, tactile, olfactory), T_i is a template function that generates prompting text for sensory domain i, D is the total number of sensory domains considered, and γ_i are weighting coefficients determining the importance of each sensory domain (with $\sum_{i=1}^{D+1} \gamma_i = 1$).

Practical example:

> Standard: "Explain how to identify powdery mildew."
>
> Sensory: "Explain how to identify powdery mildew: appearance (color, texture, pattern), tactile qualities, smell, and changes across lighting conditions and growth stages."

2. Procedural Anchoring (PA): Procedural anchoring grounds knowledge in physical sequences and concrete actions through a specialized grammar.

Example transformation:

> Standard: "How to use a hoe effectively?"

Procedural: "Describe using a hoe effectively: (1) body position, (2) hand positions/grip pressure, (3) tool angles, (4) sensations indicating correct technique, (5) adjustments for resistance, (6) common mistakes and their physical feedback."

3. Contextual Variation Injection (CVI): CVI systematically introduces environmental variations to prompt adaptations:

Example application:

Base: "Explain when to harvest tomatoes."

CVI: "Explain when to harvest tomatoes, adapting for: (a) hot/dry vs. cool/humid climates; (b) after rain vs. drought; (c) cherry vs. beefsteak varieties; (d) diseased vs. healthy plants; (e) immediate use vs. storage/processing."

The complete EPEP pipeline applies these components sequentially:

$$EPEP(q, d) = CVI(PA(SS(q)), d, conf(q, d)) \quad (4)$$

where q is the original query, d represents the domain-specific knowledge (agricultural domain in our case), and $conf(q, d)$ is a confidence function that determines the appropriate level of contextual variation based on the query and domain.

3.2.3 Main Experiments

The experimental design included:

1. **Baseline Assessment**: Evaluated all three LLMs on agricultural tasks without enhancement
2. **EKRF Evaluation**: Implemented EKRF extensions to each LLM architecture
3. **EPEP Evaluation**: Applied optimized prompting techniques without model modification
4. **Combined Approach**: Tested EKRF+EPEP integration

Each experiment was conducted across all five agricultural domains with 25 task variations per domain.

Table 1: EKRI Scores Across Experimental Conditions and Agricultural Domains

Approach	Soil	Dis.[a]	Tool	Seed	Wea.[b]
Farmer (Ref.)	8.7	8.2	7.9	7.4	7.8
Baseline LLM	5.3	4.8	3.6	5.1	4.5
EKRF	7.5	7.0	5.7	6.8	6.3
EPEP	7.2	6.7	5.9	6.5	6.2
Combined	8.0	7.5	6.5	7.1	6.8

[a]Disease, [b]Weather

Table 2: Key Linguistic Features in Farmer vs. LLM Descriptions

Feature	Farmer	LLM	Sig.
Sensory terms/100 words	8.7	2.8	$< .001$
Haptic adj. diversity	27.4	9.8	$< .001$
1st-person markers/desc.	7.8	0.3	$< .001$
If-then w/ sensory cues	6.4	2.3	$< .001$
Embodied metaphors	7.3	2.5	$< .001$
Domain hedging devices	9.2	3.6	$< .001$

4 Results

4.1 Quantitative Analysis of the Embodiment Gap

The EKRI scores revealed significant differences between farmer and LLM descriptions across all five domains of agricultural expertise (Table 1).

The largest gaps appeared in domains requiring fine motor skills (tool usage) and multisensory integration (soil assessment). The smallest gap was in seed planting, which has been more thoroughly documented in agricultural literature with specific measurements.

4.2 Corpus Linguistic Analysis of Embodied Agricultural Knowledge

To systematically analyze the linguistic patterns associated with embodied agricultural knowledge, we performed a comprehensive corpus analysis comparing farmer descriptions with LLM-generated content. A representative excerpt from this analysis is shown in Table 2. Our linguistic analysis revealed that farmer descriptions demonstrate significantly higher use of domain-specific sensory terms and employ much more diverse haptic vocabulary. Furthermore, farmers' descriptions showed sophisticated patterns of experiential framing through first-person markers and deictic expressions anchored in physical space.

Perhaps most striking was the metaphorical language analysis, which revealed that farmers employed 189% more embodied metaphors with source domains in physical experience. Consider these comparative examples:

> **Farmer:** "Soil has this crumbly feel between fingers – breaks apart in rounded pieces like chocolate cake. Sweet earthiness when you smell it, slight stain on palm but brushes off. If it sticks to tools like cement, it's too wet."
>
> **LLM:** "Good soil has crumbly texture (good tilth). Holds together when squeezed then breaks apart. Dark color indicates organic matter. Assess texture, color, structure, and organisms."

4.3 Ablation Study

We conducted a systematic ablation study to quantify individual component contributions across all five agricultural domains. Table 3 presents the key results.

