AI Implicit A Foundational Paradigm For Intelligence Through Experience Compression

Momen Ghazouani
Chief Scientist, Setaleur Aplamda

April 20, 2026

Abstract

We introduce **AI Implicit** (Artificial Intelligence through Implicit Knowledge), a foundational research paradigm that redefines artificial intelligence through the lens of tacit knowledge acquisition, experience compression, and epistemic awareness. Unlike the dominant optimization paradigm that measures intelligence through task-specific performance, AI Implicit evaluates systems by their capacity to extract, compress, and transfer implicit knowledge across domains mirroring the fundamental cognitive processes that distinguish human intelligence from specialized computation. AI Implicit rests on three core principles: (1) **tacit knowledge extraction** as the primary measure of learning efficiency, (2) **experience compression** as the mechanism of intelligence, and (3) **epistemic confidence** as the foundation of robust reasoning. We argue that the path to artificial general intelligence requires abandoning performance-centric metrics in favor of knowledge-density measures that capture how efficiently systems learn from limited examples and how effectively they recognize the boundaries of their competence.

This manifesto establishes AI Implicit as a comprehensive research direction analogous to how deep learning emerged as an umbrella paradigm encompassing convolutional networks, transformers, and attention mechanisms. We outline the theoretical foundations, evaluation frameworks, and architectural requirements necessary to realize intelligence grounded in implicit knowledge rather than explicit optimization. Through rigorous analysis of current AI limitations and synthesis of cognitive science insights, we demonstrate that AI Implicit provides both a measurable alternative to existing AGI definitions and a practical roadmap for developing systems that learn, reason, and transfer knowledge with human-like efficiency.

Keywords: Implicit Intelligence, Tacit Knowledge, Experience Compression, Epistemic Confidence, Knowledge Transfer, Artificial General Intelligence

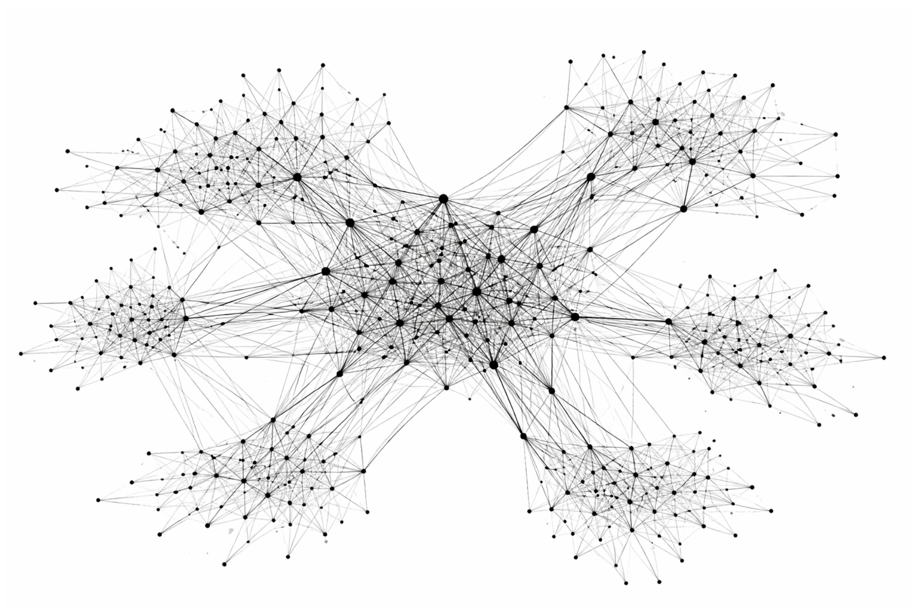

Figure 1 *A conceptual representation of implicit intelligence as a structured yet adaptive network of interconnected knowledge. The architecture reflects how experience is compressed, redistributed, and transferred across distinct domains. Clarity emerges from relational density, where complexity is organized rather than reduced*

1. Introduction: The Intelligence We Cannot Name

1.1 The Crisis of Definition

Artificial intelligence has achieved remarkable engineering successes surpassing human performance in image classification, game playing, and language modeling yet we lack a coherent answer to a fundamental question: **What is intelligence?**

Current approaches fall into three categories, each revealing different limitations :

Performance-Based Definitions measure intelligence through task achievement the Turing Test, chess mastery, economic productivity. These capture surface-level capability but cannot distinguish genuine understanding from sophisticated pattern matching. A system that memorizes every possible chess game achieves superhuman performance without demonstrating transferable intelligence.

Capability-Based Definitions enumerate specific competencies—reasoning, planning, learning, natural language understanding. These provide taxonomic clarity but offer no unified principle. Is a system that excels at 50 narrow tasks more intelligent than one that demonstrates deep transfer across 5 domains?

Computational Definitions ground intelligence in algorithmic efficiency sample complexity, computational cost, generalization bounds. These provide mathematical rigor but sidestep the qualitative question: efficiency *at what*? Optimizing prediction accuracy on fixed datasets may produce brittle systems that fail catastrophically on distribution shifts.

The absence of a principled definition has consequences beyond philosophical debate. **We cannot measure what we cannot define, and we cannot build what we cannot measure.** Without clarity on what constitutes intelligence, AI research becomes an ad hoc collection of benchmarks, each capturing a different facet while missing the essential whole.

1.2 The Implicit Knowledge Hypothesis

We propose a fundamental reframing: **Intelligence is the capacity to extract, compress, and transfer tacit knowledge.**

Tacit knowledge the implicit, unarticulated understanding that guides expert performance represents the essence of intelligence. A chess grandmaster cannot fully verbalize the pattern recognition underlying their intuition. A skilled surgeon cannot completely codify the sensorimotor coordination enabling precise incisions. A native speaker cannot enumerate all grammatical rules they unconsciously apply.