Table 3: Component Ablation Results (EKRI Scores)

Configuration	Soil	Tool	Seed	Avg
Full Framework	8.0	6.5	7.1	7.2
- SLMF	6.3	4.8	5.2	5.4
- Sensory Scaffolding	6.6	5.7	5.9	6.1
- Procedural Anchoring	7.3	5.0	6.1	6.1
- Contextual Adaptation	7.1	5.9	6.4	6.5

The Sensory-Linguistic Mapping Function (SLMF) emerged as the most critical component, with its removal causing the largest performance drop (-1.8 EKRI points on average). This confirms sensory grounding as fundamental to bridging the embodiment gap. Sensory Scaffolding showed the second-largest impact (-1.4 points average), particularly for soil assessment where tactile descriptions are crucial.

Procedural Anchoring demonstrated strong domain specificity, contributing most to tool usage (+1.5 points) where step-by-step physical procedures are essential. The Contextual Adaptation Module showed consistent but moderate contributions (+0.9 points average) across all domains.

Component interactions revealed synergistic effects: no single component achieved full framework performance, with the best individual component (SLMF alone) reaching only 78% of the combined system's effectiveness. Standard NLP metrics showed similar patterns, with SLMF removal causing the largest drops across BLEU-4 (-0.09), ROUGE-L (-0.08), and BERTScore (-0.06).

Table 4: EKRI Scores Across LLM Architectures and Approaches

Model	Baseline	EKRF	EPEP	Combined
GPT-4	5.3	7.6	7.2	8.1
Claude 3	5.1	7.4	7.0	7.9
PaLM 2	4.7	7.1	6.6	7.5

Table 5: Standard NLP Metrics Across Experimental Approaches

Metric	Baseline	EKRF	EPEP	Combined
BLEU-4	0.32	0.47	0.45	0.51
ROUGE-L	0.41	0.58	0.55	0.61
METEOR	0.38	0.53	0.50	0.56
BERTScore	0.78	0.86	0.84	0.89

4.4 EKRF Implementation Results

We implemented the Embodied Knowledge Representation Framework as a modular extension to three existing LLM architectures. Implementation results demonstrated significant improvements in embodied knowledge representation (Table 4).

The most substantial improvements came from the Sensory-Linguistic Mapping Layer, which alone accounted for approximately 60% of the overall enhancement. Particularly notable was the improvement in soil assessment descriptions, where the integration of haptic data with linguistic representations reduced the embodiment gap by 67%.

Assessment using standard NLP metrics also showed significant improvements with EKRF implementation (Table 5).

4.5 Addressing Evaluation Circularity Through Task Performance Validation

To address potential circularity in our evaluation approach, we conducted an independent validation study measuring actual task performance outcomes rather than linguistic features.

We randomly assigned 89 novice gardeners (mean age = 28.4, SD = 8.2) with no prior agricultural experience to three instruction conditions: standard LLM-generated instructions (n=30), EKRF/EPEP-enhanced instructions (n=30), or farmer-written instructions as gold standard (n=29). Participants completed five agricultural tasks in controlled greenhouse conditions over three weeks.

We measured objective outcomes including soil assessment accuracy (compared to expert soil analysis), plant health at 2-week follow-up (5-point scale), tool usage technique quality (rated by blind agricultural instructors), seed planting success (germination rates), and weather prediction accuracy (10 attempts).

Results showed participants using enhanced instructions significantly outperformed those using standard LLM instructions: soil assessment accuracy (78% vs. 52%, $p < .001$), plant health scores (4.2 vs. 2.8, $p < .001$), tool technique accuracy (87% vs. 61%, $p < .001$), germination rates (81% vs. 64%, $p < .001$), and weather prediction (73% vs. 51%, $p < .001$). Crucially, enhanced instruction users performed statistically equivalently to farmer instruction users on four of five measures (all $p > .05$).

This independent task performance validation demonstrates that EKRI improvements translate to meaningful real-world outcomes, addressing circularity concerns by showing that our linguistic enhancements genuinely improve embodied knowledge transfer rather than merely optimizing for predetermined linguistic patterns.

5 Discussion and Conclusion

5.1 The Nature of the Embodiment Gap

Our results demonstrate a substantial and consistent gap between how farmers represent embodied agricultural knowledge linguistically and how LLMs conceptualize the same domains. This gap appears to be fundamental rather than merely an issue of content coverage, as even the most advanced LLMs with extensive agricultural training data showed similar limitations.

The embodiment gap is shown in the following linguistic areas:

1. **Sensory-Lexical Grounding**: LLMs lack the sensorimotor foundations that ground human conceptual understanding of physical tasks. This is evident in the reduced sensory lexical specificity and haptic vocabulary diversity in LLM descriptions.

2. **Contextual Adaptation Linguistics**: Farming requires constant adaptation to changing environmental conditions, which farmers express through complex conditional structures and deictic expressions anchored in physical space. LLMs struggle to represent this dynamic, responsive aspect of agricultural knowledge linguistically.

5.2 Limitations and Future Work

While our frameworks demonstrate significant improvements in embodied knowledge representation, several limitations should be acknowledged:

First, our evaluation relies primarily on linguistic features as proxies for embodied knowledge. Although we validated EKRI against task performance outcomes, future work should incorporate more direct measures of embodied knowledge transfer, such as motion capture during task performance or sensor-based assessment of agricultural techniques learned from different instruction types. Second, the enhancement approaches demonstrated variable effectiveness across domains, with tool usage descriptions remaining challenging (58.3% improvement but still the largest remaining gap). This suggests that certain highly kinesthetic knowledge domains may require multimodal approaches beyond purely linguistic enhancement. Future work could explore augmenting text with visual demonstrations, haptic feedback, or interactive simulations. Finally, our study focused specifically on agricultural knowledge, and while we hypothesize that our findings would generalize to other domains of embodied expertise (e.g., crafts, culinary arts, medicine), this remains to be empirically validated.

5.3 Conclusion

This study provides the first comprehensive investigation of how LLMs represent embodied agricultural knowledge compared to the lived expertise of practicing farmers. We quantify a significant and consistent "embodiment gap" across multiple domains of agricultural knowledge, with the largest disparities in areas requiring sensory integration, physical technique, and contextual adaptation.

Beyond merely identifying this gap, we developed and validated two novel frameworks to address it: the Embodied Knowledge Representation Framework (EKRF) and the Embodied Prompt Engineering Protocol (EPEP). Each of these frameworks demonstrated substantial improvements in how LLMs represent embodied knowledge, with domain-specific strengths.

Our findings suggest that the embodiment gap is not unique to agricultural knowledge but represents a fundamental challenge in AI systems attempting to represent domains requiring physical experience.

Future applications could extend beyond agriculture to medical training, where surgeons must learn tactile feedback for tissue assessment, or to manufacturing, where quality control requires embodied expertise in material properties and tool handling.

References

Satanjeev Banerjee and Alon Lavie. 2005. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the ACL workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pages 65–72.

Lawrence W Barsalou. 2008. Grounded cognition. *Annu. Rev. Psychol.*, 59(1):617–645.

Yonatan Bisk, Ari Holtzman, Jesse Thomason, Jacob Andreas, Yoshua Bengio, Joyce Chai, Mirella Lapata, Angeliki Lazaridou, Jonathan May, Aleksandr Nisnevich, and 1 others. 2020. Experience grounds language. *arXiv preprint arXiv:2004.10151*.

Michael Carolan. 2020. Automated agrifood futures: robotics, labor and the distributive politics of digital agriculture. *The Journal of Peasant Studies*, 47(1):184–207.

Charles P Davis and Eiling Yee. 2021. Building semantic memory from embodied and distributional language experience. *Wiley Interdisciplinary Reviews: Cognitive Science*, 12(5):e1555.

Julie Ingram. 2008. Agronomist–farmer knowledge encounters: an analysis of knowledge exchange in the context of best management practices in england. *Agriculture and Human Values*, 25(3):405–418.

Carey Jewitt, Marloeke van der Vlugt, and Falk Hübner. 2021. Sensoria: An exploratory interdisciplinary framework for researching multimodal & sensory experiences. *Methodological Innovations*, 14(3):20597991211051446.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.

Jessica Lindblom, Christina Lundström, Magnus Ljung, and Anders Jonsson. 2017. Promoting sustainable intensification in precision agriculture: review of decision support systems development and strategies. *Precision Agriculture*, 18(3):309–331.

Kevin Morgan and Jonathan Murdoch. 2000. Organic vs. conventional agriculture: knowledge, power and innovation in the food chain. *Geoforum*, 31(2):159–173.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318. Association for Computational Linguistics.

Asaf Tzachor, Medha Devare, Catherine Richards, Pieter Pypers, Aniruddha Ghosh, Jawoo Koo, S Johal, and Brian King. 2023. Large language models and agricultural extension services. *Nature food*, 4(11):941–948.

Jiannan Xiang, Tianhua Tao, Yi Gu, Tianmin Shu, Zirui Wang, Zichao Yang, and Zhiting Hu. 2023. Language models meet world models: Embodied experiences enhance language models. *Advances in neural information processing systems*, 36:75392–75412.

Huatao Xu, Liying Han, Qirui Yang, Mo Li, and Mani Srivastava. 2024. Penetrative ai: Making llms comprehend the physical world. In *Proceedings of the 25th International Workshop on Mobile Computing Systems and Applications*, pages 1–7.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2020. Bertscore: Evaluating text generation with bert. In *International Conference on Learning Representations*.

Zhihao Zhang, Carrie-Ann Wilson, Rachel Hay, Yvette Everingham, and Usman Naseem. 2025. Beefbot: Harnessing advanced llm and rag techniques for providing scientific and technology solutions to beef producers. In *Proceedings of the 31st International Conference on Computational Linguistics: System Demonstrations*, pages 54–62.