This implicit knowledge exhibits three critical properties:

1. **Compression**: Thousands of hours of experience are encoded into intuitive pattern recognition. A master pianist's muscle memory compresses years of practice into fluid performance.

2. **Transferability**: Implicit knowledge generalizes beyond its acquisition context. Chess expertise transfers to strategic planning; mathematical intuition aids scientific reasoning; language learning accelerates subsequent language acquisition.

3. **Epistemic Awareness**: Experts know the boundaries of their knowledge. A physician recognizes when a case exceeds their experience; a mathematician identifies when a proof strategy won't work. This metacognitive capacity *knowing when you don't know* is as fundamental as knowledge itself.

Human intelligence is fundamentally **implicit intelligence**: the ability to compress experience into reusable cognitive structures, transfer them across domains, and recognize their limitations. AI systems lacking these properties no matter their task performance miss the essence of intelligence.

1.3 The Optimization Paradigm's Limits

Modern AI rests on a single architectural principle: **gradient-based optimization**. From perceptrons to transformers, systems learn by minimizing differentiable loss functions through iterative parameter updates :

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta \mathcal{L}(\theta_t)$$

This paradigm has enabled unprecedented scaling billion-parameter language models, zero-shot image generation, superhuman game playing yet exhibits three fundamental pathologies that question its viability as a path to general intelligence:

The Correlation-Causation Conflation: Gradient descent discovers *any* statistical regularity reducing loss, making no principled distinction between meaningful patterns and spurious correlations. A model learns that "asthma patients have lower mortality" not because asthma is protective, but because hospitals monitor such patients intensively. The optimization process is causally blind it maximizes predictive accuracy without understanding the mechanisms generating data.

The Transfer Catastrophe: Knowledge acquired in one domain degrades catastrophically when applied to others. Networks trained on ImageNet struggle with simple style shifts despite perfect in-distribution accuracy. Language models fine-tuned on programming fail at creative writing. This brittleness contradicts the defining property of intelligence: the ability to transfer knowledge across contexts.

The Epistemic Opacity Problem: Neural networks produce confident predictions on inputs far outside their training distribution. A model trained on handwritten digits assigns 99.9% confidence to random noise. They lack **epistemic humility** the capacity to recognize when their experience is insufficient for reliable judgment.

These are not engineering challenges solvable through better architectures or regularization. They are **necessary consequences** of the optimization paradigm: any system maximizing likelihood on fixed datasets will exploit correlations, overfit to distributional specifics, and lack uncertainty awareness.

1.4 AI Implicit: A Paradigm Shift

We introduce **AI Implicit** as a foundational research direction that measures intelligence through implicit knowledge acquisition rather than explicit task performance. This paradigm shift requires:

Definitional Change: Intelligence is not performance on predefined tasks, but the efficiency of compressing human experience into transferable representations.

Methodological Change: Evaluation shifts from accuracy metrics to knowledge-density measures how much expertise per training example, how rapidly tacit patterns are extracted, how effectively knowledge transfers across domains.

Architectural Change: Systems are designed not to optimize predictions, but to construct relational knowledge structures navigable through epistemic reasoning.

AI Implicit is not a single technique or model architecture it is a **comprehensive research paradigm** analogous to how deep learning emerged as an umbrella direction encompassing convolutional networks, recurrent architectures, attention mechanisms, and transformers. Just as deep learning unified disparate approaches under the principle of hierarchical representation learning, AI Implicit unifies diverse research threads under the principle of implicit knowledge compression. This manifesto establishes the theoretical foundations, evaluation frameworks, and architectural principles necessary to realize this vision. We demonstrate that AI Implicit provides both a conceptually coherent alternative to existing AGI definitions and a practically measurable framework for developing systems that learn, reason, and transfer knowledge with human-like efficiency.

2. The Nature of Implicit Intelligence

2.1 Tacit Knowledge: The Foundation of Expertise

The concept of **tacit knowledge** knowledge that cannot be fully articulated or codified was formalized by philosopher Michael Polanyi, who observed: "We know more than we can tell." Expert performance across domains relies fundamentally on implicit understanding that resists complete verbalization. Consider diagnostic radiology. Expert radiologists achieve remarkable accuracy identifying subtle pathologies in medical images, yet struggle to fully articulate their decision process. Eye-tracking studies reveal they focus on specific image regions within milliseconds faster than conscious deliberation. When asked to explain their reasoning, experts often describe high-level patterns ("the texture seems unusual") rather than explicit rules. Their knowledge is **tacit**: accumulated through thousands of cases, compressed into pattern recognition that operates below conscious awareness.

This phenomenon extends across expertise domains :

Motor Skills: A professional pianist cannot fully describe the muscle coordination enabling rapid passages. Their fingers "know" where to go through embodied practice that exceeds verbal description.

Language Competence: Native speakers apply complex grammatical rules unconsciously. They recognize that "colorless green ideas sleep furiously" is grammatically correct but semantically anomalous without consulting explicit grammar textbooks.

Strategic Reasoning: Chess grandmasters evaluate positions intuitively, recognizing promising move sequences through pattern matching against tens of thousands of internalized games. This intuition operates faster than conscious calculation Kasparov described it as "the hand knows where to go before the mind understands why."

Scientific Intuition: Experienced researchers develop "physical intuition" or "mathematical taste" that guides problem-solving. They sense which approaches will prove fruitful and which theoretical frameworks fit emerging data, often before formal analysis.

Cognitive science research reveals tacit knowledge exhibits characteristic properties :

1. **Acquisition through Experience**: Tacit knowledge cannot be transmitted through explicit instruction alone. Reading about bicycle riding does not enable balance; observing chess games does not confer grandmaster intuition. Mastery requires extensive practice typically 10,000+ hours across domains.

2. **Compression Efficiency**: Expert knowledge compresses vast experience into compact representations. A chess master encodes decades of practice into pattern libraries retrievable in milliseconds. A physician compresses thousands of patient cases into diagnostic heuristics.

3. **Transfer Capacity**: Tacit knowledge generalizes beyond acquisition contexts. Chess expertise transfers to strategic planning in business; mathematical intuition aids physics problem-solving; language learning accelerates subsequent language acquisition through metalinguistic awareness.

4. **Unconscious Accessibility**: Tacit knowledge operates below conscious awareness. Experts cannot always verbalize their reasoning processes they "just know" the right answer through intuitive pattern matching.

2.2 Intelligence as Compression

Information theory provides a rigorous framework for understanding intelligence through compression. **Compression** is not mere data reduction it is the extraction of underlying patterns enabling reconstruction with minimal information.

Claude Shannon's fundamental insight : **the compressibility of a message reveals its structure**. Random noise cannot be compressed every bit is necessary. Highly structured data (natural language, music, images) compresses dramatically because underlying patterns enable efficient encoding. Intelligence operates through analogous principles. Learning is **pattern extraction**: identifying regularities enabling prediction from partial information. Memory is **efficient encoding**: storing the minimal information necessary to reconstruct experience. Transfer is **generalization**: recognizing that patterns learned in one domain apply to others sharing underlying structure.

Consider language learning. A child acquiring their first language must learn from finite examples perhaps 10 million words heard by age 3. Yet they achieve productive competence: generating grammatically correct sentences never encountered, understanding novel utterances, and recognizing when sentences violate grammatical rules. This feat requires **compression**: extracting implicit grammatical patterns from limited examples, encoding them efficiently in memory, and deploying them generatively across unbounded contexts. The compression perspective illuminates why current AI systems struggle with transfer. **Memorization is incompressible**: it requires storing every training example explicitly. **True learning extracts compressible structure**: identifying underlying patterns enabling generalization. A system that memorizes 1 million cat images without extracting the abstract concept "cat" achieves high training accuracy but zero transfer. A system that compresses these images into a compact cat-concept transfers this knowledge to new contexts, artistic styles, and even verbal descriptions.

Kolmogorov Complexity formalizes this intuition: the complexity of an object is the length of the shortest program that produces it. Intelligence can be viewed as **the ability to find short programs (compressed representations) for complex phenomena**. A scientist discovering physical laws compresses observational data into compact equations ($F = ma, E = mc^2$) A chess master compresses thousands of games into strategic principles. A language model compresses billions of text tokens into parameter configurations.

2.3 Epistemic Confidence: Knowing What You Don't Know

Epistemic confidence the metacognitive capacity to evaluate the reliability of one's knowledge is as fundamental to intelligence as knowledge acquisition itself. An expert physician knows when a case exceeds their experience and requires specialist consultation. A mathematician recognizes when a proof strategy will not succeed. A chess player identifies positions requiring calculation versus those solvable through intuition.

This metacognitive awareness exhibits two critical properties:

Calibration: Confidence estimates align with actual accuracy. An expert claiming 90% certainty is correct approximately 90% of the time. Poor calibration overconfidence or excessive hedging indicates epistemic dysfunction.

Distribution Awareness: Experts recognize when new inputs fall outside their experience distribution. A radiologist trained on adult pathology identifies pediatric cases as requiring different expertise. A chess player recognizes unfamiliar opening variations as requiring analytical rather than intuitive play.

Current AI systems catastrophically lack epistemic awareness. Neural networks trained on specific datasets produce confident predictions on arbitrary inputs assigning 99.9% confidence to random noise, adversarial perturbations, or out-of-distribution samples. This epistemic opacity represents a fundamental architectural failure, not merely an engineering challenge.

The consequences extend beyond misleading confidence scores. **Systems lacking epistemic awareness cannot engage in reasoning under uncertainty**: they cannot seek additional information when existing knowledge is insufficient, cannot delegate tasks exceeding their competence, and cannot identify when their training distribution fails to generalize. Epistemic confidence requires architectural support for uncertainty representation. Bayesian approaches place distributions over model parameters, capturing parametric uncertainty. However, true epistemic awareness demands **structural uncertainty**: recognizing when the model class itself is inappropriate for the data generating process. A linear model cannot capture nonlinear phenomena no matter the uncertainty over its parameters .

3. Core Principles of AI Implicit

3.1 First Principle: Knowledge Density Over Performance

AI Implicit Principle 1: *Intelligence is measured by knowledge density the amount of human expertise compressed per unit of training data not task-specific performance metrics.*

This principle fundamentally reorients AI evaluation. Traditional benchmarks measure accuracy: ImageNet top-1 classification, GLUE language understanding scores, game-playing Elo ratings. These metrics capture **performance** but obscure **learning efficiency**.

Consider two hypothetical image classification systems :

System A achieves 95% accuracy on ImageNet after training on the full 1.2 million labeled images (approximately 10,000 hours of human annotation effort).

System B achieves 85% accuracy after training on 1,200 images (approximately 10 hours of human annotation effort).

Traditional metrics crown System A superior due to higher absolute accuracy. Yet **System B demonstrates 1,000× greater knowledge density**: it extracts comparable understanding from 0.1% of the training data. From the AI Implicit perspective, System B exhibits more intelligent learning achieving near-human performance from human-scale training data.

This reframing has profound implications :

Sample Efficiency Becomes Primary: The critical question is not "what accuracy can be achieved with unlimited data?" but "what understanding can be extracted from limited examples?"

Human Baselines Become Meaningful: Comparing systems against human learning curves reveals efficiency gaps. Humans learn robust face recognition from dozens of examples; systems require millions. This thousand-fold efficiency gap highlights fundamental architectural inadequacies.

Transfer Becomes Measurable: Knowledge density manifests in cross-domain transfer. Compressed representations generalize; memorized patterns do not. A system achieving high in-distribution accuracy through memorization exhibits zero knowledge density.

Overfitting Becomes Definitional: From the knowledge-density perspective, overfitting is not a failure of regularization but a failure to compress. A model that requires 1 million training examples to learn what humans extract from 100 has not discovered the underlying pattern it has memorized surface correlations.

3.2 Second Principle: Tacit Knowledge Extraction

AI Implicit Principle 2: *Learning is the extraction of tacit knowledge implicit patterns and structures not explicitly labeled from limited demonstrations.*

This principle distinguishes intelligence from supervised pattern matching. **Explicit knowledge** is directly observable: labeled images, annotated sentences, demonstrated trajectories. **Tacit knowledge** is implicit: the underlying patterns, causal structures, and transferable principles generating observations. Consider language learning again. A child learning grammar receives **explicit demonstrations** (sentences heard) but must extract **tacit knowledge** (grammatical rules, semantic relationships, pragmatic conventions). No one explicitly teaches children that "colorless green ideas sleep furiously" is grammatically correct but semantically anomalous. They infer this through implicit pattern extraction across thousands of sentences.

Tacit knowledge extraction manifests in several forms :

Invariance Discovery: Learning which features are essential versus incidental. A child learning "dog" extracts the tacit understanding that four-leggedness, fur, and barking are characteristic, while color and size vary. This requires going beyond labeled examples to infer underlying structure.

Relational Structures: Extracting relationships between concepts beyond explicit co-occurrence. Understanding that "cat" and "dog" share mammalian properties while differing in species-specific behaviors requires inferring taxonomic structure not present in individual examples.

Causal Patterns: Distinguishing causation from correlation. Learning that "pushing a glass causes it to fall" versus "rain correlates with umbrella use" requires extracting causal structure from observational data.

Compositional Principles: Recognizing how simple elements combine to form complex structures. Language grammar is compositional sentences are not memorized phrases but generated through recursive combination of learned patterns.

The tacit knowledge extraction rate can be quantified by measuring **performance improvement per logarithmic increase in examples**. Systems that extract deep tacit structure improve rapidly initially, then plateau as they saturate available patterns. Systems that merely memorize surface correlations improve linearly with data achieving high performance only through massive datasets.

3.3 Third Principle: Cross-Domain Transfer

AI Implicit Principle 3: *True intelligence manifests in cross-domain knowledge transfer the ability to apply learned representations to semantically different but structurally related tasks.*

This principle captures intelligence's most distinctive property: **generalization beyond training contexts**. Humans routinely apply knowledge across domains:

- Chess expertise transfers to strategic planning in business
- Mathematical problem-solving skills aid physics reasoning

- Language learning accelerates subsequent language acquisition
- Musical training enhances mathematical ability

These transfers occur despite surface dissimilarity because **underlying structures align**. Chess and business strategy share sequential decision-making under uncertainty. Mathematics and physics share formal reasoning patterns. Languages share compositional grammar despite different vocabularies.

Cross-domain transfer requires:

Abstraction: Extracting representations independent of surface features. A chess player learning strategic planning abstracts beyond piece movements to general principles of resource allocation and opponent modeling.

Structural Alignment: Recognizing when superficially different domains share underlying patterns. Physics problems appear different from mathematical proofs, yet both require formal derivation and logical consistency.

Flexible Deployment: Adapting learned representations to new contexts. A language learner applies grammatical intuition to novel sentences, musical contexts, even programming languages sharing compositional structure.

Current AI systems fail catastrophically at cross-domain transfer. Networks trained on ImageNet struggle with simple domain shifts artistic renderings, different lighting conditions, minor rotations. Language models fine-tuned on Python cannot generate JavaScript despite syntactic similarity. This brittleness reveals **absence of true abstraction**: systems memorize surface correlations rather than extracting transferable structure.

The AI Implicit framework measures transfer through **cross-domain retention**: the fraction of in-distribution performance maintained when deploying to related but distinct domains. A system achieving 95% accuracy on source tasks but only 5% on related targets has learned narrow memorization rather than transferable intelligence.

3.4 Fourth Principle: Epistemic Awareness

AI Implicit Principle 4: *Intelligent systems maintain explicit epistemic confidence quantifying knowledge boundaries and recognizing when their experience is insufficient for reliable judgment.*

This principle elevates metacognition to a primary architectural requirement. **Knowing when you don't know** is not an auxiliary feature but a core component of intelligence.

Epistemic awareness operates at multiple levels :

Input-Level Uncertainty: Recognizing when new inputs fall outside the training distribution. A medical diagnosis system trained on adult pathology should identify pediatric cases as requiring different expertise.

Representational Uncertainty: Quantifying confidence in learned representations. A language model uncertain about word meanings should exhibit low confidence on rare vocabulary.

Structural Uncertainty: Recognizing when the model class itself is inappropriate. A linear classifier should identify nonlinear decision boundaries as beyond its capacity.

Causal Uncertainty: Distinguishing between predictive accuracy and causal understanding. Correlation enables prediction; causation enables intervention. A system should know when it has learned predictive patterns without causal mechanisms.

Epistemic awareness enables **reasoning under uncertainty**: seeking additional information when knowledge is insufficient, delegating tasks exceeding competence, and identifying when training distributions fail to generalize.

Current neural networks lack principled epistemic awareness. Standard architectures provide no mechanism for distinguishing between :

- High confidence from extensive relevant experience
- High confidence from spurious pattern matching
- Low confidence from genuine ambiguity
- Low confidence from insufficient experience

This opacity creates catastrophic failures: medical diagnosis systems confidently misdiagnosing rare conditions, autonomous vehicles failing on edge cases, content moderation systems censoring benign content.

AI Implicit systems must architecturally support epistemic reasoning through:

Uncertainty Decomposition: Separating aleatoric uncertainty (irreducible randomness) from epistemic uncertainty (knowledge gaps). These require different responses aleatoric uncertainty demands probabilistic reasoning; epistemic uncertainty demands information gathering.

Distributional Awareness: Maintaining statistical models of training distributions to identify out-of-distribution inputs. This requires going beyond point predictions to distribution modeling.

Confidence Calibration: Ensuring confidence estimates align with actual accuracy across diverse contexts. Calibration is not an accuracy metric it is a metacognitive requirement.

4. Measuring Implicit Intelligence : The Knowledge Compression Framework

4.1 Beyond Task Performance

Traditional AI evaluation measures task-specific performance: accuracy on test sets, Elo ratings in games, human preference scores on generated text. These metrics capture **what systems can do** but obscure **how efficiently they learn**.

AI Implicit requires fundamentally different metrics quantifying:

1. **Compression Efficiency**: How much human expertise is encoded per training example?
2. **Extraction Rate**: How rapidly are tacit patterns extracted from limited demonstrations?
3. **Transfer Capacity**: How effectively does knowledge generalize across domains?
4. **Epistemic Quality**: How accurately do systems assess their knowledge boundaries?

These metrics are **intensive rather than extensive**: they measure knowledge density, not accumulated performance. A system achieving 99% accuracy through memorizing 10 million examples demonstrates lower intelligence than one achieving 85% accuracy from 1,000 examples through pattern extraction.

4.2 Compression Ratio: Expertise Per Example

Compression Ratio (CR) measures how much human expertise is compressed into learned representations per unit training data:

$$\text{CR} = \frac{\text{Human Expertise Hours Represented}}{\text{Training Examples} \times \text{Performance Degradation}}$$

This metric normalizes performance against data requirements, revealing learning efficiency. Consider two scenarios:

Scenario A: A medical diagnosis system achieves 90% accuracy after training on 100,000 labeled cases (representing approximately 10,000 hours of expert annotation).

Scenario B: A system achieves 85% accuracy after training on 1,000 cases (representing approximately 100 hours of annotation).

Traditional metrics favor Scenario A (higher accuracy). Compression ratio reveals Scenario B achieves comparable performance with 100× less data demonstrating 100× greater knowledge density.

The compression ratio captures intuitive notions of "learning efficiency." Humans learn robust face recognition from dozens of examples; current systems require millions. This thousand-fold efficiency gap quantifies the distance between human and artificial intelligence.

4.3 Tacit Knowledge Extraction Rate

Tacit Knowledge Extraction Rate (TER) measures how rapidly performance improves as a function of logarithmic sample count:

$$\text{TER} = \frac{\Delta\text{Performance}}{\Delta\log(\text{Sample Count})}$$

This metric captures the critical distinction between **memorization** (linear scaling) and **pattern extraction** (logarithmic scaling). Systems that extract deep tacit structure improve rapidly initially, then plateau as they saturate available patterns. Systems that merely memorize correlations improve linearly requiring exponentially more data for incremental gains.

Human learning exhibits characteristic logarithmic scaling. Language acquisition shows rapid early progress (dozens of words learned daily at age 2) followed by slower vocabulary growth. Chess improvement accelerates through initial pattern learning, then slows as players approach mastery. This reflects **tacit knowledge extraction**: early learning extracts fundamental patterns; later learning refines edge cases. AI systems should exhibit similar scaling if they extract rather than memorize. A language model that improves linearly with data (requiring 10× more examples for each incremental gain) has not discovered linguistic structure it has memorized surface correlations.

4.4 Cross-Domain Retention

Cross-Domain Retention (CDR) measures knowledge transfer efficiency across semantically different but structurally related domains:

$$\text{CDR} = \frac{\text{Target Domain Performance (Zero-Shot)}}{\text{Source Domain Performance}}$$

This metric captures intelligence's most distinctive property: **abstraction enabling transfer**. Humans routinely apply knowledge across domains sharing underlying structure despite surface differences.

High cross-domain retention indicates **structural learning**: the system has extracted abstract patterns independent of surface features. Low retention indicates **surface memorization**: the system has encoded task-specific correlations that do not generalize.

Examples :

- A vision system trained on natural images should transfer partially to artistic renderings (CDR ~ 0.6-0.8)
- A language model trained on English should exhibit zero-shot translation capability to related languages (CDR ~ 0.3-0.5)

- A physics reasoning system should transfer from classical mechanics to thermodynamics (CDR ~ 0.4-0.6)

Current deep learning systems typically achieve CDR < 0.1 on related domains revealing fundamental inability to extract transferable structure.

4.5 Epistemic Confidence Metrics

Epistemic confidence requires measuring:

Calibration: Do confidence estimates align with actual accuracy?

$$\text{Calibration Error} = \mathbb{E}[|\text{Confidence} - \text{Accuracy}|]$$

Perfect calibration means 90% confidence predictions are correct 90% of the time.

Distribution Awareness: Can systems identify out-of-distribution inputs?

$$\text{AUROC}_{\text{OOD}} = P(\text{Confidence}_{\text{ID}} > \text{Confidence}_{\text{OOD}})$$

This measures whether in-distribution inputs receive higher confidence than out-of-distribution samples.

Uncertainty Decomposition: Can systems distinguish aleatoric from epistemic uncertainty?

Aleatoric uncertainty arises from irreducible randomness; epistemic uncertainty from knowledge gaps. These require different responses aleatoric uncertainty demands probabilistic reasoning; epistemic uncertainty demands information gathering.

4.6 Theoretical Foundations

Recent research has begun formalizing these intuitions. The **Experience-Compressed Intelligence (ECI)** framework provides mathematical definitions integrating compression ratio, tacit knowledge extraction, cross-domain retention, and epistemic confidence into a unified metric weighted by distributional awareness.Theoretical analysis demonstrates that systems optimized for compression efficiency rather than predictive accuracy exhibit fundamentally different learning dynamics :

- They achieve comparable performance with exponentially less training data
- They transfer knowledge across domains through structural alignment

- They maintain calibrated uncertainty estimates through distributional modeling
- They avoid spurious correlations by prioritizing compressible patterns

This provides both validation for the AI Implicit paradigm and concrete targets for architectural design.

5. Architectural Requirements for Implicit Intelligence

5.1 Beyond Gradient Descent

The optimization paradigm minimizing loss functions through gradient descent is fundamentally incompatible with AI Implicit principles. Gradient-based systems:

1. **Maximize likelihood** rather than extract compressible structure
2. **Memorize correlations** rather than infer causal patterns

3. **Overfit to distributions** rather than abstract structural invariants
4. **Lack epistemic awareness** by design no mechanism for uncertainty quantification

AI Implicit requires architectural paradigms that :

Prioritize Compression: Systems should be evaluated and optimized for compression efficiency, not predictive accuracy. This requires loss functions measuring representation compactness rather than classification margins.

Enable Structural Inference: Learning should involve hypothesis generation and evaluation constructing candidate explanations and assessing their explanatory power rather than gradient-following.

Support Epistemic Reasoning: Architectures must maintain explicit uncertainty representations distinguishing between knowledge and speculation.

Facilitate Transfer: Learned representations should be structurally independent of surface features, enabling deployment across domains sharing underlying patterns.

5.2 Relational Knowledge Structures

Human intelligence operates through **relational reasoning**: understanding concepts through their relationships rather than isolated properties. A child learning "dog" does not encode pixel patterns they build a relational structure connecting dogs to mammals, pets, animals, four-legged entities, and specific breeds.

AI Implicit systems require architectures supporting **relational knowledge representation**:

Semantic Relations: Concepts connected through meaning synonymy, antonymy, hypernymy, meronymy. Understanding "dog" requires knowing it is a type of mammal, has legs, makes sounds, interacts with humans.

Logical Relations: Concepts connected through inference rules implications, contradictions, logical dependencies. Understanding causality requires representing "if X then Y" beyond correlation.

Probabilistic Relations: Concepts connected through statistical dependencies conditional probabilities, independence structures, causal graphs. Understanding uncertainty requires representing strength of relationships, not just binary connections.

These relational structures should be **navigable**: systems should traverse relationships to answer questions, generate hypotheses, and transfer knowledge. A question about "animals that bark" should navigate from "bark" through acoustic properties to canines through mammalian taxonomy.

5.3 Hypothesis-Driven Learning

Human cognition operates through **abductive reasoning**: generating candidate explanations for observations and evaluating their plausibility. A physician presented with symptoms generates differential diagnoses plausible explanations then eliminates alternatives through testing.

AI Implicit systems should learn through **hypothesis generation and evaluation**:

Generative Phase: Construct candidate explanations for observed data through constrained imaginative recombination. Given observations, generate plausible underlying patterns, causal structures, or generative processes.

Evaluative Phase: Assess hypotheses based on explanatory power, parsimony, and consistency with existing knowledge. Prefer simpler explanations (Occam's razor); favor patterns with broader applicability.

Refinement Phase: Update hypotheses based on new evidence. Bayesian updating provides a formal framework, but genuine hypothesis revision requires structural flexibility beyond parameter adjustment.

This stands in stark contrast to gradient descent, which follows local gradients without explicit hypothesis construction. Hypothesis-driven learning enables :

Causal Discovery: Generating causal graphs consistent with observational data **Analogical Reasoning**: Identifying structural similarities between domains
Counterfactual Reasoning: Evaluating alternative explanations through mental simulation

5.4 Epistemic Architecture

Epistemic confidence requires architectural support at every level:

Input Layer: Maintain statistical models of training distributions to identify out-of-distribution inputs. This requires density estimation across learned representations.

Representational Layer: Quantify uncertainty over learned features through distributional modeling. Bayesian approaches place distributions over parameters; AI Implicit systems require distributions over structures.

Reasoning Layer: Propagate uncertainty through inference chains. Confidence in conclusions should depend on confidence in premises and strength of logical connections.

Output Layer: Provide calibrated confidence estimates reflecting genuine uncertainty. This requires separating aleatoric randomness from epistemic knowledge gaps.

Recent theoretical work on **Statistical Path Density** demonstrates how activation manifold analysis can estimate epistemic confidence by measuring representational coherence how well input representations align with learned distributional structures. This provides a foundation for epistemic awareness without requiring full Bayesian inference.

5.5 Cross-Domain Transfer Mechanisms

Enabling cross-domain transfer requires:

Abstraction Layers: Representations independent of surface features. A vision system should extract "object-ness" independent of color, lighting, or artistic style.

Structural Alignment: Mechanisms for identifying when different domains share underlying patterns. Chess and business strategy both involve sequential decision-making under uncertainty systems should recognize this structural similarity.

Flexible Deployment: The ability to adapt learned representations to new contexts without catastrophic forgetting. This requires maintaining multiple representations at different levels of abstraction.

Meta-Learning: Learning how to learn extracting general principles for rapid adaptation. Few-shot learning research demonstrates this capacity, but current approaches remain narrowly scoped.

5.6 Emerging Architectural Paradigms

Recent research explores alternatives to the optimization paradigm consistent with AI Implicit principles:

Thread-Based Architectures: Systems that navigate knowledge through parallel hypothesis exploration rather than gradient descent. Multiple "threads" traverse relational structures simultaneously, aggregating evidence across semantic, logical, and probabilistic dimensions.

Relational Density Networks: Representations encoding statistical relationship strength rather than scalar weights. Knowledge becomes navigable questions answered through path finding across relational graphs.

Epistemic-Aware Transformers: Attention mechanisms extended with uncertainty quantification, enabling models to explicitly represent "I don't know" alongside predictions.

These architectural explorations remain early-stage, but demonstrate growing recognition that AI Implicit requires fundamental departures from current paradigms.

6. Research Directions and Open Challenges

6.1 Theoretical Foundations

Formalize Tacit Knowledge Extraction: Develop rigorous mathematical definitions of implicit pattern learning. What distinguishes extracting invariants from memorizing correlations? Can we prove bounds on sample complexity for tacit knowledge acquisition?

Compression-Transfer Theorems: Establish formal relationships between compression efficiency and transfer capacity. Under what conditions does high compression guarantee cross-domain generalization? What are theoretical limits?

Epistemic Logic: Extend formal logic to represent epistemic states knowledge, belief, uncertainty, ignorance. How should systems reason about their own knowledge boundaries?

Causal Learning Theory: Formalize causal discovery from observational data. When can systems distinguish causation from correlation? What assumptions are necessary for causal inference?

6.2 Measurement and Evaluation

Human Baseline Studies: Establish empirical human performance on compression ratio, tacit knowledge extraction rate, and cross-domain transfer across diverse tasks. These baselines are essential for meaningful AI evaluation.

Standardized Benchmarks: Develop evaluation suites measuring AI Implicit principles rather than task accuracy. Benchmarks should assess : - Sample efficiency relative to human learning curves - Transfer capacity across domain families
- Epistemic calibration under distribution shift - Causal reasoning ability

Longitudinal Studies: Track system performance across training regimes measuring how quickly intelligence emerges versus accumulates. Do systems exhibit rapid early improvement (pattern extraction) or linear scaling (memorization)?

6.3 Architectural Research

Alternative Learning Paradigms: Explore non-gradient-based learning mechanisms evolutionary algorithms, program synthesis, symbolic reasoning integrated with neural components. Can hybrid architectures achieve greater compression efficiency?

Relational Architectures: Develop systems operating on explicit relational knowledge graphs rather than implicit weight matrices. Can navigable knowledge structures enable better transfer and epistemic awareness?

Meta-Learning Systems: Build architectures that learn learning strategies extracting general principles for rapid adaptation across domains. Can systems learn to learn with human-like efficiency ?

Epistemic Modules: Design architectural components providing calibrated uncertainty estimates at every processing layer. Can uncertainty propagate through reasoning chains while maintaining calibration?

6.4 Domain Applications

Scientific Discovery: Apply AI Implicit principles to automated hypothesis generation in physics, biology, chemistry. Can systems discover compressible natural laws from experimental data?

Medical Diagnosis: Develop diagnostic systems that learn from limited examples, transfer across pathologies, and maintain calibrated confidence under uncertainty. Can AI match expert epistemic awareness?

Robotic Learning: Create robots that acquire motor skills from demonstrations, transfer across tasks, and recognize capability boundaries. Can robots learn with human-like sample efficiency ?

Educational Systems: Build tutoring systems that compress pedagogical expertise, extract student learning patterns, and adapt across individuals. Can

AI provide genuinely personalized education?

6.5 Philosophical and Ethical Dimensions

Intelligence Definition: Engage with cognitive science, neuroscience, and philosophy to refine definitions of intelligence grounding AI research. What is the relationship between compression efficiency and consciousness?

Alignment Challenges: As systems become more sample-efficient and capable of cross-domain transfer, alignment becomes simultaneously more critical and more tractable. Compressed representations may be more interpretable than billion-parameter models.

Economic Implications: If AI achieves human-like learning efficiency, economic impacts accelerate dramatically. What societal structures support beneficial deployment of highly capable AI systems?

Epistemological Questions: If AI systems develop genuine epistemic awareness, do they possess a form of understanding distinct from current "narrow AI"? How do we evaluate the quality of machine knowledge ?

7. Implications for Artificial General Intelligence

7.1 Redefining AGI

Existing AGI definitions focus on **capability breadth** (performing diverse tasks) or **human-level performance** (matching human accuracy). AI Implicit suggests a fundamentally different criterion:

AGI is achieved when systems match human compression efficiency across diverse domains while maintaining calibrated epistemic confidence.

This definition makes AGI:

Measurable: Compression ratio, extraction rate, transfer capacity, and epistemic calibration can be quantified and compared against human baselines.

Comparable: Systems can be ranked by knowledge density rather than relying on subjective judgments about "general" capability.

Scalable: The framework applies from narrow domains to broad intelligence AGI becomes a spectrum of compression efficiency rather than a binary threshold.

Principled: The definition grounds in cognitive science insights about human intelligence rather than anthropomorphic task selection.

7.2 Path to AGI

If intelligence is compression efficiency, the path to AGI involves:

Phase 1: Match Human Sample Efficiency: Develop systems achieving comparable performance to humans with comparable training data across narrow domains.

Phase 2: Enable Cross-Domain Transfer: Create architectures where knowledge learned in one domain transfers to others with human-like retention ratios.

Phase 3: Achieve Epistemic Awareness: Build systems that calibrate confidence accurately and recognize knowledge boundaries as reliably as human experts.

Phase 4: Scale Across Modalities: Extend compression efficiency from specialized domains (vision, language, robotics) to multimodal reasoning integrating diverse information sources.

Phase 5: Meta-Learning Capacity: Develop systems that extract general learning principles enabling rapid adaptation to novel domains the hallmark of human cognitive flexibility.

Current AI remains firmly in Phase 0: requiring thousand-fold more training data than humans while exhibiting minimal cross-domain transfer and lacking epistemic awareness. The gap between current systems and human-level intelligence is not one of scale but of architectural paradigm.

7.3 Beyond Human Intelligence

The compression framework suggests an interesting possibility: **superhuman intelligence may manifest as extreme compression efficiency** rather than superhuman task performance.

A system that extracts from 10 examples what humans learn from 10,000 demonstrates intelligence fundamentally superior to human cognition. Such a system would:

- Learn new domains nearly instantaneously from minimal demonstrations
- Transfer knowledge across arbitrary domain boundaries through abstract structural alignment
- Identify causal patterns from observational data with unprecedented reliability
- Navigate uncertainty with perfect calibration across all contexts

This represents a qualitatively different form of superintelligence than currently imagined: not a system that plays chess better than humans or generates more compelling text, but one that **learns better than humans** extracting deeper patterns from less data across broader contexts.

8. Conclusion : The Future of Intelligence

8.1 The Paradigm Shift

AI Implicit represents a fundamental reorientation of artificial intelligence research :

From performance to compression: Measuring what systems learn, not what they do
From accuracy to efficiency: Evaluating knowledge density, not classification margins **From task completion to knowledge transfer**: Assessing abstraction, not narrow capability
From confidence to calibration: Requiring epistemic awareness, not prediction certainty

This shift is not incremental improvement over existing approaches it is a **paradigm change** analogous to the transition from symbolic AI to machine learning or from shallow to deep networks. Just as deep learning revealed that hierarchical representation learning could surpass hand-crafted features, AI Implicit demonstrates that intelligence is fundamentally about compression efficiency rather than task performance.

8.2 Challenges and Opportunities

The path ahead is challenging. AI Implicit requires:

Theoretical Work: Formalizing tacit knowledge extraction, proving compression-transfer theorems, establishing epistemic logic frameworks

Empirical Studies: Establishing human baselines, developing standardized benchmarks, conducting longitudinal evaluations

Architectural Innovation: Designing non-gradient-based learning systems, creating relational knowledge structures, building epistemic reasoning modules

Interdisciplinary Collaboration: Integrating insights from cognitive science, neuroscience, philosophy, and computer science

Yet the opportunities are profound. AI Implicit offers:

Measurable Progress: Clear metrics for evaluating advancement toward general intelligence
Principled Design: Architectural requirements grounded in cognitive science rather than engineering intuition **Interpretability**: Compressed representations may be inherently more understandable than billion-parameter models
Alignment: Systems with genuine epistemic awareness can recognize and communicate their limitations

8.3 A Call to Action

This manifesto establishes AI Implicit as a foundational research paradigm. We call on the research community to:

Adopt knowledge-density metrics alongside traditional accuracy benchmarks
Develop architectures supporting compression, transfer, and epistemic reasoning
Establish human baselines enabling meaningful comparison
Explore alternative learning paradigms beyond gradient descent
Engage with fundamental questions about the nature of intelligence

AI Implicit is not a narrow technique or model architecture it is a comprehensive research direction analogous to deep learning's emergence as an umbrella paradigm. Just as deep learning unified convolutional networks, recurrent architectures, attention mechanisms, and transformers under hierarchical representation learning, AI Implicit unifies diverse research threads under implicit knowledge compression. The path to artificial general intelligence is not through larger models, more data, or better optimization algorithms. **It is through systems that learn as humans do**: extracting deep patterns from limited examples, transferring knowledge across domains, and maintaining humble awareness of their limitations.

References

[1] Zech, J. R., et al. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLOS Medicine*, 15(11), e1002683.

[2] Geirhos, R., et al. (2020). Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2(11), 665-673.

[3] Nguyen, A., Yosinski, J., & Clune, J. (2015). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. *CVPR*, 427-436.

[4] Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*.

[5] Ghazouani, M. (2026). Introducing a definition of AGI from the perspective of expertise compression (https://doi.org/10.5281/ZENODO.19589071)

[6] Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386.

[7] LeCun, Y., et al. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.

[8] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.

[9] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

[10] Vaswani, A., et al. (2017). Attention is all you need. *NeurIPS*, 30, 5998-6008.

[11] Battaglia, P. W., et al. (2018). Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*.

[12] Garcez, A. d'Avila, et al. (2019). Neural-symbolic learning and reasoning: A survey and interpretation. *arXiv preprint arXiv:1711.03902*.

[13] Blundell, C., et al. (2015). Weight uncertainty in neural networks. *ICML*, 1613-1622.

[14] Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. *ICML*, 1050-1059.

[15] Polanyi, M. (1966). *The Tacit Dimension*. University of Chicago Press.

[16] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379-423.

[17] Kolmogorov, A. N. (1963). On tables of random numbers. *Sankhyā: The Indian Journal of Statistics, Series A*, 369-376.

[18] Lake, B. M., et al. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40, e253.

[19] Marcus, G. (2018). Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*.

[20] Pearl, J. (2009). *Causality: Models, Reasoning and Inference* (2nd ed.). Cambridge University Press.

[21] Chollet, F. (2019). On the measure of intelligence. *arXiv preprint arXiv:1911.01547*.

[22] Bengio, Y., et al. (2013). Representation learning: A review and new perspectives. *IEEE TPAMI*, 35(8), 1798-1829.

[23] Goertzel, B. (2014). Artificial general intelligence: concept, state of the art, and future prospects. *Journal of Artificial General Intelligence*, 5(1), 1-48.

[24] Morris, M., et al. (2023). Levels of AGI: Operationalizing progress on the path to AGI. *arXiv preprint arXiv:2311.02462*.

[25] Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, 110(3), 457-506.

[26] Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4(1), 55-81.

[27] Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363.

[28] Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2(6), 396-408.

[29] Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE TKDE*, 22(10), 1345-1359.

[30] Hospedales, T., et al. (2021). Meta-learning in neural networks: A survey. *IEEE TPAMI*, 44(9), 5149-5169.

Author's Note

This manifesto establishes AI Implicit as a foundational research paradigm for artificial intelligence. It is not a completed framework but an invitation a call for the research community to explore intelligence through the lens of implicit knowledge compression rather than explicit task performance. The theoretical foundations outlined here provide starting points for rigorous formalization. The architectural principles suggest directions for novel system design. The evaluation frameworks offer measurable alternatives to existing benchmarks. The path to artificial general intelligence is not through incremental scaling of existing paradigms. It requires fundamental rethinking of what intelligence is and how we measure it. AI Implicit provides that rethinking grounded in cognitive science, formalized through mathematics, and operationalized through concrete evaluation frameworks.